Neutral Mechanisms: On the Feasibility of Information Sharing*

Ernesto Rivera Mora†

October 26, 2025

Abstract

I analyze a broad class of information-sharing mechanisms called neutral mechanisms. They include cheap talk, noisy communication, mediation, money burning, and transfer schemes. This paper develops a reduced-form approach that characterizes agents' payoffs through belief-dependent utilities, yielding two main results. First, if a supermodularity condition between the state and the agents' hierarchies of beliefs is satisfied, then there exists a neutral mechanism that fully reveals the private information of the informed party. Second, if a submodularity condition is satisfied, sharing information becomes impossible via neutral mechanisms. The paper provides applications related to policymaking, industrial organization, and survey design.

*I thank Amanda Friedenberg, Andreas Blume, Inga Deimen, Stanley Reynolds, Derek Lemoine, Ina Taneva, Tilman Börgers, Heng Liu, David Miller, Harry Pei, Oleg Muratov, Alessandro Pavan, Quitze Valenzuela, Elliot Lipnowski, Ian ball, and Jacopo Perego for their helpful comments.

†University of Colorado, Boulder. Email: ernesto.riveramora@colorado.edu

1 Introduction

Throughout history, adverse outcomes have resulted from the concealment of information. For instance, the tobacco industry took measures 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, contributing to the emergence of the so-called "opioid epidemic" [Meier, 2018]. These recurring negative outcomes raise important questions regarding the introduction of new products. Can policymakers design mechanisms that encourage firms to reveal harmful properties of their products? Which mechanisms incentivize information sharing? Which mechanisms are destined to fail?

This paper studies the extent to which a broad class of *neutral* mechanisms can or cannot induce an informed party to share its private information. To address this question, the paper analyzes a general model with two agents: an expert and a layman. In the absence of means to share information, the agents play a game in which payoffs depend on their actions and the state of the world. For example, the game may involve a regulator deciding whether to ban or approve a potentially addictive product of a pharmaceutical firm. A designer seeks to construct a mechanism in which the expert shares information prior to the game being played. A legislator (the designer) might aim to establish legal frameworks that incentivize the firm to disclose information about their product's safety to a regulator. This paper investigates whether it is feasible or infeasible to construct a neutral mechanism in which the expert reveals the state to the layman.

Neutral mechanisms allow agents to exchange messages, privately receive signals about the behavior of the other agent, and exchange monetary transfers. These mechanisms are extensive forms that satisfy four independence conditions. First, they satisfy *structural independence*. That is, action sets, information sets, and action correspondences are independent of the state. So, the expert has no actions that directly reveal her private information. This rules out hard evidence and disclosure [Milgrom, 1981, Grossman, 1981]. Second, they satisfy *statistical independence*. That is, the likelihoods of chance moves do not directly depend on the state. So, the layman only receives information about the state through the expert's actions in the mechanism. This rules out Blackwell [1953] experiments used in the information design literature [Kamenica and Gentzkow, 2011, Rayo and Segal, 2010, Taneva, 2019, Bergemann and Morris, 2019]. Third, they 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, in which the realized transfer depends on the state. As a consequence, Spence-style signaling [Spence, 1978] is ruled out. Fourth, they 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 behavior in the after-game through how information affects the agents' posterior beliefs. So, while the agents 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], in which the transfers depend on the actions chosen in the after-game.

Mechanisms that satisfy these four requirements 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 [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 but not on the state) [Myerson, 1982, Krishna and Morgan, 2008], among others. Their popularity stems from the fact that they do not require the strong assumptions that other mechanisms need: In many environments, (1) the expert lacks verifiable evidence; (2) the designer cannot access Blackwell experiments; (3) the state is unverifiable, so state-contingent transfers are not feasible; (4a) the designer cannot 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?

Determining the feasibility of information sharing in a neutral mechanism requires an analysis of dynamic games with asymmetric information. Each dynamic game is composed of a neutral mechanism—in which agents potentially share information—

[1] Transfers contingent on the after-game require payments to occur after actions are chosen; therefore, the credibility of such transfers depends on enforceable contracts. However, in many settings, behavior in the after-game may not be contractible due to verification challenges or legal constraints.

followed by the after-game. Typically, dynamic games are solved using backward induction: The analyst first characterizes behavior at the last move and proceeds backwards in the tree. However, in our context, employing backward induction is not trivial because the dynamic game may lack observable actions. In particular, the information conveyed to the layman through the mechanism may not be commonly known at the conclusion of the mechanism. (For example, in mediation, the expert does not observe the interaction between the layman and the mediator.) Hence, each mechanism may lead to a non-trivial Bayesian game that requires its own analysis.

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. So, for the purpose of evaluating a given mechanism, it is not important to understand the behavior in the after-game, but rather the value of the information that the mechanism conveys. With this in mind, the paper develops a *reduced-form approach*, using belief-dependent utilities to summarize how information impacts the agents' payoffs in the after-game. Specifically, the paper analyzes an auxiliary mechanism design problem with belief-dependent preferences. (See Rivera Mora [2024].) In this auxiliary problem, the designer chooses a neutral mechanism in which the agents interact (but do not engage in any after-game) and receive belief-dependent payoffs. So, in this sense, each mechanism induces a dynamic psychological game [Geanakoplos, Pearce, and Stacchetti, 1989, Battigalli and Dufwenberg, 2009]. The paper characterizes the possibility or impossibility of information sharing by linking equilibria of this auxiliary problem to equilibria of the original mechanism design problem.

Implementing the reduced-form approach involves some subtleties. First, the belief-dependent 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]. Second, if there are multiple equilibria of the after-game, a single profile of belief-dependent utilities may not be sufficient to characterize equilibrium payoffs in the after-game. Hence, we will look at a set of belief-dependent utility profiles, called a *reduced-form representation*.

The paper uses the reduced-form approach to provide two results regarding information sharing. It does so by introducing conditions on reduced forms that 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. Conversely, the submodularity condition captures the idea that an expert that observes a low state has strictly higher incentives to "be perceived" as having observed a high state.

The first theorem is a positive result. Loosely speaking, if the expert's reduced form satisfies the supermodularity condition, then full revelation of the state is possible. 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. In multiple settings, full-revelation of the state often maximizes a given objective of the designer. In such cases, the positive result directly solves the designer's problem.[2]

The second theorem is a negative result. Loosely speaking, sharing any payoff-relevant information 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 information) can be transmitted. Thus, the agents behave as if they had not interacted in the mechanism at all. Consequently, this result directly solves the designer problem, as there is nothing that the designer can achieve within this class of mechanisms.

In applications, it is often simple to construct reduced forms and to verify supermodularity and submodularity conditions. To illustrate this, the paper provides a characterization of information sharing in two distinct parametrized classes of games. Section 8.1 analyzes a class in which only the layman takes an action and Section 8.2 analyzes a class in which 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 result, the analysis provides a complete taxonomy of the parameters that enable or prevent information sharing, while offering insights into how these parameters influence information transmission.

[2]Moreover, if the agents' outside options are sufficiently low, the designer can fully reveal the state using a budget-balanced transfer scheme, as the agents would be willing to cover the required transfers. If their outside options are high, the designer may need to use either money burning or a budget breaker.

The first application involves a policymaker and a bureaucrat. The bureaucrat knows the state of the world. The policymaker is responsible for selecting policies across multiple tasks, each of which is relevant to both agents. The application characterizes the types of disagreement that allow for or preclude information sharing. The key is that the absolute level of disagreement is irrelevant for information sharing. Rather, what matters is how changes in the state influence the "direction" of the agents' preferred vector of policies. Applying the main results, full revelation of the state is possible if and only if the agents face "directional agreement," i.e., if their preferred policy vectors move in a "similar direction" in response to changes in the state. Conversely, if agents face "directional disagreement," i.e., if the agents' preferred policy vectors move in "opposite directions" as the state changes, then no relevant information can be transmitted.

The second application analyzes the interaction between two firms competing in a duopoly market that competes either in prices or quantities. One of the firms observes the industry demand. Should an antitrust agency worry about firms sharing information? Or does the strategic interaction between firms prevent any information sharing from happening? First, when the firms' actions are strategic complements (such as in price competition), full revelation of the state is possible. Second, when the firms' actions are strategic substitutes (such as in quantity competition), information sharing becomes infeasible.[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 a bad-news firm. Hence, the good-news firm has higher incentives to induce "optimistic beliefs," and the supermodularity condition is satisfied. 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.

A third application broadens the scope of the paper by analyzing settings with intrinsic belief-dependent preferences. In the spirit of Warner [1965], a researcher attempts to elicit a subject's private trait. However, the subject has image concerns and cares about how he is perceived. The application characterizes when the researcher

[3]While these conclusions bear resemblance to results in Vives [1984], Gal-Or [1985], and Raith [1996], they focus on non-neutral mechanisms that permit verifiable signals. An exception is Ziv [1993], who investigates neutral mechanisms for sharing information about production costs. Note that our second application considers a different source of uncertainty: demand shocks.

can learn the subject's true trait. Full revelation is possible when the subject with the acceptable trait has at least the same incentives to be perceived positively as the subject with the stigmatized trait. Conversely, no information can be shared when the stigmatized subject has strictly more incentives to be perceived positively.

Literature review 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], Ziv [1993], Myerson [1982], Austen-Smith and Banks [2000], Aumann and Hart [2003], Blume, Board, and Kawamura [2007], Krishna and Morgan [2008], Goltsman, Hörner, Pavlov, and Squintani [2009] for examples of neutral mechanisms and Spence [1978], Milgrom [1981], Grossman [1981], Dessein [2002], Rayo and Segal [2010], Kamenica and Gentzkow [2011], Taneva [2019], Bergemann and Morris [2019], McClellan and Rappoport [2024] for examples of non-neutral mechanisms.

Much of the literature focuses on a particular mechanism and a particular after-game and shows that *only* partial information sharing is feasible. 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 fully reveal the state. Ottaviani [2000], Austen-Smith and Banks [2000], Kartik [2007], and Krishna and Morgan [2008] focus on environments in which only the layman can choose an action and the agents' payoffs are supermodular in the action and state. (Their result is related to Application 8.1.) 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 in *each* neutral mechanism and *each* equilibrium, no payoff-relevant information is transmitted.

This paper contributes to a growing literature of mechanisms with after-games [Calzolari and Pavan, 2006a,b, 2009]. A significant strand of this literature employs the *communication revelation principle* [Myerson, 1982] to analyze mechanisms with after-games. Although this seminal result characterizes the set of implementable equilibrium outcomes, it neither identifies sufficient conditions for full revelation nor identifies conditions that prevent sharing payoff-relevant information. A second strand of this literature uses belief-dependent utility functions to summarize payoffs of after-games.[4] For instance, Dworczak [2020] employs belief-dependent utilities to capture bidders' payoffs in auctions with aftermarkets. Morris [2001] and Ottaviani and Sørensen [2006] use belief-dependent utilities to summarize reputational payoffs in cheap talk models with future interactions. These papers focus on mechanisms in which the outcome of the mechanism is publicly observed. Consequently, first-order beliefs uniquely determine higher-order beliefs, and belief-dependent utilities need only depend on first-order beliefs. In contrast, this work studies a wide class of mechanisms, including those in which the mechanism outcome is not publicly observable (such as mediation). In these mechanisms, first-order beliefs do not determine higher-order beliefs. Because higher-order beliefs can impact the after-game, belief-dependent utilities must account for the full hierarchy of beliefs.

The results in this paper share similarities with classical results on implementation as such in Spence [1974], Mirrlees [1976], and Rochet [1987]. However, several challenges limit their direct application here. First, it is unclear how those results apply in environments in which both agents take actions. In particular, it is unclear whether verifying supermodularity between the state and the layman's action is sufficient, or whether the expert's action and its interaction with both the state and the layman's action must also be considered.[5] Second, even if only the layman is active, classical results may be difficult to apply when the action space is multidimensional and lacks a complete order.[6] Third, even within a belief-based framework, classical results introduce additional subtleties. Unlike material allocations—which the designer can choose directly—hierarchies of beliefs emerge endogenously from agents' strategic interaction in the mechanism. Moreover, hierarchies are infinite-dimensional, making it unclear how to order or compare them, a gap addressed here by introducing acute statistics. (See Section 6.1.)

At first glance, the supermodularity condition might resemble the supermodularity condition in Van Zandt and Vives [2007]. However, the conditions are quite different.

[4] The reduced-form approach is particularly useful in applications with infinite action spaces (e.g., Section 8.2). Notice, the communication revelation principle requires analyzing double deviations in which agents both misreport information and disobey actions recommended by the mechanism. If action spaces are infinite, the analysis may become intractable, as it requires examining all possible combinations of misreports and all possible deviations for each recommended action.

[5] In application 2, all the co-modular relations between the state, the expert's action, and the layman's action are relevant. The belief-based approach simplifies the analysis by aggregating these relations into a single condition between the state and the expert's hierarchy of beliefs.

[6] Application 1 illustrates how the results here generate a simple geometric condition characterizing information sharing in settings with multidimensional actions.

While their notion is defined in terms of actions and a single parameter that captures both payoff and belief types, here it is defined in terms of payoff-relevant states and hierarchies.[7]

The paper is organized as follows. Section 2 presents an illustrative example. Section 3 introduces the model. Section 4 develops the auxiliary problem with belief-dependent preferences. Sections 5 and 6 establish the two central results. Section 7 elaborates on the reduced-form approach. Section 8 provides three applications, and Section 9 discusses some final remarks. All proofs not included in the main text are included in the Appendix.

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 $\mu(\overline{\theta}) \in (0,1)$. 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 when the painkiller is *approved* and safe. When the painkiller is *approved* and addictive, profit is c. The parameter c can be greater than 1 if the firm's profit increases when the painkiller is addictive. However, 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. Importantly, c is assumed to be exogenous.

The firm knows whether the painkiller is safe ($\overline{\theta}$) or addictive ($\underline{\theta}$), while the regulator lacks this information. Given the prior, the regulator would *ban* the painkiller.

[7]Their paper establishes 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 may be unclear how to interpret the order on belief types. By contrast, here, the order on hierarchies is directly inherited from the order on states.

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, regardless 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 mechanisms: 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. The main text will consider general neutral mechanisms.

After the agents interact in the mechanism—that is, after the firm chooses its message and the regulator observes it—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.[8]

The regulator's behavior depends on the posterior belief $p(m)$. The reduced-form approach directly references those beliefs. Specifically, the approach characterizes the agents' equilibrium payoffs in terms of posterior beliefs, transforming the original problem into an auxiliary mechanism design problem with belief-dependent preferences. In this auxiliary problem, agents first interact in the mechanism, then update their beliefs, and then receive belief-dependent utilities.[9] This approach lets us an-

[8]Since the message is publicly observed, the firm assigns probability one to the regulator assigning probability $p(m)$ to $\overline{\theta}$ (safe). So, $p(m)$ captures all the information about hierarchies of beliefs.

[9]Formally, a game with belief-dependent preferences is called a psychological game. While psychological games typically model psychological motivations, in this example they serve as a tool in which belief-dependent payoffs summarize the payoffs of the regulation after-game.

alyze the distribution of beliefs that can emerge in equilibrium, revealing whether information sharing is feasible or not.

With this in mind, we begin by characterizing the agents' equilibrium payoffs in terms of beliefs. Note, *ban* is optimal if and only if $p(m) \in \left[0, \frac{1}{2}\right]$ and *approve* is optimal if and only if $p(m) \in \left[\frac{1}{2}, 1\right]$. Thus, the payoff the regulator can get as a function of p—i.e., the posterior that the state is $\overline{\theta}$—is $u_{\text{reg}}(p) = \max\{1-p, p\}$. The firm's payoffs depend on the likelihood that the firm assigns to the regulator choosing *approve*. Because this depends on the regulator's posterior, this likelihood can be summarized by a mapping approve : $[0,1] \to [0,1]$ with

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

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-dependent utility function is

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

Any such pair of functions $(u_{\text{firm}}, u_{\text{reg}})$ defines a psychological game and is called a *reduced form* of the regulation after-game.[10] The main results speak to the possibility of information sharing based on the reduced form of the informed party. Loosely:

Main Theorem:

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

(2) If u_{firm} satisfies a submodularity condition, then no (payoff-relevant) information can be transmitted.

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

[10]Each value $x \in [0,1]$ defines a different reduced form. (The regulator mixes in case of indifference).

corresponds to the standard notion, i.e.,

$$u_{\text{firm}}(\overline{\theta}, 1) - u_{\text{firm}}(\overline{\theta}, 0) \geq u_{\text{firm}}(\underline{\theta}, 1) - u_{\text{firm}}(\underline{\theta}, 0).$$

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 in which the firm is fully revealed as either safe ($p(m) = 1$) or addictive ($p(m) = 0$). The idea will be introduced below.

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_{firm} 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 involves money burning. The firm directly chooses 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 in which the safe firm ($\overline{\theta}$) chooses $\overline{m}$ and the addictive firm ($\underline{\theta}$) chooses $\underline{m}$. (The fact that the incentive constraints are satisfied follows from u_{firm} 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 of messages M and a cost function $y : M \to \mathbb{R}$. The negative result will follow from two equilibrium properties:

(i) The regulator's likelihood of *approval* is weakly higher 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*.

We show that there is no equilibrium in which the regulator is strictly more likely to *approve* after observing a message from the safe firm. It suffices to show that in any equilibrium where the firm of type $\overline{\theta}$ chooses message $\overline{m}$ with positive probability and the firm of type $\underline{\theta}$ chooses message $\underline{m}$ with positive probability, it must be that $\text{approve}(p(\underline{m})) \geq \text{approve}(p(\overline{m}))$. If this condition holds, then the message does not affect the likelihood of *approval* (by (i)), and thus no relevant information is transmitted (by (ii)).

Assume, by contradiction, that $\text{approve}(p(\overline{m})) > \text{approve}(p(\underline{m}))$. Since $c > 1$,

$$\begin{aligned} u_{\text{firm}}(\underline{\theta}, p(\overline{m})) - u_{\text{firm}}(\underline{\theta}, p(\underline{m})) &= c \cdot \left[\text{approve}(p(\overline{m})) - \text{approve}(p(\underline{m}))\right] \\ &> \text{approve}(p(\overline{m})) - \text{approve}(p(\underline{m})) \\ &= u_{\text{firm}}(\overline{\theta}, p(\overline{m})) - u_{\text{firm}}(\overline{\theta}, p(\underline{m})), \end{aligned}$$

capturing a form of strict submodularity of u_{firm}: firm ($\underline{\theta}$) has a strictly higher marginal benefit of approval and, hence, a strictly higher incentive to induce a high posterior $p(m)$. Since firm $\overline{\theta}$ chooses $\overline{m}$, $u_{\text{firm}}(\overline{\theta}, p(\overline{m})) + y(\overline{m}) \geq u_{\text{firm}}(\overline{\theta}, p(\underline{m})) + y(\underline{m})$. Hence, these two inequalities imply that

$$u_{\text{firm}}(\underline{\theta}, p(\overline{m})) + y(\overline{m}) > u_{\text{firm}}(\underline{\theta}, p(\underline{m})) + y(\underline{m}),$$

contradicting the fact that firm $\underline{\theta}$ chooses $\underline{m}$. Thus, $\text{approve}(p(\underline{m})) \geq \text{approve}(p(\overline{m}))$.

The impossibility 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, including those that garble the firm's behavior through noise and 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 only observe signals provided by the mechanism and the firm will influence the distribution of signals through its behavior. Notice, since $u_{\text{firm}}(\theta, p)$ is linear in $\text{approve}(p)$, the argument above still applies. The addictive firm is the one that has a (strictly) higher willingness to pay for high-posterior signals. Hence, it selects a signal distribution with a weakly higher expected value of $\text{approve}(p)$ than the distribution selected by the safe firm. However, under Bayesian updating, signals with a higher value of approve(p) are more likely to come from the safe firm (by (i)). As a consequence, the state does not impact the likelihood of *approval* and no information can be shared (by (ii)).

Remarkably, the feasibility of information sharing is discontinuous in the parameter c. Small changes in c lead to big changes in what information can be transmitted. This suggests that, when c is close to 1, a designer may want to use non-neutral mechanisms to lower c.

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. [11] 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 $G = \big((A_i, \pi_i) : i \in \{e, \ell\}\big)$ for that game. The game G is fixed throughout the analysis.

3.1 Neutral Mechanisms

A neutral mechanism is an extensive form that is played after the expert learns the state but before the agents play the game of interest G. These mechanisms allow agents to share 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 the game G. Since only neutral mechanisms are analyzed in this paper, they will be referred to simply as mechanisms.

A mechanism $\mathcal{M}$ is an extensive form played by the expert, the layman, and chance.[12] (Appendix A.2 provides a full description.) The key ingredients are the non-

[11]Section 9.2 extends the analysis to state spaces beyond $\mathbb{R}$.

[12]The definition of extensive form used here generalizes the one in Osborne and Rubinstein [1994] (Section 6.3.2), which allows for sequential and simultaneous-move games. By introducing informational partitions, the extended framework here captures imperfect information with perfect recall. (See Friedenberg and Rivera Mora [2025].)

terminal information sets, the terminal information sets, and the associated transfers. 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 correspond to the start of G. Each $i \in \{e, \ell, \mathrm{c}\}$ has an information partition on V, given by $\mathcal{I}_i \subseteq 2^V$. (The partition $\mathcal{I}_i$ covers both non-terminal and terminal nodes in the extensive form.) Each partition $\mathcal{I}_i$ satisfies three conditions. First, it satisfies no absentmindedness, i.e., if $I_i \in \mathcal{I}_i$, $\{v, v'\} \subseteq I_i$, and $v \neq v'$, then $v \not\succsim v'$ and $v' \not\succsim v$. Second, each information partition $\mathcal{I}_i$ is such that i has perfect recall. Third, the mechanism has an observable end, i.e., if $I_i \in \mathcal{I}_i$ and $I_i \cap Z \neq \emptyset$, 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 $\mathcal{T}_i \subseteq \mathcal{I}_i$. There is a transfer function $\gamma_i : \mathcal{T}_i \to \mathbb{R}$ that associates each terminal information set of $i \in \{e, \ell\}$ with a transfer that i receives. Observe, this implicitly assumes that i observes her transfer $\gamma_i(T_i)$.

The definition of a mechanism $\mathcal{M}$ captures the four independence properties. Notice, the set of nodes V does not contain information about the realization of the state. So, the information sets, the action sets, and the action correspondences do not depend on the realization of the state Θ. Hence, $\mathcal{M}$ satisfies structural independence. Second, the strategy of chance does not depend on the realization of Θ. Hence, $\mathcal{M}$ is statistically independent. Third, the transfer mapping does not depend on the realization of Θ. Hence, $\mathcal{M}$ is outcome independent. Fourth, the mechanism does not make reference to G and so cannot change G itself. Hence, $\mathcal{M}$ is game independent.

3.2 The Supergame

A 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: First, Nature chooses state θ, which is observed by the expert. Next, the agents play $\mathcal{M}$, after which each agent i observes a terminal information set $T_i \in \mathcal{T}_i$ and obtains a transfer $y_i = \gamma_i(T_i)$. Finally, the agents play G. The payoffs are quasilinear in the outcome of G and the transfer. So, the payoff for i from a state θ, an action profile a, and a transfer y_i is $\pi_i(\theta, a) + y_i$.

Each agent has the option to participate in the supergame $(\mathcal{M}, G)$ or select an exogenous outside option. The expert's outside option is a state-contingent mapping

$\underline{\pi}_e : \Theta \to \mathbb{R} \cup \{-\infty\}$. The layman's outside option is a value $\underline{\pi}_\ell \in \mathbb{R} \cup \{-\infty\}$.[13]

4 The Auxiliary Problem

This section analyzes an auxiliary mechanism design problem in which agents have belief-dependent preferences. Specifically, the analysis here omits explicit reference to the game G and uses belief-dependent utility functions as primitives that capture the agents' payoffs. Section 7 formalizes the connection between G and the belief-dependent utilities, showing how equilibria in this auxiliary problem characterize the equilibria in the original mechanism design problem.

4.1 The Psychological Game

Describing the auxiliary problem requires introducing the agents' hierarchies of beliefs about the state Θ. Since the expert knows the state, the expert's first-order belief h_e^1 is trivial; the layman's first-order belief h_ℓ^1 describes the probability that the layman assigns to the state Θ; the expert's second-order belief h_e^2 describes the probability that the expert assigns to the layman's first-order beliefs; the layman's second-order belief h_ℓ^2 describes the probability that the layman assigns to both, the state Θ and the expert's first-order beliefs; and so on for higher-order beliefs. Write H_i for i's set of collectively coherent hierarchies of beliefs and call $H := H_e \times H_\ell$ the **belief structure**. (See Appendix A.1 for the construction of H.) Write $h_i = (h_i^1, h_i^2, ...) \in H_i$ for a **hierarchy of beliefs** of agent i and call h_i^k the **kth-order belief** of i.

A **belief-dependent utility** for the expert is a measurable function $u_e : \Theta \times H_e \to \mathbb{R}$. Similarly, a **belief-dependent utility** for the layman is a measurable function $u_\ell : H_\ell \to \mathbb{R}$. Notice, since the layman does not observe the state, the layman's belief-dependent utility does not directly depend on Θ; instead it depends on his first-order beliefs h_ℓ^1 about Θ.

The mechanism $\mathcal{M}$ and belief-dependent utilities (u_e, u_ℓ) induce the **psychological game** $(\mathcal{M}, u_e, u_\ell)$. The timing is as follows: First, Nature chooses state $\theta \in \Theta$, which is observed by the expert. Next, the agents play $\mathcal{M}$, after which each agent i observes a terminal information set $T_i \in \mathcal{T}_i$ and obtains a transfer $y_i = \gamma_i(T_i)$.

[13] The value $-\infty$ indicates no viable outside option. While the layman's outside option could also depend on the state, the layman does not observe it. Hence, $\underline{\pi}_\ell$ is the expected value.

The payoffs of each agent i are quasilinear in the belief-dependent utility u_i and the transfer y_i. So, if the expert observes state θ, has hierarchy h_e, and transfer y_e, the expert's utility is $u_e(\theta, h_e) + y_e$. If the layman has hierarchy h_ℓ and transfer y_ℓ, the layman's utility is $u_\ell(h_\ell) + y_\ell$.

A **behavior strategy** for the expert is a mapping ρ_e from states Θ and information sets $\mathcal{I}_e \backslash \mathcal{T}_e$ to distributions of actions available at $\mathcal{I}_e \backslash \mathcal{T}_e$. Likewise, a **behavior strategy** for the layman is a mapping ρ_ℓ from information sets $\mathcal{I}_\ell \backslash \mathcal{T}_\ell$ to distributions of actions available at $\mathcal{I}_\ell \backslash \mathcal{T}_\ell$. (See Appendix A.2 for a formal description.)

4.2 Perfect Bayesian Equilibrium

This paper uses Perfect Bayesian Equilibrium as its solution concept, appropriately defined for psychological games. This solution concept builds on two ingredients: strategies and interim belief mappings. The interim belief mappings specify the endogenous beliefs that agents hold at each node of the mechanism. An **interim belief mapping** for the expert is a function $\beta_e : \Theta \times \mathcal{I}_e \to \Delta(V)$ such that $\beta_e(\theta, I_e)(I_e) = 1$. So, β_e specifies e's beliefs about which node has been reached. Likewise, an **interim belief mapping** for the layman is a function $\beta_\ell : \mathcal{I}_\ell \to \Delta(\Theta \times V)$ such that $\beta_\ell(I_\ell)(\Theta \times I_\ell) = 1$. So, β_ℓ specifies ℓ's beliefs about which state is realized and which node has been reached. Observe, interim belief mappings satisfy $\beta_e(\theta, T_e)(Z) = 1$ for each $(\theta, T_e) \in \Theta \times \mathcal{T}_e$ and that $\beta_\ell(T_\ell)(Z) = 1$ for each $T_\ell \in \mathcal{T}_\ell$. So, at each terminal node, both agents know that the mechanism has ended.

The pair of belief mappings $\beta = (\beta_e, \beta_\ell)$ induces a hierarchy $h_e \in H_e$ for each $(\theta, T_e) \in \Theta \times \mathcal{T}_e$ and a hierarchy $h_\ell \in H_\ell$ for each $T_\ell \in \mathcal{T}_\ell$. (See Appendix A.1 for their formalization.) Write $\delta_e : \Theta \times \mathcal{T}_e \to H_e$ and $\delta_\ell : \mathcal{T}_\ell \to H_\ell$ for the **hierarchy mappings** induced by β. So, if the state is θ and the expert observes T_e, the expert has hierarchy $\delta_e(\theta, T_e)$. If the layman observes T_ℓ, the layman has hierarchy $\delta_\ell(T_\ell)$.[14]

In principle, the interim belief mappings may not be consistent with the strategy profile played. The notion of perfect Bayesian equilibrium requires that they are. Write S_i for i's set of pure strategies in $\mathcal{M}$. For each $(\theta, v) \in \Theta \times V$, write $P(\theta, v \mid s_i, \rho_{-i})$ for the ex-ante probability that θ occurs and the path goes through v, provided that i plays the pure strategy s_i and $-i$ plays the behavior strategy ρ_{-i}. (See Appendix A.2 for the calculation.)

[14]Although δ_e and δ_ℓ depend on β, the reference to β is suppressed for notational convenience.

Definition 4.1. *The interim beliefs* (β_e, β_ℓ) *are* ***consistent*** *with* (ρ_e, ρ_ℓ) *if the following hold:*

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

$$\beta_e(\theta, I_e)(v) \sum_{v' \in I_e} P(\theta, v' \mid s_e, \rho_\ell) = P(\theta, v \mid s_e, \rho_\ell).$$

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

$$\beta_\ell(I_\ell)(\theta, v) \sum_{(\theta', v') \in \Theta \times I_\ell} P(\theta', v' \mid s_\ell, \rho_e) = P(\theta, v \mid s_\ell, \rho_e).$$

Consistency requires that e's (resp. ℓ's) beliefs are derived by the chain rule of conditional probability. This condition imposes the implicit requirement that interim beliefs satisfy *own-action independence:* The probability that ℓ assigns to each (θ, v) is independent of the pure strategy s_ℓ that is used (provided that ℓ uses a strategy s_ℓ that allows for I_ℓ). Likewise, the probability that e assigns to each v is independent of the pure strategy s_e that is used (provided that e uses a strategy s_e that allows for I_e). So, if β is consistent with (ρ_i, ρ_{-i}) and i deviates from ρ_i, i still believes that $-i$ is playing according to ρ_{-i}.

Write $P(T_e \mid \theta, I_e, \rho, \beta_e)$ for the probability that e assigns to T_e given that the state is θ, the information set I_e is reached, ρ is played, and e has interim beliefs β_e. Likewise, write $P(T_\ell \mid I_\ell, \rho, \beta_\ell)$ for the probability that ℓ assigns to T_ℓ given that I_ℓ is reached, ρ is played, and ℓ has interim beliefs β_ℓ. (See Appendix A.2 for the calculation.) So, if the interim beliefs $\beta = (\beta_e, \beta_\ell)$ are associated with the hierarchy mappings (δ_e, δ_ℓ), then the expert's interim expected payoffs at (θ, I_e) are

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

and the layman's interim expected payoffs at I_ℓ are

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

Thus, interim beliefs impact both the probability of reaching each terminal node and the belief-dependent utility that agents get at each terminal node.

Definition 4.2. *An assessment* (ρ, β) *satisfies* ***sequential rationality*** *if the following hold:*

(i) For each $(\theta, I_e) \in \Theta \times \mathcal{I}_e$ *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 $I_\ell \in \mathcal{I}_\ell$ *and* ρ'_ℓ, $\mathcal{U}_\ell(\rho_e, \rho_\ell \mid I_\ell, \beta) \geq \mathcal{U}_\ell(\rho_e, \rho'_\ell \mid I_\ell, \beta)$.

The assessment (ρ, β) satisfies sequential rationality if the behavior strategy ρ_i maximizes i's interim expected payoffs at each information set, provided that the agents face interim beliefs β.

Definition 4.3. *Call* (ρ, β) *a* ***perfect Bayesian equilibrium (PBE)*** *of* $(\mathcal{M}, u_e, u_\ell)$ *if* (ρ, β) *satisfies sequential rationality and belief mappings* β *are consistent with* ρ.

We will focus on perfect Bayesian equilibria that incentivize agents to participate in $\mathcal{M}$. A PBE (ρ, β) is **individually rational** if (1) for each state $\theta \in \Theta$, $\mathcal{U}_e(\rho \mid \theta, \{\emptyset\}, \beta) \geq \underline{\pi}_e(\theta)$, and (2) $\mathcal{U}_\ell(\rho \mid \{\emptyset\}, \beta) \geq \underline{\pi}_\ell$. So, at the initial information set $\{\emptyset\}$ (i.e., after the expert learns the state but before $\mathcal{M}$ is played) the expected utility that i gets from (ρ, β) must be higher than i's outside option. In this sense, individual rationality requires that each agent has incentives to participate.

5 Fully-Revealing Preferences

This section establishes the first central result, offering a sufficient condition for belief-dependent preferences to allow full revelation of the state. A pair of utility functions (u_e, u_ℓ) is fully revealing if there exists some mechanism $\mathcal{M}$ that induces the expert to fully reveal the state to the layman.

To formally describe fully-revealing preferences, we introduce some notation. Following [Brandenburger and Dekel, 1993], there exists a canonical homeomorphism between H_e and $\Delta(H_\ell)$ that identifies each hierarchy $h_e \in H_e$ with a unique probability measure $h_e^\infty \in \Delta(H_\ell)$. The measure h_e^∞ captures the expert's beliefs regarding the layman's hierarchy of beliefs. Similarly, there exists a canonical homeomorphism between H_ℓ and $\Delta(\Theta \times H_e)$ that identifies each hierarchy $h_\ell \in H_\ell$ with a unique probability measure $h_\ell^\infty \in \Delta(\Theta \times H_e)$. The measure h_ℓ^∞ captures the layman's beliefs about both the state and the expert's hierarchies of beliefs. (See Appendix A.1.) Denote h_i^∞ as the **extension** of h_i.

For each $\theta \in \Theta$, there is a unique hierarchy profile (h_e, h_ℓ) that satisfies $h_e^\infty(h_\ell) = h_\ell^\infty(\theta, h_e) = 1$. At such profile, there is common belief that "the layman believes that the state is θ." Say $h_i \in H_i$ has **common degenerate belief for** θ if there is some $h_{-i} \in H_{-i}$ such that (h_i, h_{-i}) satisfies these equalities. Write $\text{CDB}_i \subsetneq H_i$ for the set of common degenerate beliefs of i for some state in Θ. If $h_i \in \text{CDB}_i$ (resp. $h'_i \in \text{CDB}_i$) has common degenerate belief for θ (resp. θ'), write $h_i \succsim h'_i$ if and only if $\theta \geq \theta'$.

Definition 5.1. *A PBE (ρ, β) of $(\mathcal{M}, u_e, u_\ell)$ is* ***fully revealing*** *if, for each $(T_e, T_\ell) \in \mathcal{T}_e \times \mathcal{T}_\ell$ with $T_e \cap T_\ell \neq \emptyset$, there is some $\theta \in \Theta$ so that $(\delta_e(\theta, T_e), \delta_\ell(T_\ell))$ have common degenerate belief for θ.*

Call (u_e, u_ℓ) **fully revealing** if there is some mechanism $\mathcal{M}$ and some individually rational PBE (ρ, β) of $(\mathcal{M}, u_e, u_\ell)$ that is fully revealing. That is, if it is feasible to construct a mechanism and a PBE thereof in which the layman fully learns the state.

Definition 5.2. *Say that u_e is* ***supermodular on common degenerate beliefs*** *if, for each $\theta, \theta' \in \Theta$ with $\theta \geq \theta'$ and each $h_e, h'_e \in \text{CDB}_e$ with $h_e \succsim h'_e$,*

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

So, u_e is supermodular on common degenerate beliefs if the expert has a weakly higher incentive to induce "higher" degenerate hierarchies when the true state is high versus when it is low. The following result shows that this condition is sufficient for fully revealing the state.

Theorem 5.1. *Fix belief-based utilities (u_e, u_ℓ). If u_e is supermodular on common degenerate beliefs, then (u_e, u_ℓ) is fully revealing.*

The proof constructs a simple mechanism in which the expert directly reveals the state and receives a transfer that depends on their report.[15] Observe, however, that belief-dependent preferences introduce a subtlety not present in similar classical results: although the mechanism specifies transfers for each report, it cannot directly determine the agents' endogenous beliefs that arise from each report. Instead, agents form beliefs based on the strategy they expect the expert to follow. The key is that, under the expert's truthful reporting strategy, each report induces a hierarchy profile with common degenerate belief in the reported state. Moreover, this property

[15]Theorem 5.1 provides only a sufficient condition for perfect revelation. Section 9.2 expands it by introducing a cyclical monotonicity condition that is necessary and sufficient for full revelation of the state. However, the supermodularity condition is often easier to verify.

holds even when the expert deviates from truthful reporting.[16] Given this property, supermodularity ensures that higher states provide the expert with stronger incentives to induce higher degenerate hierarchies and, consequently, to make higher reports.

Example 5.1. *Consider the example of Section 2. Observe, if $c \leq 1$, then the belief-dependent utility u_{firm} satisfies increasing differences. This implies that u_{firm} is supermodular on common degenerate beliefs. Therefore, $(u_{\text{firm}}, u_{\text{reg}})$ is fully revealing.*

6 Concealing Preferences

This section states the second central result, providing a sufficient condition for belief-dependent preferences to preclude the transmission of payoff-relevant information.

It will be convenient to introduce some notation. Fix a mechanism $\mathcal{M}$ with a set of terminal nodes Z and terminal information sets $\mathcal{T}_i$ for each agent i. Let (ρ, β) be a consistent assessment of $\mathcal{M}$. The strategy profile ρ and the prior μ induce a probability distribution $\mathbb{P} \in \Delta(\Theta \times Z)$. (See Appendix A.2). Then $\mathcal{P} = (\Theta \times Z, \mathbb{P})$ is the **terminal probability space** induced by $\mathcal{M}$ and (ρ, β).[17]

The analysis will rely on random variables $\mathbf{X} : \Theta \times Z \to \mathbb{R}$ on $\mathcal{P}$. Since $\Theta \times Z$ is finite, each random variable has finite moments. There will be three random variables of interest, specifying transfers, states, and hierarchies as a function of the state and terminal node. First, let $\mathbf{Y}_i : \Theta \times Z \to \mathbb{R}$ be such that $\mathbf{Y}_i(\theta, z) = \gamma_i(T_i[z])$, where $T_i[z]$ is the unique $T_i \in \mathcal{T}_i$ such that $z \in T_i$. Second, let $\mathbf{\Theta}$ be the projection of $\Theta \times Z$ onto Θ. Third, let $\mathbf{H}_i : \Theta \times Z \to H_i$ be such that $\mathbf{H}_e(\theta, z) = \delta_e(\theta, T_e[z])$ and $\mathbf{H}_\ell(\theta, z) = \delta_\ell(T_\ell[z])$.

Absent information sharing, the agents' hierarchies are described by a profile of hierarchies $\left(\tilde{h}_e, \tilde{h}_\ell\right) \in H_e \times H_\ell$ that is induced by the prior μ. (See Appendix A.1.) We use $\left(\tilde{h}_e, \tilde{h}_\ell\right)$ as a benchmark to determine which equilibria impact the agents' utility functions.

Definition 6.1. *Fix a psychological game $(\mathcal{M}, u_e, u_\ell)$, and a PBE (ρ, β) thereof. Let*

[16] This property implies that Theorem 5.1 holds regardless of the expert's utility on non-degenerate hierarchies. This is because the equilibrium construction ensures that each report—whether truthful or not—induces common degenerate hierarchies for each reported state.

[17] The space $\Theta \times Z$ is endowed with the discrete σ-algebra. Random variables are written in bold.

$\mathcal{P}$ be the terminal probability space they induce. Say $\mathcal{P}$ ***impacts*** *(u_e, u_ℓ) if*

$$\mathbb{P}\big[u_e(\mathbf{\Theta}, \mathbf{H}_e) = u_e\big(\mathbf{\Theta}, \tilde{h}_e\big) \ \textit{ and } \ u_\ell(\mathbf{H}_\ell) = u_\ell\big(\tilde{h}_\ell\big)\big] < 1.$$

A pair of utility functions (u_e, u_ℓ) is **concealing** if for each mechanism $\mathcal{M}$ and each PBE (ρ, β) of $(\mathcal{M}, u_e, u_\ell)$, the induced probability space $\mathcal{P}$ does not impact (u_e, u_ℓ). In other words, (u_e, u_ℓ) is concealing if no mechanism can transmit payoff-relevant information. Notice, this definition does not rule out the possibility that the layman learns information about Θ while leaving ex-post payoffs unchanged.[18]

We now introduce a sufficient condition for concealing preferences. Call a probability space $\mathcal{P} = (\Theta \times Z, \mathbb{P})$ **feasible** if there is a mechanism $\mathcal{M}$ and a consistent assessment (ρ, β) thereof that induces $\mathcal{P}$.

Definition 6.2. *Fix belief-dependent utility functions (u_e, u_ℓ). Say (u_e, u_ℓ) is* ***statistically submodular*** *if for each feasible terminal probability space $\mathcal{P}$ that impacts (u_e, u_ℓ), there are some states $\overline{\theta} > \underline{\theta}$ such that*

$$\mathbb{E}\big[u_e(\overline{\theta}, \mathbf{H}_e) - u_e(\underline{\theta}, \mathbf{H}_e) \mid \mathbf{\Theta} = \overline{\theta}\big] < \mathbb{E}\big[u_e\big(\overline{\theta}, \mathbf{H}_e\big) - u_e\big(\underline{\theta}, \mathbf{H}_e\big) \mid \mathbf{\Theta} = \underline{\theta}\big]. \qquad (1)$$

The belief-dependent utilities (u_e, u_ℓ) are statistically submodular if whenever there is an equilibrium that shares payoff-relevant information, there are states $\overline{\theta} > \underline{\theta}$ such that the following hold: an expert's marginal benefit from inducing the hierarchy associated with $\overline{\theta}$ over the hierarchy associated with $\underline{\theta}$ is higher for $\underline{\theta}$ over $\overline{\theta}$.

Lemma 6.1. *Fix a psychological game $(\mathcal{M}, u_e, u_\ell)$, a consistent assessment (ρ, β), and let $\mathcal{P}$ be the terminal probability space they induce. If (ρ, β) is a PBE of $(\mathcal{M}, u_e, u_\ell)$, then for each $\theta, \theta' \in \Theta$,*

$$\mathbb{E}\big[u_e(\theta, \mathbf{H}_e) + \boldsymbol{Y}_e \mid \mathbf{\Theta} = \theta\big] \geq \mathbb{E}\big[u_e(\theta, \mathbf{H}_e) + \boldsymbol{Y}_e \mid \mathbf{\Theta} = \theta'\big]. \qquad (2)$$

Lemma 6.1 identifies the restrictions that PBE imposes on the conditional distributions of the expert's hierarchies and transfers. The proof uses the revelation

[18] Section 2 argues that if $c > 1$, then $(u_{\text{firm}}, u_{\text{reg}})$ is concealing. However, in equilibrium, the regulator can learn information: consider a prior probability of $\frac{1}{4}$ that the firm is safe and cheap talk mechanism in which firm $\overline{\theta}$ sends $\overline{m}$ with probability $\frac{3}{4}$ and firm $\underline{\theta}$ sends $\overline{m}$ with probability $\frac{5}{12}$. The regulator's posteriors are $p(\overline{m}) = \frac{3}{8}$ and $p(\underline{m}) = \frac{1}{8}$, both below $\frac{1}{2}$ and thus do not affect approval. Hence, this Bayesian equilibrium impacts the beliefs without affecting the payoffs.

principle for settings with belief-dependent preferences to capture the incentive compatible constraints associated with (ρ, β). (See Rivera Mora [2024].) Equation (2) requires that the expected payoff of an expert of type θ following ρ (the left term) must be weakly higher than her payoff from mimicking a type θ' (the right term).

Theorem 6.1. *If (u_e, u_ℓ) is statistically submodular, then (u_e, u_ℓ) is concealing.*

Proof. Fix a mechanism $\mathcal{M}$ and a PBE (ρ, β) of $(\mathcal{M}, u_e, u_\ell)$. Write $\mathcal{P}$ for the probability space induced by $\mathcal{M}$ and (ρ, β). To show that (u_e, u_ℓ) is concealing, it suffices to show that $\mathcal{P}$ does not impact (u_e, u_ℓ).

We proceed by contradiction. Suppose that $\mathcal{P}$ impacts (u_e, u_ℓ). So, by statistical submodularity, there are some states $\overline{\theta} > \underline{\theta}$ that satisfy Equation (1). In addition, Lemma 6.1 states that

$$\mathbb{E}\left[u_e(\overline{\theta}, \mathbf{H}_e) + \mathbf{Y}_e \mid \mathbf{\Theta} = \overline{\theta}\right] - \mathbb{E}\left[u_e(\overline{\theta}, \mathbf{H}_e) + \mathbf{Y}_e \mid \mathbf{\Theta} = \underline{\theta}\right] \geq 0, \quad \text{and}$$
$$\mathbb{E}\left[u_e(\underline{\theta}, \mathbf{H}_e) + \mathbf{Y}_e \mid \mathbf{\Theta} = \underline{\theta}\right] - \mathbb{E}\left[u_e(\underline{\theta}, \mathbf{H}_e) + \mathbf{Y}_e \mid \mathbf{\Theta} = \overline{\theta}\right] \geq 0.$$

Adding these inequalities implies that

$$\mathbb{E}\left[u_e(\overline{\theta}, \mathbf{H}_e) - u_e(\underline{\theta}, \mathbf{H}_e) \mid \mathbf{\Theta} = \overline{\theta}\right] - \mathbb{E}\left[u_e(\overline{\theta}, \mathbf{H}_e) - u_e(\underline{\theta}, \mathbf{H}_e) \mid \mathbf{\Theta} = \underline{\theta}\right] \geq 0,$$

which violates Equation (1). Therefore, $\mathcal{P}$ does not impact (u_e, u_ℓ), as desired. ∎

Intuitively, statistical submodularity states that some high-state expert has stronger incentives to appear as a low-state expert. However, incentive compatibility requires the opposite—the transfer scheme must induce experts to act as their true types rather than mimic other types. Therefore, statistical submodularity makes it impossible to share any payoff-relevant information.

6.1 Acute Statistics

This subsection shows how projecting the belief structure H onto a single dimension offers a simple method for identifying statistical submodularity. A **statistic** is a measurable mapping $f : H \to \mathbb{R}$ that summarizes the information revealed by the mechanism (captured by the profile of hierarchies of beliefs) into a single real number.

Fix a mechanism $\mathcal{M}$, a consistent assessment (ρ, β), and a statistic f. Let $\mathcal{P} = (\Theta \times Z, \mathbb{P})$ be the terminal probability space induced by $\mathcal{M}$ and (ρ, β). For each

$(\theta, z) \in \Theta \times Z$, write $\mathbf{F}(\theta, z) := f\big(\mathbf{H}_e(\theta, z), \mathbf{H}_\ell(\theta, z)\big)$. So, $\mathbf{F} : (\Theta \times Z) \to \mathbb{R}$ is the random variable in $\mathcal{P}$ that captures the distribution of the statistic f.

Definition 6.3. *The statistic $f : H \to \mathbb{R}$ is **acute** if, for each feasible terminal probability space $\mathcal{P}$, the associated random variable $\mathbf{F}$ satisfies either $\mathbb{P}\big[\mathbf{F} = f\big(\tilde{h}_e, \tilde{h}_\ell\big)\big] = 1$ or $\mathrm{Cov}[\mathbf{F}, \boldsymbol{\Theta}] > 0$.*

Recall that $\tilde{h}_e$ and $\tilde{h}_\ell$ are the hierarchies induced by the prior. So, a statistic f is acute if either (i) the information transmitted has no impact on f, or (ii) high values of f signal high values of θ. There are a plethora of acute statistics, including monotone transformations of the layman's expectation of the state, the agents' higher-order expectations of the state, and positive linear transformations of them. (See Lemmata B.7 - B.10.) The next example illustrates one important acute statistic:

Example 6.1. *Recall h^1_ℓ denotes the first-order beliefs of ℓ and write $f^1_\ell(h_e, h_\ell) = \sum_{\theta \in \Theta} \theta \cdot h^1_\ell(\theta)$. The statistic $f^1_\ell : H \to \mathbb{R}$ captures the layman's first-order expectation of the state given a profile $(h_e, h_\ell) \in H$.*[19] *Lemma B.7 shows that f^1_ℓ is acute. So, in each feasible probability space $\mathcal{P}$, either the layman's expectation of the state remains constant (i.e., the mechanism provides no relevant information for ℓ to update his conditional expectation) or it positively correlates with the state.*

Write $\mathbf{F}^1_\ell$ for the random variable associated with f^1_ℓ. Intuitively, a higher conditional expectation signals a higher state, so $\mathrm{Cov}[\mathbf{F}^1_\ell, \boldsymbol{\Theta}]$ should be non-negative. Geometrically, this follows from the fact that the conditional expectation $\mathbf{F}^1_\ell$ is a projection of $\boldsymbol{\Theta}$ into some subspace. (See Durrett [2019].) So, either $\mathbf{F}^1_\ell$ is a constant or the "angle" between $\boldsymbol{\Theta}$ and $\mathbf{F}^1_\ell$ is acute.[20]

Acute statistics are useful because they identify the Bayesian restrictions of how information flows between the agents. If agents are Bayesian, then no mechanism can "deceive," in the sense of inducing (on average) low values of f for high values of θ. This provides a method to verify statistical submodularity. To formalize this, fix an acute statistic f and a belief-dependent utility (u_e, u_ℓ). Say f is **essential** for (u_e, u_ℓ) if for each feasible probability space $\mathcal{P}$ either $\mathbb{P}\big[\mathbf{F} = f\big(\tilde{h}_e, \tilde{h}_\ell\big)\big] < 1$ or $\mathcal{P}$ does not

[19] Recall that statistics are defined here as mappings of both h_e and h_ℓ.

[20] Acute statistics have a geometric interpretation in terms of the angle induced by $\boldsymbol{\Theta}$ and $\mathbf{F}$. Write $L^2(\mathcal{P})$ for the quotient normed space of $\mathcal{P}$ in which two random variables $\mathbf{X}$ and $\mathbf{X}'$ are identified if there exists a constant c such that $\mathbf{X} - \mathbf{X}' = c$ almost surely. Since $\mathrm{Cov}[\cdot, \cdot]$ is an inner product, the angle between $\mathbf{X}$ and $\mathbf{X}'$ depends on the sign and relative magnitude of $\mathrm{Cov}[\mathbf{X}, \mathbf{X}']$. So, if f is acute, then either $\mathbf{F}$ is constant or the angle between $\boldsymbol{\Theta}$ and $\mathbf{F}$ is acute (less than 90 degrees).

impact (u_e, u_ℓ). So, f is essential if changing the agents' payoffs requires "moving" $\mathbf{F}$ with positive probability.

Lemma 6.2. *Fix a pair belief-dependent utilities (u_e, u_ℓ) and let f be an acute statistic that is essential for (u_e, u_ℓ). If, for each pair of states $\overline{\theta} > \underline{\theta}$ there are $c_1 \in \mathbb{R}$ and $c_2 < 0$ such that*

$$u_e(\overline{\theta}, h_e) - u_e(\underline{\theta}, h_e) = c_1 + c_2 \int_{H_\ell} f(h_e, h_\ell)\ dh_e^\infty,$$

then (u_e, u_ℓ) is statistically submodular.

Lemma 6.2 provides a simple condition to verify that a pair of belief-dependent utilities (u_e, u_ℓ) satisfy statistical submodularity. This condition requires that (1) f is essential for their payoffs, and (2) the difference $u_e\big(\overline{\theta}, \cdot\big) - u_e\big(\underline{\theta}, \cdot\big)$ decreases with respect to e's expectation of f. These conditions mirror decreasing differences in a stochastic setting: Higher values of the state $\mathbf{\Theta}$ (and thus higher values of $\mathbf{F}$) correspond to lower values of the difference $u_e\big(\overline{\theta}, \mathbf{H}_e\big) - u_e\big(\underline{\theta}, \mathbf{H}_e\big)$.

Example 6.2. *Consider the example of Section 2 with a parameter $c > 1$. The statistic $f(h_{\text{firm}}, h_{\text{reg}}) = \text{approve}\big(h^1_{\text{reg}}(\overline{\theta})\big)$ is acute (Lemma B.9) and captures the likelihood of approval. So, f is essential for $(u_{\text{firm}}, u_{\text{reg}})$. In addition, $u_{\text{firm}}(\cdot, \cdot)$ satisfies Equation (10) for $c_1 = 0$ and $c_2 = 1 - c < 0$. Thus, $(u_{\text{firm}}, u_{\text{reg}})$ is statistically submodular and concealing. (See Lemma 6.2 and Theorem 6.1.)*

7 The Reduced-Form Approach

This section describes the reduced-form approach, connecting Theorems 5.1 and 6.1 with the game G. The analysis identifies two classes of games: those that allow full revelation of the state and those that preclude sharing information.

7.1 Perfect Bayesian Equilibrium of Supergames

Fix a mechanism $\mathcal{M}$. The agents' optimization in $(\mathcal{M}, G)$ is based on their payoffs at each information set $I_i \in \mathcal{I}_i$. A **behavior strategy** for e is a mapping σ_e from states Θ and information sets $\mathcal{I}_e$ to distributions of actions available at $\mathcal{I}_e$. Likewise, a **behavior strategy** for ℓ is a mapping σ_ℓ from information sets $\mathcal{I}_\ell$ to distributions of actions available at $\mathcal{I}_\ell$. Each strategy profile (σ_e, σ_ℓ) specifies the agents' actions in G at each terminal node of $\mathcal{M}$. So, if $T_e \in \mathcal{T}_e$ and $T_\ell \in \mathcal{T}_\ell$, then $\sigma_e(\theta, T_e) \in \Delta(A_e)$ and $\sigma_\ell(T_\ell) \in \Delta(A_\ell)$. (See Appendix A.2 for a full description of the strategies.)

Fix a profile $\sigma = (\sigma_e, \sigma_\ell)$ and write $\mathcal{T} := \{(T_e, T_\ell) \in \mathcal{T}_e \times \mathcal{T}_\ell : T_e \cap T_\ell \neq \emptyset\}$ for the feasible pairs of terminal information sets. The expected payoff for i in G at state θ and profile (T_e, T_ℓ) is

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

Write $P(T_e, T_\ell \mid \theta, I_e, \sigma, \beta_e)$ for the probability that e assigns to (T_e, T_ℓ) being reached given that the state is θ, information set I_e is reached, σ is played, and e has interim beliefs β_e. Likewise, write $P(\theta, T_e, T_\ell \mid I_\ell, \sigma, \beta_\ell)$ for the probability that ℓ assigns to (θ, T_e, T_ℓ) given that I_ℓ is reached, σ is played, and ℓ has interim beliefs β_ℓ. (See Appendix A.2.) The agents' interim payoffs at (θ, I_e) and I_ℓ are

$$U_e(\sigma \mid \theta, I_e, \beta_e) := \sum_{(T_e, T_\ell) \in \mathcal{T}} [\Pi_e(\sigma \mid \theta, T_e, T_\ell) + \gamma_e(T_e)]\, P(T_e, T_\ell \mid \theta, I_e, \sigma, \beta_e), \text{ and}$$
$$U_\ell(\sigma \mid I_\ell, \beta_\ell) := \sum_{\theta \in \Theta} \sum_{(T_e, T_\ell) \in \mathcal{T}} [\Pi_\ell(\sigma \mid \theta, T_e, T_\ell) + \gamma_\ell(T_\ell)]\, P(\theta, T_e, T_\ell \mid I_\ell, \sigma, \beta_\ell).$$

So, the interim payoffs include both, the transfers from the mechanism $\mathcal{M}$ and the expected payoffs from G. (Observe, while $\mathcal{U}_i$ denotes i's interim payoffs of the psychological game, U_i denotes i's interim payoffs of the supergame.)

Definition 7.1. *An assessment $(\sigma_e, \sigma_\ell, \beta_e, \beta_\ell)$ satisfies **sequential rationality** if:*

(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, $U_\ell(\sigma_e, \sigma_\ell \mid I_\ell, \beta_\ell) \geq U_\ell(\sigma_e, \sigma'_\ell \mid I_\ell, \beta_\ell)$.

Observe, in contrast to the psychological game, here sequential rationality requires optimization at both the terminal and the non-terminal information sets of $\mathcal{M}$.

Definition 7.2. *The interim beliefs (β_e, β_ℓ) are **consistent** with (σ_e, σ_ℓ) if:*

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

$$\beta_e(\theta, I_e)(v) \sum_{v' \in I_e} P(\theta, v' \mid s_e, \sigma_\ell) = P(\theta, v \mid s_e, \sigma_\ell).$$

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

$$\beta_\ell(I_\ell)(\theta, v) \sum_{(\theta', v') \in \Theta \times I_\ell} P(\theta', v' \mid s_\ell, \sigma_e) = P(\theta, v \mid s_\ell, \sigma_e).$$

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 perfect Bayesian equilibrium (σ, β) is **individually rational** if (1) for each state $\theta \in \Theta$, $U_e(\sigma \mid \theta, \{\emptyset\}, \beta_e) \geq \underline{\pi}_e(\theta)$, and (2) $U_\ell(\sigma \mid \{\emptyset\}, \beta_\ell) \geq \underline{\pi}_\ell$. So, in the same way as Section 4, individual rationality requires that each agent has incentives to participate in $\mathcal{M}$.

7.2 The Induced Bayesian Game

The mechanism $\mathcal{M}$ and behavior thereof induce an information structure for the game G, thereby inducing a Bayesian game. To describe the induced Bayesian game, it will be useful to point out two observations. First, for the purposes of analyzing the induced Bayesian game, the only relevant aspect of the mechanism is the description of its terminal information sets: they impact the agents' information and so beliefs of the associated Bayesian game. With this in mind, 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, belief mappings are only important if they can arise from the agents' updating. With this in mind, write cons($\mathcal{M}$) for the set of interim belief mappings in $\mathcal{M}$ that are consistent with some strategy profile.

Fix a supergame $(\mathcal{M}, G)$ with $\mathcal{M} = (\cdot, \mathcal{T}_e, \mathcal{T}_\ell)$ and a profile of interim belief mappings $\beta \in \text{cons}(\mathcal{M})$. After the agents interact in $\mathcal{M}$, each agent i learns information associated with a terminal information set of $\mathcal{M}$. 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 $\beta_e(\theta, T_e)(Z) = \beta_\ell(T_\ell)(\Theta \ \times Z) = 1$. Hence, 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 $\beta_\ell(T_\ell)(\{\theta\} \times T_e)$. Write BG$(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$ for the **induced Bayesian game**.

Within $\text{BG}(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$, 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 $\hat{\sigma}_\ell : \mathcal{T}_\ell \to \Delta(A_\ell)$, and The agent's expected payoff of profile $\hat{\sigma} = (\hat{\sigma}_e, \hat{\sigma}_\ell)$ given (θ, T_e, β_e) and (T_ℓ, β_ℓ) are

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

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

Definition 7.3. *Call the profile $\hat{\sigma} = (\hat{\sigma}_e, \hat{\sigma}_\ell)$ a **Bayesian equilibrium** of $\text{BG}(\mathcal{T}_e, \mathcal{T}_\ell, \beta_e, \beta_\ell)$ 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 $\hat{\sigma}'_\ell$, $\Pi_\ell(\hat{\sigma}_e, \hat{\sigma}_\ell \mid T_\ell, \beta_\ell) \geq \Pi_\ell(\hat{\sigma}_e, \hat{\sigma}'_\ell \mid T_\ell, \beta_\ell)$.

7.3 Reduced Forms

Reduced forms formalize the idea of using a pair of belief-dependent utility functions (u_e, u_ℓ) to capture the equilibrium payoffs in the induced Bayesian games.

Definition 7.4. *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 $(\theta, T_e) \in \Theta \times \mathcal{T}_e$, $u_e(\theta, \delta_e(\theta, T_e)) = \Pi_e(\hat{\sigma} \mid \theta, T_e, \beta_e)$.

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

The pair (u_e, u_ℓ) is a reduced form for G if it captures equilibrium payoffs of each induced Bayesian game by making reference to the agents' hierarchies of beliefs. So, in this sense (u_e, u_ℓ) captures the preferences for information that are induced by G. Note, each PBE of $(\mathcal{M}, G)$ has two essential outputs: the transfers exchanged in the mechanism and the payoffs from the induced Bayesian game. Definition 7.4 refers to the latter but not the former. With this in mind, write $Y_e(\theta, \mathcal{M}, \sigma)$ (resp. $Y_e(\theta, \mathcal{M}, \rho)$) for the expert's expected transfer under σ (resp. ρ) conditional on observing θ. Likewise, write $Y_\ell(\mathcal{M}, \sigma)$ (resp. $Y_\ell(\mathcal{M}, \rho)$) for the layman's expected transfer under σ (resp. ρ). (See Appendix A.2.)

Definition 7.5. *A PBE (σ, β) of a supergame $(\mathcal{M}, G)$ and a PBE (ρ, β') of a psychological game $(\mathcal{M}', u_e, u_\ell)$ are **equivalent** if the following hold:*

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

(ii) $U_\ell(\sigma \mid \{\emptyset\}, \beta_\ell) = \mathcal{U}_\ell(\rho \mid \{\emptyset\}, \beta')$, $Y_\ell(\mathcal{M}, \sigma) = Y_\ell(\mathcal{M}', \rho)$.

So, under equivalence, the agents' payoffs and the transfers used in the mechanism are equal in the supergame and in the psychological game.[21]

Lemma 7.1. *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* $(\mathcal{M}, G)$ *which is equivalent to* (ρ, β)*.*

Lemma 7.1 shows that each equilibrium of an associated psychological game $(\mathcal{M}, u_e, u_\ell)$ induces an equivalent equilibrium of the supergame $(\mathcal{M}, G)$. So, for a given reduced form, the set of equilibria across all associated psychological games captures a subset of equilibria across all supergames. However, the result is silent about whether a single reduced form captures all equilibria of all supergames. If each induced Bayesian game has a unique equilibrium, then a reduced form captures all equilibria. (See Sections 8.1 and 8.2.) If some induced Bayesian games have multiple equilibria, then a single reduced form may not capture all equilibria. (See Section 2.)

Definition 7.6. *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 characterizes each equilibrium of each supergame as an equilibrium of some psychological game. So, the PBE of the class of supergames $\big\{(\mathcal{M}, G) : \mathcal{M} \text{ is a mechanism}\big\}$ are equivalent to the PBE of the class of psychological games $\big\{(\mathcal{M}, u_e, u_\ell) : \mathcal{M} \text{ is a mechanism}, (u_e, u_\ell) \in \mathrm{RF}\big\}$. Intuitively, different reduced forms in RF capture differences in behavior for fixed hierarchies of beliefs. So, by considering each $(u_e, u_\ell) \in \mathrm{RF}$, the class of psychological games effectively captures all equilibria. Section 9.1 discusses the existence of reduced-form representations.

7.4 Main Result

Theorems 5.1 and 6.1 identify which belief-dependent preferences are fully-revealing or concealing. This section builds on these results to identify which games are fully-revealing or concealing.

[21] Notice, equivalence is allowed even if the mechanisms $\mathcal{M}$ and $\mathcal{M}'$ are different.

First we define fully-revealing games. Fix a mechanism $\mathcal{M} = (\cdot, \mathcal{T}_e, \mathcal{T}_\ell)$. A PBE (σ, β) of a supergame $(G, \mathcal{M})$ is fully revealing if, for each $(T_e, T_\ell) \in \mathcal{T}_e \times \mathcal{T}_\ell$ with $T_e \cap T_\ell \neq \emptyset$, there is some $\theta \in \Theta$ such that $h_e = \delta_e(\theta, T_e)$ and $h_\ell = \delta_\ell(T_\ell)$ have common degenerate belief for θ. A game G is **fully revealing** if there is a mechanism $\mathcal{M}$ and individually rational PBE (σ, β) of $(\mathcal{M}, G)$ that is fully revealing.

To define concealing games, it is useful to consider the benchmark in which the expert and the layman do not interact before playing G. So, the expert observes the state and, subsequently, each agent i selects actions from A_i. (No transfers are sent or received and the layman does not observe the state.) In this silent Bayesian game, the expert's strategy is $\sigma_e^s : \Theta \to \Delta(A_e)$ and the layman's strategy is $\sigma_\ell^s \in \Delta(A_\ell)$. Call the pair $\Pi_e^s : \Theta \to \mathbb{R}$ and $\Pi_\ell^s \in \mathbb{R}$ **silent payoffs** if there is some Bayesian equilibrium σ^s of the silent game whose payoffs are given by Π_e^s and Π_ℓ^s.[22] A game G is **concealing** if, for each mechanism $\mathcal{M}$ and each individually rational PBE $(\sigma, \beta_e, \beta_\ell)$ of the supergame $(\mathcal{M}, G)$, there are silent payoffs (Π_e^s, Π_ℓ^s) so that

$$U_e(\sigma \mid \theta, \{\emptyset\}, \beta_e) = \Pi_e^s(\theta) + Y_e(\theta, \mathcal{M}, \sigma) \quad \text{and} \quad U_\ell(\sigma \mid \{\emptyset\}, \beta_\ell) = \Pi_\ell^s + Y_\ell(\mathcal{M}, \sigma).$$

So, a game G is concealing if any tuple of expected utilities and expected transfers that can be achieved by some perfect Bayesian equilibrium can also be achieved with a mechanism that does not reveal any information. This means that in a concealing game, the designer can only influence payoffs through monetary transfers, not through revealing information.

Theorem 7.1.

(i) *Let (u_e, u_ℓ) be a reduced form of G. If u_e is supermodular on common degenerate beliefs, then G is fully-revealing.*

(ii) *Let* RF *be a reduced-form representation of G. If each $(u_e, u_\ell) \in \text{RF}$ is statistically submodular, then G is concealing.*

8 Applications

This section presents examples of economically relevant games and characterizes when such games are fully-revealing or concealing. The first considers an environment in

[22] Formally, Π_e^s and Π_ℓ^s are silent payoffs if they describe the expected payoffs of Bayesian equilibrium of the Bayesian game induced by a mechanism that has a single terminal node and no transfers.

which only the layman takes an action, implying that only first and second-order beliefs are relevant. In the second, both players take actions, making the entire hierarchy of beliefs relevant. The third explores a setting with no after-game in which agents care intrinsically about the information that is transmitted.

8.1 Multidimensional Actions under an Inactive Expert

The expert is a bureaucrat and the layman a politician. The bureaucrat knows the state of the world, $\Theta \subsetneq \mathbb{R}$. The politician is responsible for selecting policies across N different tasks, each of which is relevant to both agents. The agents differ in their preferences regarding how the optimal policies should depend on the state. This application characterizes the environments in which information sharing is feasible, as determined by the type of disagreement between the agents over policy choices.

In this game, $G = \big((A_i, \pi_i) : i \in \{e, \ell\}\big)$, the action spaces are $A_e = \{\triangleleft\}$ for the expert (who is inactive) and $A_\ell = \mathbb{R}^N$ for the layman. The layman chooses a policy $a_\ell^k \in \mathbb{R}$ for each task $k = 1, \ldots, N$, resulting in a policy vector $a_\ell = (a_\ell^1, \ldots, a_\ell^N) \in \mathbb{R}^N$. The payoff functions are:

$$\pi_\ell(\theta, a_\ell) = -\|a_\ell - \theta\, 1_N\|^2, \qquad \text{and} \qquad \pi_e(\theta, a_\ell) = -\omega\, \|a_\ell - b_1 - \theta\, b_2\|^2,$$

where $1_N \in \mathbb{R}^N$ is the N-dimensional vector of ones, $b_1 = (b_1^1, \ldots, b_1^N) \in \mathbb{R}^N$ represents the expert's *base-level bias*, $b_2 = (b_2^1, \ldots, b_2^N) \in \mathbb{R}^N$ is the expert's *directional bias*, and $\omega \in \mathbb{R}_+$ determines the relative importance of the expert's preferences. So, while the layman's ideal policy for each task k is θ, the expert's preferred policy for such task is $b_1^k + \theta\, b_2^k$. The agents may be aligned in some tasks and disagree in others: the agents are fully aligned on task k (for all states) if and only if $b_1^k = 0$ and $b_2^k = 1$.

Proposition 8.1. *The game G has a reduced-form representation $RF = \{(u_e, u_\ell)\}$,*

$$\textit{where} \quad u_e(\theta, h_e) = -\omega \int_{H_\ell} \|b_1 + \theta\, b_2 - f_\ell^1(h_e, h_\ell)\, 1_N\|^2\, dh_e^\infty, \tag{3}$$

$$u_\ell(h_\ell) = -\int_{\Theta\times H_e} N\left(\theta - f_\ell^1(h_e, h_\ell)\right)^2 dh_\ell^\infty, \ \textit{and} \tag{4}$$

$f_\ell^1(h_e, h_\ell)$ is the layman's first-order expectation of the state.

The layman's unique equilibrium policy vector is $f_\ell^1(h_e, h_\ell)\, 1_N$, i.e., the layman's expectation of the state applied uniformly across all tasks. As a result, utility u_ℓ is given by the negative of the residual variance of $f_\ell^1(\cdot)$. Since the expert cares about the policy vector selected by ℓ, the expert's utility is determined by her interim second-order beliefs: desiring ℓ's belief about $\theta\, 1_N$ to be as close as possible to e's favorite policy vector, $b_1 + \theta\, b_2$.

To capture the relation of the agents' directional preferences, write $\langle 1_N, b_2\rangle$ for the inner product between 1_N (the direction dictating the layman's favorite policy) and b_2 (the expert's directional bias vector). We say the agents exhibit *directional agreement* if $\langle 1_N, b_2\rangle \geq 0$, i.e., if the angle between 1_N and b_2 is at most 90 degrees. That is, if the preferred policy vectors shift in a similar direction as the state changes. Conversely, the agents have *directional disagreement* if $\langle 1_N, b_2\rangle < 0$, so the agents' preferred policies move in opposite directions as the state changes.

Proposition 8.2.

(i) If the agents have directional agreement, then the game G is fully revealing.

(ii) If the agents have directional disagreement, then the game G is concealing.

Proof. We first show (i). Assume $\langle 1_N, b_2\rangle \geq 0$. Write (u_e, u_ℓ) for the reduced-form of Proposition 8.1 and write

$$g(\theta,\theta') := -\omega\, \|b_1 + \theta\, b_2 - \theta'\, 1_N\|^2 = -\omega\, \Big(\|b_1 + \theta\, b_2\|^2 + \|\theta'\, 1_N\|^2 - 2\langle b_1 + \theta\, b_2, \theta'\, 1_N\rangle\Big).$$

Observe that $\frac{\partial^2 g}{\partial\theta\partial\theta'} = 2\,\omega\,\langle 1_N, b_2\rangle \geq 0$.

Fix $\theta, \theta' \in \Theta$ and write h_e for e's hierarchy with common degenerate belief for θ'. So, $u_e(\theta, h_e) = g(\theta,\theta')$. Since g is supermodular, it follows that u_e is supermodular on common degenerate beliefs and the result follows from Theorem 7.1.

Now we show (ii). Assume $\langle 1_N, b_2\rangle < 0$. By Theorem 7.1, it suffices to show that (u_e, u_ℓ) is statistically submodular. Fix a feasible probability space $\mathcal{P}$. Notice that the statistic f_ℓ^1 is acute. (See Lemma B.7.) Fix states $\overline{\theta} > \underline{\theta}$. Proposition 8.1 implies f_ℓ^1 is essential for (u_e, u_ℓ) and that

$$u_e(\overline{\theta}, h_e) - u_e(\underline{\theta}, h_e) = c_1 + c_2 \int_{H_\ell} f_\ell^1(h_e, h_\ell) \cdot dh_e^\infty,$$

where $c_1 := \omega\,\|b_1 + \overline{\theta}\, b_2\|^2 - \omega\,\|b_1 + \underline{\theta}\, b_2\|^2$ and $c_2 := 2\,\omega\,(\overline{\theta} - \underline{\theta})\,\langle 1_N, b_2\rangle$. Since $\langle 1_N, b_2\rangle < 0$, it follows that $c_2 < 0$ and (u_e, u_ℓ) is statistically submodular. (See Lemma 6.2.) ■

Proposition 8.2 offers a characterization of the parameters that determine whether (any) information sharing is possible, showing it solely depends on the angle between the agents' directional preferences (i.e., the angle between 1_N and b_2), and is independent of the base-level bias b_1. Hence, the key factor is not the magnitude of disagreement, but whether the agents have directional agreement.

Observe that Proposition 8.2 does not address whether information sharing is desirable. For the layman, information sharing is always beneficial, as it allows to make more informed decisions. In contrast, the expert may be worse off when information is shared, depending on her bias. To analyze this trade-off, consider utilitarian welfare under an exogenous information structure.[23] Two information structures are particularly relevant: full information (where the layman learns the state completely) and no information (where the layman receives a non-informative signal). We say that full revelation (respectively, no information) is welfare maximizing if it yields the highest sum of the agents' ex-ante utilities among all possible information structures.

To simplify the exposition, let $D := \frac{2}{N}\langle 1_N, b_2\rangle$ denote the agents' *directional alignment*. This parameter measures the degree of directional alignment between the agents' preferences, normalized by the number of tasks.

Proposition 8.3.

(i) If $D \geq 1$, then full revelation maximizes the payoffs of both agents.

(ii) If $D \leq 1$, then no information maximizes the expert's payoff, and full information maximizes the layman's payoffs.

(iii) If $\omega(1 - D) \leq 1$, then full revelation maximizes welfare.

(iv) If $\omega(1 - D) \geq 1$, then no information maximizes welfare.

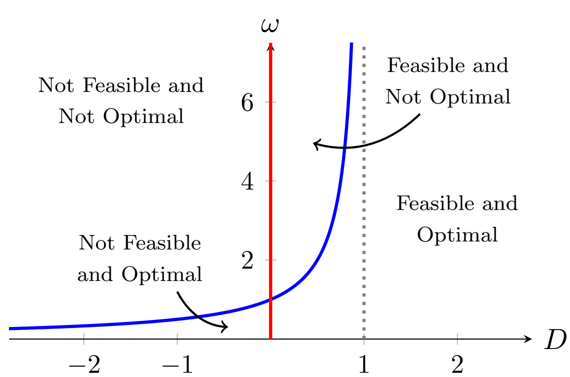

Figure 8.1. Feasibility and optimality of information sharing.

[23]An information structure consists of a set of public signals $\mathcal{S}$ and a mapping $\chi : \Theta \to \Delta(\mathcal{S})$.

Proposition 8.3 shows that the trade-off between revealing and concealing the state does not depend on the absolute level of disagreement (as captured by b_1), but rather entirely depends on the agents' directional agreement. When $D \geq 1$, the expert is sufficiently aligned with the way the layman responds to information. So, both agents benefit from full revelation of the state. When $D < 1$, revealing information is detrimental to the expert. Note, this negative effect persists even within the range $D \in [0, 1)$, where the agents still have directional agreement. In this region, the expert anticipates that the layman overreacts to information, making information sharing undesirable for the expert.

Propositions 8.2 and 8.3 identify a conflict between feasibility and optimality. The parameter space can be partitioned into four distinct regions, as illustrated in Figure 8.1. First, when $D \geq 0$ and ω is low, full revelation is both feasible and optimal. In this region, directional agreement exists and information benefits the layman, the relatively important agent. Second, when $D < 0$ and ω is high, information sharing is neither feasible nor optimal. Here, information sharing would harm the expert, the relatively important agent. Third, when $D < 0$ and ω is low, full information sharing would be optimal but is not feasible. Directional disagreement prevents information transmission that would benefit the layman, the relatively important agent. Fourth, when $1 > D \geq 0$ and ω is high, full information sharing would be feasible but is not optimal. Although the agents have directional agreement, the expert believes the layman overreacts to information. Consequently, a welfare-maximizing designer would prefer to conceal information even if full revelation is possible.

8.2 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. The second firm (the layman) does not observe θ. We analyze the extent to which an industry association (the designer) can induce information sharing via neutral mechanisms.

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

$$\pi_i(\theta, a_i, a_{-i}) = \theta\, a_i - \tfrac{1}{2}a_i^2 + \alpha\, a_i\, a_{-i}, \tag{5}$$

where $\alpha \in (-1, 1)$ is a commonly known parameter. 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.

When $\alpha < 0$, the game G captures a model of quantity competition in which a_i is the quantity supplied by firm i. Firm i has a marginal cost of c and faces a linear inverse demand given by $P_i = \theta - \frac{1}{2}a_i + \alpha a_{-i} + c$. So, 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 and constant marginal cost c. In this model, the action a_i is firm i's markup price, i.e., $a_i = p_i - c$. 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.[24]

Note, 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 the competitor's action. Hence, firms are concerned with their expectation of the state, their competitor's expectation of the state, their competitor's expectation of their expectation of the state, and so on.

To properly describe these belief-dependent preferences, we introduce hierarchies of expectations. Recall that $f_\ell^1 : H \to \mathbb{R}$ denotes the layman's first-order expectation of the state. Write $\iota(k) = e$ if k is even and $\iota(k) = \ell$ if k is odd. Fix $k \in \mathbb{N}$ and let

[24] The assumption $\alpha \in (-1, 1)$ captures all the environments in which the demand (resp. inverse demand) for firm i is more sensitive on its own action than to the competitor's action.

$i = \iota(k+1)$. Assuming that f^k_{-i} is defined, inductively define $f^{k+1}_i : H \to \mathbb{R}$ as

$$f^{k+1}_i(h_i, h_{-i}) := \int_{\hat{h}_{-i} \in H_{-i}} f^k_{-i}(h_i, \hat{h}_{-i}) \; d\text{marg}_{H_{-i}} h^\infty_i.$$

The statistic f^{k+1}_i captures i's $k+1^{th}$-*order expectation* of the state.[25] We will show that, in each induced Bayesian game of G, the agents' behavior and payoffs are characterized by the statistic $f^\alpha : H \to \mathbb{R}$, defined as:

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

As f^α is a weighted average of ℓ's hierarchies of expectations, it can be seen as an index of how the state is "commonly perceived." Lemma B.10 shows that f^α is a convergent positive sum of acute statistics, and as a consequence, it is acute.

Proposition 8.4. *G has a reduced-form representation* $\text{RF} = \{(u_e, u_\ell)\}$ *given by*

$$u_e(\theta, h_e) = \tfrac{1}{2}\left(\theta + \alpha \int_{H_\ell} f^\alpha(h_e, h_\ell) \; dh^\infty_e\right)^2 \; \textit{and}$$
$$u_\ell(h_\ell) = \tfrac{1}{2}\left(\int_{\Theta \times H_e} \theta + \alpha \, f^\alpha(h_e, h_\ell) \; dh^\infty_\ell\right)^2.$$

Proposition 8.4 provides a simple way to characterize the firms' value of information. The firms only care about how information "impacts" the statistic f^α. 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, the agents' unique equilibrium strategy is derived by an iterated substitution of their (linear) expected best responses. Consequently, the agents' strategies and their value of information is captured by f^α.

Proposition 8.5.

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

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

Proof. Write (u_e, u_ℓ) for the reduced form of Proposition 8.4. First we show (i).

[25] Notice, f^{k+1}_i depends on h_i but not on h_{-i}.

Assume $\alpha \in [0,1)$. Note, if there is common belief of θ' at (h_e, h_ℓ), then

$$f^\alpha(h_e, h_\ell) = (1+\alpha) \sum_{k=1}^{\infty} \alpha^{2k-2} \, \theta' = \sum_{k=0}^{\infty} \alpha^k \, \theta' = \tfrac{1}{1-\alpha} \, \theta'.$$

So, for each $\theta \in \Theta$, $u_e(\theta, h_e) = \frac{1}{2}\left(\theta + \frac{\alpha}{1-\alpha} \, \theta'\right)^2$, which satisfies increasing differences with respect to θ and θ'. Hence, u_e is supermodular in degenerate beliefs. The result follows from Theorem 7.1.

Now we show (ii). Assume $\alpha \in (-1, 0)$. By Theorem 7.1, it suffices to show that (u_e, u_ℓ) is statistically submodular. Note that f^α is acute and essential for (u_e, u_ℓ). (See Lemma B.10.) Moreover, for each pair of states $\overline{\theta} > \underline{\theta}$,

$$u_e(\overline{\theta}, h_e) - u_e(\underline{\theta}, h_e) = c_1 + c_2 \int_{H_\ell} f^\alpha(h_e, h_\ell) \; dh_e^\infty,$$

where $c_1 := \frac{1}{2}\big(\overline{\theta}^2 - \underline{\theta}^2\big)$ and $c_2 := \frac{1}{2}\alpha \left(\overline{\theta} - \underline{\theta}\right)$. Notice, since $\alpha < 0$, then $c_2 < 0$. Therefore, (u_e, u_ℓ) is statistically submodular. (See Lemma 6.2). ■

Proposition 8.5 provides a complete taxonomy of the parameters that allow for or preclude information sharing. Remarkably, results solely depend on the sign of α, i.e., the type of duopoly market that the firms face. Under price competition, a "good-news firm" has a higher willingness to pay for inducing optimistic beliefs and thus higher market prices. Hence, there is a message-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 for inducing pessimistic beliefs, which leads to reduced output from its competitor. Hence, the submodularity condition is satisfied and information sharing is not possible.

Proposition 8.5 is silent about whether information sharing is desirable or not. Proposition B.1 in the appendix addresses the impact of information on the firms' profit. The layman firm strictly prefers observing the state over receiving no information. First, fully revealing the state increases (resp. decreases) the profits of the expert firm if $\alpha > 0$ (resp. $\alpha < 0$). Second, fully revealing the state increases (resp. decreases) total industry profits if $\alpha > 1 - \sqrt{2}$ (resp. $\alpha < 1 - \sqrt{2}$). Third, fully revealing the state increases industry profit (over no information sharing) only if firms compete in prices ($\alpha > 0$) or if they compete in quantities and their products are weak substitutes ($1 - \sqrt{2} < \alpha < 0$).

The results identify a conflict between feasibility and optimality. In the case of quantity competition with weak substitutes ($1-\sqrt{2}<\alpha<0$) full revelation of the state increases the industry profits, but it is not feasible under any neutral mechanism: increasing industry profit requires the use of mechanisms that are not neutral.

8.3 Intrinsic Preferences for Information

This paper uses a reduced-form approach to analyze environments in which agents derive *instrumental* value from information. However, the results also extend to settings in which agents have *intrinsic* value for information.

A researcher (ℓ) seeks to elicit private information about a subject's (e) private 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 $\theta=1$ for a trait that is perceived as "good" and $\theta=0$ for a trait that has a social stigma.

The subject has image concerns—that is, the subject cares about whether the researcher perceives him as having the acceptable trait. Image concerns are modeled by a belief-dependent utility

$$u_e(\theta,h_e)=\int_{H_\ell} g\big(\theta,f_\ell^1(h_e,h_\ell)\big)\ d\mathrm{marg}_{H_\ell}h_e^\infty,$$

where f_ℓ^1 is the layman's first-order belief of the trait $\theta=1$ and $g:\Theta\times[0,1]\to\mathbb{R}$ is a mapping such that $g(\theta,\cdot)$ is increasing for each θ. So, regardless of the subject's actual trait, the subject prefers a higher value of f_ℓ^1. The researcher has a belief-based utility $u_\ell: H_\ell\to\mathbb{R}$. We assume that the researcher seeks to learn f_ℓ^1. (Hence, f_ℓ^1 is essential for (u_e,u_ℓ)). The outside options of both agents are normalized to zero.

In this context, non-neutral mechanisms are often out of reach as they depend on the subject's trait—something that, arguably, the researcher does not know. So, it is natural to use neutral mechanisms to elicit information about the subject's trait. The following result follows from verifying the supermodularity and submodularity conditions in this context:

Proposition 8.6.

(i) If $g(1,1)-g(1,0)\geq g(0,1)-g(0,0)$, then there is a neutral mechanism where the researcher learns the true state θ.

(ii) If $g : \Theta \times [0,1] \to \mathbb{R}$ has strictly decreasing differences, then in each equilibrium of each neutral mechanism the researcher's posterior equals the prior.

The extent to which the researcher can vs. cannot learn the subject's trait depends on the differences in g. 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 stigmatized subject ($\theta = 0$) has strictly higher incentives to be perceived as "good," then no information can be extracted.[26]

9 Discussion

9.1 Existence of a Reduced-Form Representation

The main results in this paper assume the existence of reduced forms and reduced-form representations. Rivera Mora [2025] shows that there are certain instances in which a reduced form and a reduced-form representation are guaranteed. The first is when G involves an inactive expert (i.e., A_e is a singleton). The second is when equilibrium behavior in G is fully captured by a pair of mappings $\varsigma_e : \Theta \times H_e \to \Delta(A_e)$ and $\varsigma_\ell : H_\ell \to \Delta(A_\ell)$. That is, if each associated Bayesian game has a unique Bayesian equilibrium that is induced by $(\varsigma_e, \varsigma_\ell)$, then G has some reduced form (u_e, u_ℓ). Moreover, the singleton $\text{RF} = \{(u_e, u_\ell)\}$ is a reduced-form representation of G.

Example 9.1. *Let Θ be a singleton, so there is no private information. The game G has two actions for each player and is described in Figure 4 of Aumann [1987]:*

$6,6$	$2,7$
$7,2$	$0,0$

Figure 9.1. Payoffs of e (first) and ℓ (second)

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 $(5,5)$ that lies outside of the convex hull of Nash equilibria payoffs.

[26] There are functions g that differ from parts (i) and (ii). In that case, the researcher may be able to extract some partial information about the state. (See Discussion 9.3.)

Rivera Mora [2025] shows that the game G has no reduced-form representation: since Θ and $H = H_e \times H_\ell$ are both singletons, the set of reduced-forms can only capture Nash equilibrium payoffs. However, there exists a neutral mechanism that generates the correlated equilibrium payoff vector $(5,5)$ by privately suggesting actions to agents. This payoff lies outside the convex hull of Nash equilibrium payoffs and thus cannot be captured by any reduced form.

Example 9.1 shows that the hierarchies of beliefs of Θ are not sufficiently rich to capture correlation generated by external signals. As a result, the game G does not admit a reduced-form representation. Nevertheless, by properly extending the agents' domain of uncertainty, an existence result can be established.[27] Within this extended domain, Rivera Mora [2025] proves the existence of reduced forms and reduced-form representations for a broad class of games, including all games with finite action sets.

9.2 Complete Characterization of Fully-Revealing Games

Theorem 7.1 provides a sufficient condition for G to be fully revealing for settings with a finite state space $\Theta \subset \mathbb{R}$. This section complements the result by providing necessary and sufficient conditions for full revelation of the state in settings with an arbitrary finite state space Θ.

Fix an arbitrary state space Θ. Following Vohra [2011], use the state space to define a completely connected network in which the set of vertices is Θ. Call a vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{n+1}) \in \Theta^{n+1}$ a **path** (of size n). A path $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{n+1})$ is a **cycle** if $\theta_{n+1} = \theta_1$. Call a cycle $(\theta_1, \ldots, \theta_{n+1}) \in \Theta^{n+1}$ **simple** if $\theta_i \neq \theta_j$ for each $i < j$ with $i, j \in \{1, \ldots, n\}$, 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. For each cycle $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{n+1})$ write

$$\mathcal{L}(g, \boldsymbol{\theta}) := \sum_{k=1}^{n} \Big(g(\theta_k, \theta_k) - g(\theta_k, \theta_{k+1}) \Big),$$

for the **length** of the cycle $\boldsymbol{\theta}$ with respect to the function g. Say g satisfies **cyclical monotonicity** if $\mathcal{L}(g, \boldsymbol{\theta}) \geq 0$ for each cycle $\boldsymbol{\theta}$ of arbitrary size.[28]

[27] This is reminiscent of the literature on redundant hierarchies. (See Ely and Peski [2006].) As Liu [2009] shows, enriching the underlying space of uncertainty can capture correlation precluded by hierarchies on Θ alone.

[28] One can verify that each cycle $\boldsymbol{\theta}$ can be "decomposed" into simple cycles $\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^m$.

Notice, cyclical monotonicity is defined even if Θ is not an ordered set. If $\Theta \subset \mathbb{R}$ and g has weakly increasing differences then g satisfies cyclical monotonicity. (See Lemma B.1.) The following example shows that the converse does not hold.

Example 9.2. *Fix a state space $\Theta = \{1,2,3\}$ and let $g : \Theta \times \Theta \to \mathbb{R}$ be such that*

$$g(n,m) = \begin{cases} 1 & \text{if } (n,m) \in \{(2,1),(2,2)\} \\ 0 & \text{otherwise.} \end{cases}$$

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 simple cycles for Θ (up to shifts). Observe 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 has no increasing differences.

Fix a reduced form (u_e, u_ℓ) and write $\eta_i : \Theta \to \mathrm{CDB}_i$ for the function that maps θ into i's hierarchy that has common degenerate belief for θ. Say u_e **satisfies cyclical monotonicity on degenerate beliefs** if the mapping $g(\theta,\theta') := u_e(\theta,\eta_e(\theta'))$ satisfies cyclical monotonicity. The following result provides a necessary and sufficient condition for full revelation of the state.

Theorem 9.1. *Assume G has a reduced-form representation* RF*. The game G is fully revealing if and only if there is $(u_e,u_\ell) \in \mathrm{RF}$ so that u_e satisfies cyclical monotonicity on degenerate beliefs.*

9.3 Partial Information Sharing

Theorems 5.1 and 6.1 provide sufficient conditions for showing that G is either fully revealing or concealing. In the applications of Section 8 the game is either fully revealing or concealing. However, not all games fit into these two categories. The following example illustrates this.

Example 9.3. *Let $\Theta = \{1,2,3\}$, with $\mu(\theta) = \frac{1}{3}$ for each θ. Consider a game in which the expert is inactive and $A_\ell = \mathbb{R}$. The agents' payoff functions are $\pi_\ell(\theta,a_\ell) = -(\theta - a_\ell)^2$ and $\pi_e(\theta,a_\ell) = -(\mathrm{b}(\theta) - a_\ell)^2$, where $\mathrm{b}(1) = 1, \mathrm{b}(2) = 3,$ and $\mathrm{b}(3) = 2$.*

Lemma B.13 shows that G is neither fully revealing nor concealing. Intuitively, full revelation of the state is not feasible, as an expert that observes the state $\theta = 2$ (resp. $\theta = 3$) will try to mimic an expert that observes the state $\theta = 3$ (resp. $\theta = 2$).

However, there is a neutral mechanism and a PBE involving cheap talk messages in which the expert directly reveals if $\theta = 1$ or $\theta \in \{2, 3\}$, thereby changing the layman's action and the payoffs of both agents. (See Lemma B.13.)

References

Charalambos D. Aliprantis and Kim C. Border. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. 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.

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.

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

Giacomo Calzolari and Alessandro Pavan. Monopoly with resale. *The RAND Journal of Economics*, 37(2):362–375, 2006a.

Giacomo Calzolari and Alessandro Pavan. On the optimality of privacy in sequential contracting. *Journal of Economic theory*, 130(1):168–204, 2006b.

Giacomo Calzolari and Alessandro Pavan. Sequential contracting with multiple principals. *Journal of Economic Theory*, 144(2):503–531, 2009.

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.

Amanda Friedenberg and Ernesto Rivera Mora. Notes on sequential games and perfect recall. Manuscript available at `www.ernestoriveramora.com/materials`, 2025.

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.

Andrew McClellan and Daniel Rappoport. Signaling good faith by taking stands. *Available at SSRN 4510675*, 2024.

Barry Meier. Origins of an epidemic: Purdue pharma knew its opiods were widely abused. *The New York Times*, 2018.

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

James A Mirrlees. Optimal tax theory: A synthesis. *Journal of public Economics*, 6 (4):327–358, 1976.

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.

Martin J Osborne and Ariel Rubinstein. *A course in game theory*. MIT press, 1994.

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. Mechanism design with belief-dependent preferences. *Journal of Economic Theory*, 216:105782, 2024.

Ernesto Rivera Mora. On the existence of reduced-form representations. Working Paper, 2025.

Jean-Charles Rochet. A necessary and sufficient condition for rationalizability in a quasi-linear context. *Journal of mathematical Economics*, 16(2):191–200, 1987.

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

A Michael Spence. Market signaling: Informational transfer in hiring and related screening processes. *(No Title)*, 1974.

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.

Stanley L Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.

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.

Supplemental Appendix

Appendix A Additional Definitions

A.1 Hierarchies of Beliefs

Let C_1 and C_2 be two compact metric spaces endowed with their Borel sigma algebra and let $\varphi : C_1 \to C_2$ be a measurable mapping. Write $\underline{\varphi} : \Delta(C_1) \to \Delta(C_2)$ for the function that maps each measure in $\Delta(C_1)$ to the image measure under $\varphi : C_1 \to C_2$. Notice that $\underline{\varphi}$ is measurable. (See Theorem 15.14 in Aliprantis and Border [2006].)

Write $D^1_\ell := \Theta$ (resp. $D^1_e := \{\diamond\}$) for the first-order domain of uncertainty for the layman (resp. the expert.)[29] 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. Write $D^{k+1}_i := D^k_i \times H^k_{-i}$ for the $(k+1)$-order domain of uncertainty of agent i and write $H^{k+1}_i := \left\{(h^1_i, \ldots, h^{k+1}_i) \in H^k_i \times \Delta(D^{k+1}_i) : \operatorname{marg}_{D^k_i} h^{k+1}_i = h^k_i\right\}$ for the set of collectively coherent $(k+1)$-order beliefs of agent i. Note that, if $(h^1_i, \ldots, h^{k+1}_i) \in H^{k+1}_i$, then $(h^1_i, \ldots, h^n_i) \in H^n_i$ for all $n \le k$; that is, each $(h^1_i, \ldots, h^{k+1}_i) \in H^{k+1}_i$ is coherent. Write

$$H_i = \left\{ (h^1_i, h^2_i, \ldots) \in \prod_{k=1}^{\infty} \Delta(D^k_i) : (h^1_i, \ldots, h^k_i) \in H^k_i \text{ for each } k \in \mathbb{N} \right\},$$

for the set of i's **collectively coherent hierarchies of beliefs**. So, $h_i = (h^1_i, h^2_i, \ldots) \in H_i$ is a hierarchy of beliefs for agent i. Call $H = H_e \times H_\ell$ the **belief structure**.

Brandenburger and Dekel [1993] constructs the canonical homeomorphism between the spaces H_i and $\Delta(D_i \times H_{-i})$. For each $h_i = (h^1_i, h^2_i, \ldots) \in H_i$, the extension h^∞_i is the unique probability measure in $\Delta(D_i \times H_{-i})$ so that $\operatorname{marg}_{D^k_i} h^\infty_i = h^k_i$ for each $k \in \mathbb{N}$. Conversely, for each $h^\infty_i \in \Delta(D_i \times H_{-i})$, there is a unique hierarchy $h_i = (h^1_i, h^2_i, \ldots)$ in H_i so that for each k, $\operatorname{marg}_{D^k_i} h^\infty_i = h^k_i$. Observe, the prior $\mu \in \Delta(\Theta)$ induces a hierarchy profile $(\tilde{h}_e, \tilde{h}_\ell) \in H$ that describe the agents' hierarchies absent any information sharing. This profile $(\tilde{h}_e, \tilde{h}_\ell)$ is the unique element of H that satisfies $\operatorname{marg}_{H_{-i}} \tilde{h}^\infty_i(\tilde{h}_{-i}) = 1$ for each $i \in \{e, \ell\}$ and $\operatorname{marg}_\Theta \tilde{h}^\infty_\ell = \mu$.

Fix a mechanism $\mathcal{M}$ and interim belief mappings $\beta_e : \Theta \times \mathcal{I}_e \to \Delta(V)$ and

[29] Notice that the expert has no uncertainty, and so, her first-order domain of uncertainty is trivial.

$\beta_\ell : \mathcal{I}_\ell \to \Delta(\Theta \times V)$. The mappings β_e and β_ℓ induce terminal belief mappings $\hat{\beta}_e : \Theta \times \mathcal{T}_e \to \Delta(Z)$ and $\hat{\beta}_\ell : \mathcal{T}_\ell \to \Delta(\Theta \times Z)$.[30]

Define mappings $\varphi_e^1 : Z \to D_e^1$ and $\varphi_\ell^1 : \Theta \times Z \to D_\ell^1$, so that $\varphi_\ell^1(\theta, z) = \theta$.[31] Note that φ_i^1 is measurable for each $i \in \{e, \ell\}$. Assume that the measurable maps φ_e^k and φ_ℓ^k are defined. Let $\varphi_e^{k+1} : Z \to D_e^{k+1}$ be defined by $\varphi_e^{k+1}(z) = (\varphi_e^k(z), \underline{\varphi}_\ell^k(\hat{\beta}_\ell(T_\ell)))$ where $T_\ell \in \mathcal{T}_\ell$ is the unique terminal information set of ℓ such that $z \in T_\ell$. Likewise, let $\varphi_\ell^{k+1} : \Theta \times \mathcal{T}_e \to D_\ell^{k+1}$ be defined by $\varphi_\ell^{k+1}(\theta, z) = (\varphi_\ell^k(\theta, z), \underline{\varphi}_e^k(\hat{\beta}_e(\theta, T_e)))$, where $T_e \in \mathcal{T}_e$ is the unique terminal information set of e such that $z \in T_e$. Since Θ and Z are both finite, the mappings φ_i^{k+1} are measurable for each $k \in \mathbb{N}$.

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

A.2 Neutral Mechanisms

Each neutral mechanism $\mathcal{M}$ has a set of actions X. Each agent $i \in \{e, \ell, \text{c}\}$ 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 $\mathcal{I}_i$ and the action correspondence $\mathcal{A}_i$ are such that i satisfies *no absentmindedness* and *perfect recall*.[32] Chance's behavior is described by a exogenous behavior strategy $\rho_\text{c} \in \prod_{I_\text{c} \in \mathcal{I}_\text{c} \backslash \mathcal{T}_\text{c}} \Delta(\mathcal{A}_\text{c}(I_\text{c}))$. Since the set of actions at a given information set can be a singleton, the definition does not require simultaneous moves. Note, no absentmindedness implies that each $i \in \{e, \ell, \text{c}\}$ knows the start of the game, i.e., $\{\emptyset\} \in \mathcal{I}_i$.

Write $S_i := \prod_{I_i \in \mathcal{I}_i \backslash \mathcal{T}_i} \mathcal{A}_i(I_i)$ for the set of pure strategies for i and write $S := S_e \times S_\ell \times S_\text{c}$. Write $\psi : S \rightrightarrows V$ for the **path correspondence** of $\mathcal{M}$. So, $\psi(s)$ denotes the set of nodes of V that constitute the path of play under $s \in S$. Write $\zeta : S \times V \to Z$ for the **end-node mapping**. So, $\zeta(s \mid v) \in Z$ is the end node that would be realized if the interaction starts at v and actions are subsequently played

[30]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 $\beta_\ell(T_\ell)(\Theta \times Z) = 1$.

[31]Notice, $D_e^1 = \{\diamond\}$ so φ_e^1 is trivial.

[32]In this setting, the definitions of no absentmindedness and perfect recall require care. (See Friedenberg and Rivera Mora [2025] for formal definitions and a discussion.) The results here only make use of no absentmindedness and perfect recall in so far as the solution concepts only fit for that environment.

according s.

behavior strategies of psychological games. Fix a psychological game $(\mathcal{M}, u_e, u_\ell)$. The expert's behavior is is described by a behavior strategy $\rho_e : \Theta \to \prod_{I_e \in \mathcal{I}_e \setminus \mathcal{T}_e} \Delta(\mathcal{A}_e(I_e))$. The layman's behavior is described by a behavior strategy $\rho_\ell \in \prod_{I_\ell \in \mathcal{I}_c \setminus \mathcal{T}_\ell} \Delta(\mathcal{A}_\ell(I_\ell))$. Observe, each pair (θ, ρ_e) (resp. each ρ_ℓ) induces a probability distribution over S_e (resp. S_ℓ) given by

$$P(s_e \mid \theta, \rho_e) := \prod_{I_e \in \mathcal{I}_e \setminus \mathcal{T}_e} \rho_e(\theta)(\text{proj}_{I_e} s_e) \qquad \text{and} \qquad P(s_\ell \mid \rho_\ell) := \prod_{I_\ell \in \mathcal{I}_\ell \setminus \mathcal{T}_\ell} \rho_\ell(\text{proj}_{I_\ell} s_\ell).$$

Likewise, write $P(s_\text{c} \mid \rho_\text{c})$ for the probability distribution over S_c that ρ_c induces.

The profile (ρ_e, ρ_ℓ) and the prior $\mu \in \Delta(\Theta)$ induce a distribution of paths over $\Theta \times V$. So, for each $(\theta, v) \in \Theta \times V$, the ex-ante probability that θ occurs and the path goes through v, given that (s_e, ρ_ℓ) (resp. (ρ_e, s_ℓ)) is played, is

$$P(\theta, v \mid s_e, \rho_\ell) := \sum_{(s_\ell, s_c) \in S_\ell \times S_c} \mu(\theta) \cdot P(s_\ell \mid \rho_\ell) \cdot P(s_c \mid \rho_c) \cdot \mathbb{1}[v \in \psi(s_e, s_\ell, s_c)], \quad \text{and}$$
$$P(\theta, v \mid \rho_e, s_\ell) := \sum_{(s_e, s_c) \in S_e \times S_c} \mu(\theta) \cdot P(s_e \mid \theta, \rho_e) \cdot P(s_c \mid \rho_c) \cdot \mathbb{1}[v \in \psi(s_e, s_\ell, s_c)].$$

For each $s = (s_e, s_\ell, s_c) \in S$, write $P(s \mid \theta, \rho) := P(s_e \mid \theta, \rho_e) \cdot P(s_\ell \mid \rho_\ell) \cdot P(s_c \mid \rho_c)$. The probability of g $(\theta, v) \in \Theta \times V$ under ρ is $P(\theta, v \mid \rho) := \sum_{s \in S} \mu(\theta) \cdot P(r \mid \theta, \rho) \cdot \mathbb{1}[v \in \psi(s_e, s_\ell, s_c)]$, and the probability of reaching v conditional on state θ is $P(v \mid \rho, \theta) := P(\theta, v \mid \rho) \cdot \mu(\theta)^{-1}$. Write

$$P(T_e, T_\ell \mid \theta, I_e, \rho, \beta_e) := \sum_{v \in I_e} \sum_{s \in S} \beta_e(\theta, I_e)(v) \cdot P(r \mid \theta, \rho) \cdot \mathbb{1}[\zeta(s, v) \in T_e \cap T_\ell]$$

for the probability that the e assigns to (T_e, T_ℓ) given that the state is θ, information set I_e is reached, ρ is played and e has interim beliefs β_e. Likewise, write

$$P(\theta, T_e, T_\ell \mid I_\ell, \rho, \beta_\ell) := \sum_{\theta \in \Theta} \sum_{v \in I_e} \sum_{s \in S} \beta_\ell(I_\ell)(\theta, v) \cdot P(r \mid \theta, \rho) \cdot \mathbb{1}[\zeta(s, v) \in T_e \cap T_\ell]$$

for the probability that ℓ assigns to (θ, T_e, T_ℓ) given that I_ℓ is reached, ρ is played, and ℓ has interim beliefs β_ℓ. The expected transfers induced by ρ are

$$Y_e(\theta, \mathcal{M}, \rho) := \sum_{T_e \in \mathcal{T}_e} \gamma_e(T_e) \cdot P(T_e \mid \rho, \theta), \quad \text{and} \quad Y_\ell(\mathcal{M}, \rho) := \sum_{T_\ell \in \mathcal{T}_\ell} \gamma_\ell(T_\ell) \cdot P(\Theta \times T_\ell \mid \rho).$$

Finally, for each $(\theta, v) \in \Theta \times Z$, write $\mathbb{P}(\theta, v) := P(\theta, v \mid \rho)$. So, each strategy profile ρ induces a probability measure $\mathbb{P} \in \Delta(\Theta \times Z)$ that specifies the distribution of states and terminal nodes.

Behavior strategies of supergames. Fix a mechanism $\mathcal{M}$. To define the strategies of the supergame $(\mathcal{M}, G)$, it is useful to extend the action correspondence from $\mathcal{A}_i : \mathcal{I}_i \backslash \mathcal{T}_i \rightrightarrows X$ to $\mathcal{A}_i : \mathcal{I}_i \rightrightarrows X \cup A_i$ so that (i) for each terminal information set $T_i \in \mathcal{T}_i$, $\mathcal{A}_i(T_i) = A_i$, and (ii) for each $I_i \in \mathcal{I}_i \backslash \mathcal{T}_i$, the action set $\mathcal{A}_i(I_i)$ corresponds to what it was originally defined in $\mathcal{M}$. A **behavior 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 **behavior strategy** for the layman is a vector $\sigma_\ell \in \prod_{I_\ell \in \mathcal{I}_\ell} \Delta(\mathcal{A}_\ell(I_\ell))$.

The probabilities of states and paths $P(\cdot)$, the expected transfers $Y_i(\cdot)$, and the terminal probability distribution $\mathbb{P} \in \Delta(\Theta \times Z)$ that σ induces are defined in analogous way as above.

Appendix B Omitted Proofs

B.1 Proofs of Section 5

Lemma B.1. *If $g : \Theta \times \Theta \to \mathbb{R}$ has 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 $(\theta_1, \ldots, \theta_{n+1})$ with $\theta_{n+1} = \theta_1$. Without loss, assume that $\theta_n = \max\{\theta_1, \ldots, \theta_{n+1}\}$. (Otherwise shift the indexes of the cycle.) Now, consider the cycle $(\theta_1, \ldots, \theta_{n-1}, \theta_{n+1})$ 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

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

Thus $\sum_{k=1}^{n} g(\theta_k, \theta_k) - g(\theta_k, \theta_{k+1}) \geq 0$. So, the result holds for n. ■

Lemma B.2. *The function $g : \Theta \times \Theta \to \mathbb{R}$ 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 $\mathrm{Id} : \Theta \to \Theta$ 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]. ■

Proof of Theorem 5.1. Fix a reduced form (u_e, u_ℓ) so that u_e is supermodular on common degenerate beliefs. We construct a mechanism $\mathcal{M}$ and an equilibrium thereof where the layman learns the state at all terminal nodes. In this mechanism, the layman and chance are inactive, and the expert selects to report an element from the set of pure strategies $S_e := \Theta$. If the expert selects a report $s_e = \theta$, then the layman observes s_e and the expert receives transfer $y_e(\theta)$. The layman receives a transfer y_ℓ regardless of the expert's report.

Write $\eta_i : \Theta \to \mathrm{CDB}_i$ for the function that maps each state θ to i's hierarchy in which there is common degenerate belief for θ. Set the transfer of the layman as $y_\ell := \underline{\pi}_\ell - \sum_{\theta \in \Theta} u_\ell\big(\eta_\ell(\theta)\big) \cdot \mu(\theta)$. Write $g : \Theta \times \Theta \to \mathbb{R}$ for the function given by $g(\theta, \theta') = u_e(\theta, \eta_e(\theta'))$. Since u_e is supermodular on common degenerate beliefs, g has increasing differences. By Lemmata B.1 and B.2, there is some function $z : \Theta \to \mathbb{R}$ such that, for each $\theta, \theta' \in \Theta$,

$$g(\theta, \theta) + z(\theta) \geq g(\theta, \theta') + z(\theta'). \tag{6}$$

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

$$c + g(\theta, \theta) + z(\theta) - \underline{\pi}_e(\theta) \geq 0. \tag{7}$$

Set $y_e(\theta) := z(\theta) + c$ for each $\theta \in \Theta$ for the transfers of the expert.

If suffices to show that (ρ, β) is an individually rational PBE of $(\mathcal{M}, u_e, u_\ell)$ that fully reveals the state. Consider the strategy $\rho_e : \Theta \to \Delta(S_e)$ so that $\rho_e(\theta)(\theta) = 1$ for each θ. Let ρ_ℓ the layman's (trivial) strategy and write β for beliefs consistent with $\rho = (\rho_e, \rho_\ell)$. Observe, if the state is θ and the expert reports $s_e = \theta'$, then $\delta_e(\theta, \{s_e\}) = \eta_e(\theta')$ and $\delta_\ell(\{s_e\}) = \eta_\ell(\theta')$. So, under (ρ, β), the layman fully learns the state. Moreover, the expert's payoff from state θ and selecting $s_e = \theta'$ is

$$u_e(\theta, \eta_e(\theta')) + y_e(\theta) = g(\theta, \theta') + y_e(\theta').$$

Notice, Equation (6) implies $g(\theta, \theta) + y_e(\theta) \geq g(\theta, \theta') + y_e(\theta')$. So, the expert that observes θ has no incentives to report a different state, and hence (ρ, β) is a PBE. In addition, Equation (7) implies $g(\theta, \theta) + y_e(\theta) \geq \underline{\pi}_e(\theta)$. Thus, (ρ, β) is IR for the expert. Finally, note that the layman's expected payoff is

$$y_\ell + \sum_{\theta \in \Theta} u_\ell\big(\eta_\ell(\theta)\big) \cdot \mu(\theta) = \underline{\pi}_\ell.$$

So, (ρ, β) is IR for the layman. ■

B.2 Proofs of Section 6

The revelation principle. We use the revelation principle in Rivera Mora [2024] for settings with belief-dependent preferences.[33]

Fix finite sets $M_e \subsetneq H_e$ and $M_\ell \subsetneq H_\ell$. Call the elements of M_i as hierarchy-messages for agent i. Say $M_e \times M_\ell$ is **belief closed** if, for each $h_i \in M_i$, $h_i^\infty(D_i^1 \times M_{-i}) = 1$. (Recall that $D_\ell^1 = \Theta$ and D_e^1 is trivial.) An **extended direct mechanism**, $\mathcal{M}^d := (\Theta, (\mathcal{Y}_i, M_i : i \in \{e, \ell\}), m)$, is a neutral mechanism in which $\mathcal{Y}_i \subsetneq \mathbb{R}$ is a finite set of transfers for agent i and $\mathcal{Y} := \mathcal{Y}_e \times \mathcal{Y}_\ell$ is the set of transfer profiles; 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 $m : \Theta \to \Delta(\mathcal{Y} \times M)$ is a **protocol** that describes the likelihood of chance selecting transfers and hierarchy-messages, given

[33] Such result states that there is no loss by focusing on a class of "extended direct mechanisms." Under an extended direct mechanism, the expert reports the state (only) to the mechanism. Given the report, the mechanism selects a transfer and private message for each agent. Each agent i (only) observes their transfer and their message. The message that each agent receives is a suggestion of the hierarchies of beliefs that the agent should have.

each report. So, the expert reports $\theta \in \Theta$ and chance selects $(y, h) \in \mathcal{Y} \times M$ according to the distribution $m(\theta)$. The set of terminal nodes is $Z = \Theta \times \mathcal{Y} \times M$, where the set Θ represents the reported state.[34] The expert's terminal information sets are of the form $I_e = \{\theta\} \times \{y_e\} \times \mathcal{Y}_\ell \times \{h_e\} \times M_\ell$ and the layman's information sets are of the form $I_\ell = \Theta \times \mathcal{Y}_e \times \{y_\ell\} \times M_e \times \{h_\ell\}$.

The set of pure strategies for the expert is $S_e = \Theta$. (The set of strategies for the layman is trivial.) So, in the psychological game, a strategy for the expert maps Θ to reporting strategies in $S_e = \Theta$. Write $\rho_e^* : \Theta \to \Delta(S_e)$ 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** $\phi \in \Delta(\Theta \times \mathcal{Y} \times M)$ is defined by $\phi(\theta, y, h) = \mu(\theta) \cdot m(\theta)(y, h)$.[35]

Definition B.1. *Fix an extended direct mechanism $\mathcal{M}^d$ and let ϕ be the ex-ante probability it induces. $\mathcal{M}^d$ is* ***believable*** *(BLV) if, for each $(\theta, y_e, h_e) \in \Theta \times Y_e \times M_e$ and each $(y_\ell, h_\ell) \in Y_\ell \times M_\ell$ the following hold:*

(i) $h_\ell^\infty(\theta, h_e) \cdot \phi(y_\ell, h_\ell) = \phi(\theta, h_e, y_\ell, h_\ell)$, and

(ii) $h_e^\infty(h_\ell) \cdot \phi(\theta, y_e, h_e) = \phi(\theta, y_e, h_e, h_\ell)$.

Believability implies 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 [2024].) So, provided that beliefs are honest, the expert's expected value of participating in $\mathcal{M}^d$ when the state is θ, and the report is θ' is

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

Likewise, provided that beliefs are honest, the layman's expected value of participating in $\mathcal{M}^d$ is

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

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

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

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

Proof of Lemma 6.1. Write $\mathcal{P} = (\Theta \times Z, \mathbb{P})$ for the induced terminal probability space induced by $\mathcal{M}$ and (ρ, β). Let $\mathcal{M}^d = \big(m, (\mathcal{Y}_i, M_i)_{i \in \{e,\ell\}}\big)$ be the induced direct mechanism of (ρ, β). (See Definition 5.1 in Rivera Mora [2024].)

Fix $\theta \in \Theta$, $y = (y_e, y_\ell) \in \mathcal{Y}$, and $h = (h_e, h_\ell) \in M$. First, we show that

$$m(\theta)(y, h) = \mathbb{P}[\mathbf{Y}_e = y_e, \mathbf{Y}_\ell = y_\ell, \mathbf{H}_e = h_e, \mathbf{H}_\ell = h_\ell \mid \mathbf{\Theta} = \theta]. \tag{8}$$

To show this, write $\mathcal{T}[y, h] := \big\{(T_e, T_\ell) \in \mathcal{T} : \delta_e(\theta, T_e) = h_e, \delta_\ell(T_\ell) = h_\ell, \gamma_e(T_e) = y_e, \gamma_\ell(T_\ell) = y_\ell\big\}$ and observe that $\mathbb{P}[\mathbf{\Theta} = \theta] = \mu(\theta)$. Hence,

$$\begin{aligned} m(\theta)(y, h) \cdot \mathbb{P}[\mathbf{\Theta} = \theta] &= \sum_{T_e \times T_\ell \in \mathcal{T}[\theta, y, h]} P(T_e \cap T_\ell \mid \theta, \rho) \cdot \mu(\theta) \\ &= \sum_{T_e \times T_\ell \in \mathcal{T}[\theta, y, h]} P(\{\theta\} \times T_e \cap T_\ell \mid \rho) \\ &= \mathbb{P}[\mathbf{Y}_e = y_e, \mathbf{Y}_\ell = y_\ell, \mathbf{H}_e = h_e, \mathbf{H}_\ell = h_\ell, \mathbf{\Theta} = \theta], \end{aligned}$$

where the first equality follows from definition of m, the second from definition of $P(\cdot \mid \cdot)$, and the last from the definition of $\mathbb{P}$. The revelation principle states that that $\mathcal{M}^d$ is believable, BIC, and IR. (See Theorem 5.1 in Rivera Mora [2024].) Moreover, notice that

$$\begin{aligned} \mathcal{V}_e(\theta, \theta' \mid \mathcal{M}^d) &= \sum_{h_e \in M_e} \sum_{y_e \in Y_e} (u_e(\theta, h_e) + y_e) \operatorname{marg}_{M_e \times Y_e} m(\theta')(y_e, h_e) \\ &= \sum_{h_e \in M_e} \sum_{y_e \in Y_e} (u_e(\theta, h_e) + y_e) \, \mathbb{P}[\mathbf{H}_e = h_e, \boldsymbol{Y}_e = y_e \mid \mathbf{\Theta} = \theta'] \\ &= \mathbb{E}[u_e(\theta, \mathbf{H}_e) + \boldsymbol{Y}_e \mid \mathbf{\Theta} = \theta'], \end{aligned}$$

where the second equality follows from Equation (8).

Fix $\theta, \theta' \in \Theta$, since $\mathcal{M}^d$ is BIC, it follows that $\mathcal{V}_e(\theta, \theta \mid \mathcal{M}^d) \geq \mathcal{V}_e(\theta, \theta' \mid \mathcal{M}^d)$. Consequently, $\mathbb{E}\big[u_e(\theta, \mathbf{H}_e) + \boldsymbol{Y}_e \mid \mathbf{\Theta} = \theta\big] \geq \mathbb{E}\big[u_e(\theta, \mathbf{H}_e) + \boldsymbol{Y}_e \mid \mathbf{\Theta} = \theta'\big]$ ■

B.3 Proofs of Section 6.1

Lemma B.3. *Fix a feasible probability space $\mathcal{P}$ induced by a mechanism $\mathcal{M}$ and profile (σ, β). Let δ_i be the hierarchy mapping induced by β. Write M_i is the range of δ_i. The following hold:*

(i) *The set $M_e \times M_\ell$ is belief closed.*

(ii) $h_e^\infty(h_\ell)\,\mathbb{P}[\mathbf{\Theta} = \theta, \mathbf{H}_e = h_e] = \mathbb{P}[\mathbf{\Theta} = \theta, \mathbf{H}_e = h_e, \mathbf{H}_\ell = h_\ell] = h_\ell^\infty(\theta, h_e)\,\mathbb{P}[\mathbf{H}_\ell = h_\ell]$.

(iii) $\text{marg}_{\,\Theta} h_\ell^\infty(\theta) \cdot \mathbb{P}[\mathbf{H}_\ell = h_\ell] = \mathbb{P}[\mathbf{\Theta} = \theta, \mathbf{H}_\ell = h_\ell]$.

(iv) $\text{marg}_{\,H_{-i}} h_i^\infty(h_j) \cdot \mathbb{P}[\mathbf{H}_i = h_i] = \mathbb{P}[\mathbf{H}_i = h_i, \mathbf{H}_j = h_j]$.

Proof. First we show (i). Write Z for the set of terminal nodes of $\mathcal{M}$. We show that $M_e \times M_\ell$ is belief closed. Fix $(h_e, h_\ell) \in M_e \times M_\ell$. Lemma C.4 in Rivera Mora [2024] states that $\text{marg}_{\,H_{-i}}\text{Supp}\,(h_i^\infty) \subseteq M_{-i}$. Hence, $M_e \times M_\ell$ is belief closed.

We now show (ii). Fix $(\theta, h_e, h_\ell) \in \Theta \times H_e \times H_\ell$ and write $\mathcal{T}_e[h_e \mid \theta] := \{T_e \in \mathcal{T}_e : \delta_e(\theta, T_e) = h_e\}$ and $\mathcal{T}_\ell[h_\ell] := \{T_\ell \in \mathcal{T}_\ell : \delta_\ell(T_\ell) = h_\ell\}$. Observe that

$$
\begin{aligned}
h_e^\infty(h_\ell) \cdot \mathbb{P}[\mathbf{\Theta} = \theta, \mathbf{H}_e = h_e] &= \sum_{T_e \in T_e[h_e|\theta]} h_e^\infty(h_\ell) \cdot \mathbb{P}(\{\theta\} \times T_e) \\
&= \sum_{T_e \in \mathcal{T}_e[h_e|\theta]} \sum_{T_\ell \in \mathcal{T}_\ell[h_\ell]} \sum_{z \in T_e \cap T_\ell} \beta_e(\theta, T_e)(z) \cdot \mathbb{P}(\{\theta\} \times T_e) \\
&= \sum_{T_e \in \mathcal{T}_e[h_e|\theta]} \sum_{T_\ell \in \mathcal{T}_\ell[h_\ell]} \sum_{z \in T_e \cap T_\ell} \mathbb{P}(\theta, z) \\
&= \mathbb{P}[\mathbf{\Theta} = \theta, \mathbf{H}_e = h_e, \mathbf{H}_\ell = h_\ell],
\end{aligned}
$$

where the second equality follows from Lemma C.4 in Rivera Mora [2024] and the third equality from consistency of (ρ, β). Showing $\mathbb{P}[\mathbf{\Theta} = \theta, \mathbf{H}_e = h_e, \mathbf{H}_\ell = h_\ell] = h_\ell^\infty(\theta, h_e) \cdot \mathbb{P}[\mathbf{H}_\ell = h_\ell]$ follows from an analogous argument.

Finally, (iii) follows from adding (ii) over $h_e \in M_e$ and (iv) from adding (ii) over $h_\ell \in M_\ell$. ■

For the following lemmata, write $\iota(k) = e$ if k is even and $\iota(k) = \ell$ if k is odd.

Lemma B.4. *Fix a feasible probability space $\mathcal{P}$. Write $\mathbf{F}^k_{\iota(k)}$ for the random variable associated with the $k-$order hierarchy of expectation $f^k_{\iota(k)}$. The following hold:*

(i) *$\mathbf{F}^1_\ell$ is a version of the conditional expectation of $\mathbf{\Theta}$ given $\mathbf{H}_\ell$.*

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

Proof. We first show part (i). Write M_ℓ for the range of δ_ℓ and fix $h_\ell \in M_\ell$. Observe that $\mathbf{F}^1_\ell(\theta, z) = \sum_{\theta\in\Theta} \theta \cdot \text{marg}_{H_e} h^\infty_\ell(\theta)$ for each (θ, z) such that $\mathbf{H}_\ell(\theta, z) = h_\ell$. hence,

$$\begin{aligned}
\mathbb{E}\big[\mathbf{F}^1_\ell \cdot \mathbb{1}[\mathbf{H}_\ell = h_\ell]\ \big] &= \sum_{\theta\in\Theta} \theta \cdot \text{marg}_{H_e} h^\infty_\ell(\theta) \cdot \mathbb{P}[\mathbf{H}_\ell = h_\ell] \\
&= \sum_{\theta\in\Theta} \theta \cdot \mathbb{P}[\mathbf{\Theta} = \theta, \mathbf{H}_\ell = h_\ell] \\
&= \mathbb{E}\big[\mathbf{\Theta} \cdot \mathbb{1}[\mathbf{H}_\ell = h_\ell]\ \big],
\end{aligned}$$

where the second equality follows from Lemma B.3 (iii). Thus, $\mathbf{F}^1_\ell$ is a version of the conditional expectation of $\mathbf{\Theta}$ given $\mathbf{H}_\ell$.

We now show part (ii). For each $h_\ell \in H_\ell$, write $E^1_\ell(h_\ell) := \sum_{\theta\in\Theta} \theta \cdot h^1_\ell(\theta)$ for the ℓ's first-order expectation given h_ℓ. Fix $k \in \mathbb{N}$ and write $i = \iota(k+1)$ and $j = \iota(k)$. Assuming that $E^k_j : H_j \to \mathbb{R}$ is defined, inductively define, $E^{k+1}_i(h_i) := \int_{H_j} E^k_j(h_j)\ d\text{marg}_{H_j} h^\infty_i$. Observe, by construction, $\mathbf{F}^k_i(\theta, z) = E^k_i(\mathbf{H}_i(\theta, z))$ for each $(\theta, z) \in \Theta \times Z$. Hence,

$$\begin{aligned}
\mathbb{E}\big[\mathbf{F}^k_i \cdot \mathbb{1}[\mathbf{H}_i = h_i]\ \big] &= E^k_i(h_i) \cdot \mathbb{P}[\mathbf{H}_i = h_i] \\
&= \sum_{h_j\in M_j} E^{k-1}_j(h_j) \cdot \text{marg}_{H_j} h^\infty_i(h_j) \cdot \mathbb{P}[\mathbf{H}_i = h_i] \\
&= \sum_{h_j\in M_j} E^{k-1}_j(h_j) \cdot \mathbb{P}[\mathbf{H}_i = h_i, \mathbf{H}_j = h_j] \\
&= \mathbb{E}\big[\mathbf{F}^{k-1}_j \cdot \mathbb{1}[\mathbf{H}_i = h_i]\ \big],
\end{aligned}$$

where the second equality follows from Equation Lemma B.3 (iv). Thus, $\mathbf{F}^k_i$ is a version of the conditional expectation of $\mathbf{F}^{k-1}_j$ given $\mathbf{H}_i$. ■

Lemma B.5. *Fix a probability space $\mathcal{P} = (\Omega, \mathcal{F}, \mathbb{P})$ in which $\mathbf{X}$ and $\mathbf{X}'$ have finite second moments.*

(i) If $\mathcal{F}' \subseteq \mathcal{F}$ is a sigma algebra and $\mathbf{X}$ is $\mathcal{F}'$-measurable, then $\text{Cov}[\mathbf{X}, \mathbf{X}'] = \text{Cov}\big[\mathbf{X}, \mathbb{E}[\mathbf{X}' \mid \mathcal{F}']\big]$.

(ii) If $g : \mathbb{R} \to \mathbb{R}$ is weakly increasing and $g(\mathbf{X})$ has finite second moments, then $\text{Cov}[\mathbf{X}, g(\mathbf{X})] \geq 0$. *Moreover, if* $\text{Cov}[\mathbf{X}, g(\mathbf{X})] = 0$, *then $g(\mathbf{X}) = g(\mathbb{E}[\mathbf{X}])$ almost surely.*

(iii) If the function $g(x) = \mathbb{E}[\mathbf{X} \mid \mathbf{X}' = x]$ is decreasing in x, then $\text{Cov}[\mathbf{X}, \mathbf{X}'] \leq 0$.

Proof. First we show (i). Notice that

$$\begin{aligned}\operatorname{Cov}[\mathbf{X},\mathbf{X}'] &= \mathbb{E}[\mathbf{X}\cdot\mathbf{X}'] - \mathbb{E}[\mathbf{X}]\cdot\mathbb{E}[\mathbf{X}'] \\ &= \mathbb{E}[\mathbb{E}[\mathbf{X}\cdot\mathbf{X}' \mid \mathcal{F}']] - \mathbb{E}[\mathbf{X}]\cdot\mathbb{E}[\mathbb{E}[\mathbf{X}' \mid \mathcal{F}']] \\ &= \mathbb{E}[\mathbf{X}\cdot\mathbb{E}[\mathbf{X}' \mid \mathcal{F}']] - \mathbb{E}[\mathbf{X}]\cdot\mathbb{E}[\mathbb{E}[\mathbf{X}' \mid \mathcal{F}']] \\ &= \operatorname{Cov}[\mathbf{X},\mathbb{E}[\mathbf{X}' \mid \mathcal{F}']],\end{aligned}$$

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].)

Now we show (ii). Write $x = \mathbb{E}[\mathbf{X}]$ and notice that

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

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

Now we show (iii). By the law of iterated expectations,

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

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

$$\operatorname{Cov}[\mathbf{X},\mathbf{X}'] = \mathbb{E}[\mathbf{X}\cdot\mathbf{X}'] - \mathbb{E}[\mathbf{X}]\cdot\mathbb{E}[\mathbf{X}'] = \mathbb{E}[\mathbf{X}'\cdot g(\mathbf{X}')] - \mathbb{E}[g(\mathbf{X}')]\cdot\mathbb{E}[\mathbf{X}'] = \operatorname{Cov}[\mathbf{X}',g(\mathbf{X}')].$$

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

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

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

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

as desired. ■

For the next lemmata, it will be convenient to write $\mathbf{F}_e^0 := \mathbf{\Theta}$. Recall that $\iota(m) = e$ if m is odd and $\iota(m) = \ell$ if m is even.

Lemma B.6. *If $k, m, n \geq 0$ satisfy $k + m \in \{2n, 2n-1\}$, then* $\mathrm{Cov}\big[\mathbf{F}_{\iota(k)}^k, \mathbf{F}_{\iota(m)}^m\big] = \mathrm{Var}\big[\mathbf{F}_{\iota(n)}^n\big]$.

Proof. Fix $n \in \mathbb{N}$. Without loss of generality, we will show that the result holds for each pair $k \geq m$ that satisfies $k + m \in \{2n, 2n-1\}$. Note, 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 such pairs (k, m) by using an induction argument over d.

Note, if $d = 0$, then $k = m = n$. Thus, $\mathrm{Cov}\big[\mathbf{F}_{\iota(k)}^k, \mathbf{F}_{\iota(k)}^k\big] = \mathrm{Var}\big[\mathbf{F}_{\iota(k)}^k\big]$, as desired.

Now, assume that the result holds for each pair (k, m) such that $k - m = d$. Consider the case $k \neq m \bmod 2$. Observe that $\mathbf{F}_{\iota(k+1)}^{k+1} = \mathbb{E}\big[\mathbf{F}_{\iota(k)}^k \mid \mathcal{F}_{\iota(k+1)}\big]$ and $\mathbf{F}_{\iota(m)}^m$ is $\mathcal{F}_{\iota(m)}$-measurable. (See Lemma B.4.) Thus,

$$\mathrm{Cov}\big[\mathbf{F}_{\iota(k+1)}^{k+1}, \mathbf{F}_{\iota(m)}^m\big] = \mathrm{Cov}\big[\mathbf{F}_{\iota(k)}^k, \mathbf{F}_{\iota(m)}^m\big] = \mathrm{Cov}\big[\mathbf{F}_{\iota(n)}^n, \mathbf{F}_{\iota(n)}^n\big] = \mathrm{Var}[\mathbf{F}_{\iota(n)}^n],$$

where the first equality follows from Lemma B.5 and the second equality follows from the induction hypothesis.

Consider the case $k = m \bmod 2$. Observe that $\mathbf{F}_{\iota(m)}^m = \mathbb{E}\big[\mathbf{F}_{\iota(m-1)}^{m-1} \mid \mathcal{F}_{\iota(m)}\big]$ and $\mathbf{F}_{\iota(k)}^k$ is $\mathcal{F}_{\iota(k)}$-measurable. (See Lemma B.4.) Thus,

$$\mathrm{Cov}\big[\mathbf{F}_{\iota(k)}^k, \mathbf{F}_{\iota(m-1)}^{m-1}\big] = \mathrm{Cov}\big[\mathbf{F}_{\iota(k)}^k, \mathbf{F}_{\iota(m)}^m\big] = \mathrm{Cov}\big[\mathbf{F}_{\iota(n)}^n, \mathbf{F}_{\iota(n)}^n\big] = \mathrm{Var}[\mathbf{F}_{\iota(n)}^n],$$

where the first equality follows from Lemma B.5 and the second equality follows from the induction hypothesis. Therefore, the result holds for $d + 1$. ■

Lemma B.7. *For each $k \in \mathbb{N}$, the statistic $f_{\iota(k)}^k$ is acute.*

Proof. Fix a feasible probability space $\mathcal{P}$ and $k \in \mathbb{N}$. Write $n = \lfloor \frac{k+1}{2} \rfloor$ and observe that $\mathrm{Cov}[\mathbf{F}_{\iota(k)}^k, \mathbf{\Theta}] = \mathrm{Var}[\mathbf{F}_{\iota(n)}^n] \geq 0$. (Use Lemma B.6, with $m = 0$.) Moreover, if $\mathcal{P}$

is such that $\mathrm{Cov}\big[\mathbf{F}^k_{\iota(k)}, \mathbf{\Theta}\big] = 0$, then $\mathrm{Var}\big[\mathbf{F}^n_{\iota(n)}\big] = 0$. This implies that $\mathbf{F}^n_{\iota(n)} = \mathbb{E}[\mathbf{\Theta}]$ almost surely. Moreover, since $k \geq n$, it follows that $\mathbf{F}^k_{\iota(k)} = \mathbb{E}[\mathbf{\Theta}] = f^k_{\iota(k)}(\tilde{h})$ a.s. ■

Lemma B.8. *If $g : \mathbb{R} \to \mathbb{R}$ is a weakly increasing function, then $g \circ f^1_\ell$ is an acute statistic.*

Proof. Notice, by Lemma B.4, $\mathbf{F}^1_\ell$ is a version of the conditional expectation of $\mathbf{\Theta}$. This implies that $\mathrm{Cov}[g(\mathbf{F}^1_\ell), \mathbf{\Theta}] = \mathrm{Cov}[g(\mathbf{F}^1_\ell), \mathbf{F}^1_\ell] \geq 0$. (See Lemma B.5.) Moreover, if $\mathrm{Cov}[g(\mathbf{F}^1_\ell), \mathbf{F}^1_\ell] = 0$, then $g(\mathbf{F}^1_\ell) = g(f^1_\ell(\tilde{h}))$ almost surely. (See Lemma B.5.) ■

Lemma B.9. *Assume $\Theta = \{\underline{\theta}, \overline{\theta}\}$ with $\overline{\theta} > \underline{\theta}$ if $f(h) = g\big(h^1_\ell(\overline{\theta})\big)$ for some increasing mapping $g : [0,1] \to \mathbb{R}$, then f is acute.*

Proof. Notice that the mapping $\hat{g}(x) = x \cdot (\overline{\theta} - \underline{\theta}) - \underline{\theta}$ is increasing. So, the composition $g \circ \hat{g}$ is also increasing. Hence, the result follows from Lemma B.8 and observing that $f = g \circ \hat{g} \circ f^1_\ell$. ■

Lemma B.10. *If $f^\alpha(h) = (1+\alpha) \sum_{k=1}^{\infty} \alpha^{2k-2} f^{2k-1}_\ell(h)$, then f^α is an acute statistic.*

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

Write $K := \max\{|\theta| : \theta \in \Theta\}$ and $K_\alpha := (1+\alpha) \sum_{k=1}^{\infty} \alpha^{2k-2} K$. 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 $|\mathbf{F}^k_\ell| \leq K$ for each $k \in \mathbb{N}$. Thus,

$$|\mathbf{X}^n| = |(1+\alpha) \textstyle\sum_{k=1}^{n} \alpha^{2k-2} \mathbf{F}^{2k-1}_\ell| \leq (1+\alpha) \sum_{k=1}^{n} \alpha^{2k-2} |\mathbf{F}^{2k-1}_\ell| \leq (1+\alpha) \sum_{k=1}^{n} \alpha^{2k-2} K$$

Thus, $|\mathbf{X}^n| \leq K_\alpha$. Moreover, since $|\mathbf{\Theta}|$ is bounded by K, it follows that for each $n \in \mathbb{N}$, $|\mathbf{\Theta} \cdot \mathbf{X}^n|$ is bounded by $K \cdot K_\alpha$. So, by the dominated convergence theorem, $\mathbb{E}[\mathbf{F}^\alpha \cdot \mathbf{\Theta}] = \mathbb{E}[\lim_{k\to\infty} \mathbf{X}^k \cdot \mathbf{\Theta}] = \lim_{k\to\infty} \mathbb{E}[\mathbf{X}^k \cdot \mathbf{\Theta}]$. Likewise, $\mathbb{E}[\mathbf{F}^\alpha] = \mathbb{E}[\lim_{k\to\infty} \mathbf{X}^k] = \lim_{k\to\infty} \mathbb{E}[\mathbf{X}^k]$. Hence,

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

where the last equality from the fact that f^k is acute for each $k \in \mathbb{N}$. (See Lemma B.7.) Thus, $\mathrm{Cov}[\mathbf{F}^\alpha, \mathbf{\Theta}] \geq 0$. Moreover, if $\mathrm{Cov}[\mathbf{F}^\alpha, \mathbf{\Theta}] = 0$, it follows that $\mathrm{Cov}[\mathbf{F}^k_\ell, \mathbf{\Theta}] = 0$ and thus $\mathbf{F}^k = f^k(\tilde{h})$ almost surely. Therefore, if $\mathrm{Cov}[\mathbf{F}^\alpha, \mathbf{\Theta}] = 0$, then $\mathbf{F}^\alpha = f^\alpha(\tilde{h})$ almost surely. Hence, f^α is acute. ∎

Proof of Lemma 6.2. Fix a psychological game $(\mathcal{M}, u_e, u_\ell)$, let (ρ, β) be a PBE of $(\mathcal{M}, u_e, u_\ell)$, and let (δ_e, δ_ℓ) be the associated hierarchy mappings. Write $\mathcal{P}$ for the terminal probability space induced by (ρ, β) and write M_i for the range of δ_i. (Note that each M_i is finite.)

Assume that $\mathcal{P}$ impacts (u_e, u_ℓ). Since f is essential, we have $\mathbb{P}\big[\mathbf{F} = f(\tilde{h})\big] < 1$. Moreover, since f is an acute statistic, we know that $\mathrm{Cov}[\mathbf{F}, \mathbf{\Theta}] > 0$. Therefore, the function $g(\theta) := \mathbb{E}[\mathbf{F} \mid \mathbf{\Theta} = \theta]$ is not weakly decreasing (see Lemma B.5 (iii)). In particular, there exist states $\overline{\theta}$ and $\underline{\theta}$ with $\overline{\theta} > \underline{\theta}$ such that

$$\mathbb{E}[\mathbf{F} \mid \mathbf{\Theta} = \overline{\theta}] > \mathbb{E}[\mathbf{F} \mid \mathbf{\Theta} = \underline{\theta}]. \tag{9}$$

By hypothesis, there exist constants $c_1 \in \mathbb{R}$ and $c_2 < 0$ such that for each $h_e \in H_e$:

$$u_e(\overline{\theta}, h_e) - u_e(\underline{\theta}, h_e) = c_1 + c_2 \int_{H_\ell} f(h_e, h_\ell)\, dh_e^\infty.$$

Note that $M_e \times M_\ell$ is belief closed (see Lemma B.3 (i)). Thus, if $h_e \in M_e$, then

$$\begin{aligned} u_e(\overline{\theta}, h_e) - u_e(\underline{\theta}, h_e) &= c_1 + c_2 \sum_{h_\ell \in M_\ell} f(h_e, h_\ell)\, h_e^\infty(h_\ell) \\ &= c_1 + c_2 \sum_{h_\ell \in M_\ell} f(h_e, h_\ell) \cdot \mathbb{P}[\mathbf{H}_\ell = h_\ell \mid \mathbf{H}_e = h_e] \\ &= c_1 + c_2 \cdot \mathbb{E}\big[f(\mathbf{H}_e, \mathbf{H}_\ell) \mid \mathbf{H}_e = h_e\big] \\ &= c_1 + c_2 \cdot \mathbb{E}\big[\mathbf{F} \mid \mathbf{H}_e = h_e\big], \end{aligned}$$

where the second equality follows from Lemma B.3 (iv). hence, for each $\theta \in \Theta$:

$$\begin{aligned} \mathbb{E}\big[u_e(\overline{\theta}, \mathbf{H}_e) - u_e(\underline{\theta}, \mathbf{H}_e) \mid \mathbf{\Theta} = \theta\big] &= c_1 + c_2 \cdot \mathbb{E}\big[\mathbb{E}[\mathbf{F} \mid \mathbf{H}_e] \mid \mathbf{\Theta} = \theta\big] \\ &= c_1 + c_2 \cdot \mathbb{E}[\mathbf{F} \mid \mathbf{\Theta} = \theta], \end{aligned} \tag{10}$$

where the second equality follows from the law of iterated expectations. (See Durrett

[2019].) Since $c_2 < 0$, Equations (9) and (10) imply that

$$\mathbb{E}\big[u_e(\overline{\theta}, \mathbf{H}_e) - u_e(\underline{\theta}, \mathbf{H}_e) \mid \mathbf{\Theta} = \overline{\theta}\big] < \mathbb{E}\big[u_e(\overline{\theta}, \mathbf{H}_e) - u_e(\underline{\theta}, \mathbf{H}_e) \mid \mathbf{\Theta} = \underline{\theta}\big],$$

as desired. ∎

B.4 Proofs of Section 7

Fix a mechanism $\mathcal{M} = (\cdot, \mathcal{T}_e, \mathcal{T}_\ell)$, interim belief mappings $\beta \in \text{cons}(\mathcal{M})$, and belief-based 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 $\text{BG}(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$ if $\hat{\sigma}_e$ (resp. $\hat{\sigma}_\ell$) is the restriction of σ_e (resp. $\hat{\sigma}_\ell$) to $\Theta \times \mathcal{T}_e$ (resp. $\mathcal{T}_\ell$). Similarly, (σ_e, σ_ℓ) 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 $\text{BG}(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$ induce the strategy profile (σ_e, σ_ℓ) of the supergame $(\mathcal{M}, G)$ if

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

Lemma B.11. *Fix a PBE (σ, β) of $(\mathcal{M}, G)$ and utility functions (u_e, u_ℓ). Assume σ induces ρ on $(\mathcal{M}, u_e, u_\ell)$ and σ induces $\hat{\sigma}$ on $\text{BG}(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$. If each tuple $(\theta, T_e, T_\ell) \in \theta \times \mathcal{T}_e \times \mathcal{T}_\ell$ satisfies*

(i) $u_e(\theta, \delta_e(\theta, T_e)) = \Pi_e(\hat{\sigma} \mid \theta, T_e, \beta_e)$, and

(ii) $u_\ell(\delta_\ell(T_\ell)) = \Pi_\ell(\hat{\sigma} \mid T_\ell, \beta_\ell)$,

then (ρ, β) is a PBE of $(\mathcal{M}, u_e, u_\ell)$ that is equivalent to (σ, β).

Proof. We first show (ρ, β) is equivalent (σ, β). Notice that for each $(\theta, I_e) \in \Theta \times \mathcal{I}_e$,

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

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

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

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

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

Proof of Lemma 7.1. Fix a PBE (ρ, β) of the psychological game $(\mathcal{M}, u_e, u_\ell)$. Let $\delta_e : \Theta \times \mathcal{T}_e \to H_e$, $\delta_\ell : \mathcal{T}_\ell \to H_\ell$ 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 with (u_e, u_ℓ) for the Bayesian game $\mathrm{BG}(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$. Notice that Lemma B.11 applies and therefore (σ, β) is a PBE of $(\mathcal{M}, G)$ that is equivalent to (ρ, β). ■

Lemma B.12. *Fix a reduced form (u_e, u_ℓ) of G and let $(\tilde{h}_e, \tilde{h}_\ell) \in H$ be the hierarchy profile absent any information. If $\Pi^s_e(\theta) = u_e(\theta, \tilde{h}_e)$ and $\Pi^s_\ell = u_\ell(\tilde{h}_\ell)$. Then, (Π^s_e, Π^s_ℓ) are silent payoffs for G.*

Proof. Fix a mechanism $\mathcal{M} = (\cdot, \mathcal{T}_e, \mathcal{T}_\ell)$ in which the set of nodes V is a singleton. So, in $\mathcal{M}$, the agents do not interact at all. Let $\beta \in \mathrm{cons}(\mathcal{M})$ and let δ be the associated hierarchies. So, $\delta_e(\theta, T_e) = \tilde{h}_e$, $\delta_\ell(T_\ell) = \tilde{h}_\ell$ and $\mathrm{BG}(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$ is the silent game of G. Since (u_e, u_ℓ) is a reduced form of G, then there is some BE $\hat{\sigma}$ of $\mathrm{BG}(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$ such that $\Pi_e(\hat{\sigma} \mid \theta, T_e, \beta_e) = u_e(\theta, \tilde{h}_e)$ and $\Pi_\ell(\hat{\sigma} \mid T_\ell, \beta_\ell) = u_e(\tilde{h}_\ell)$. Hence, the mapping $\Pi_e^s(\theta) := u_e(\theta, \tilde{h}_e)$ and the value $\Pi_\ell^s := u_\ell(\tilde{h}_\ell)$ are silent payoffs of G. ■

Proof of Theorem 7.1. First we show Part (i). Let (u_e, u_ℓ) be a reduced form of G, and assume u_e is supermodular on common degenerate beliefs. By Theorem 5.1, there is a mechanism $\mathcal{M}$ and an individually rational PBE (ρ, β) of $(\mathcal{M}, u_e, u_\ell)$ that is perfectly revealing. Since (u_e, u_ℓ) is a reduced form of G, there exists a strategy profile σ such that (σ, β) is a PBE of $(\mathcal{M}, G)$ that is equivalent to (ρ, β). (See Lemma 7.1.) Since (ρ, β) is perfectly revealing, it follows that (σ, β) is also be perfectly revealing. Therefore, the game G is perfectly-revealing.

Now we show Part (ii). Let RF be a reduced-form representation of G, and assume each $(u_e, u_\ell) \in \mathrm{RF}$ is statistically submodular. So, if (σ, β) be a PBE of $(\mathcal{M}, G)$, then there exists a mechanism $\mathcal{M}'$, a reduced form $(u_e, u_\ell) \in \mathrm{RF}$, and a PBE (ρ, β') of $(\mathcal{M}', u_e, u_\ell)$ that is equivalent to (σ, β). By Theorem 6.1, (u_e, u_ℓ) is concealing. Thus, the terminal probability space $\mathcal{P}$ induced by (ρ, β') satisfies $\mathbb{E}[u_e(\mathbf{\Theta}, \mathbf{H}_e) \mid \mathbf{\Theta} = \theta] = u_e(\theta, \tilde{h}_e)$ and $\mathbb{E}[u_\ell(\mathbf{H}_\ell)] = u_\ell(\tilde{h}_\ell)$. Moreover, since (u_e, u_ℓ) is a reduced form of G, it follows that $\Pi_e^s(\theta) = u_e(\theta, \tilde{h}_e)$ and $\Pi_\ell^s = u_\ell(\tilde{h}_\ell)$ are silent payoffs. (See Lemma B.12.) Thus, the expert's expected utility at state θ is $\mathcal{U}_e(\rho \mid \theta, \{\emptyset\}, \beta') = \Pi_e^s(\theta) + Y_e(\theta, \mathcal{M}', \sigma)$, and ℓ's expected utility is $\mathcal{U}_\ell(\sigma \mid \{\emptyset\}, \beta_\ell) = \Pi_\ell^s + Y_\ell(\mathcal{M}', \sigma)$. Since (σ, β) is equivalent to (ρ, β'), it follows that (σ, β) is also concealing. ■

B.5 Proofs of Section 8.1

Proof of Proposition 8.1. Write $\varsigma_\ell : H_\ell \to \Delta(A_\ell)$ for the mapping in which $\varsigma_\ell(h_\ell)$ assigns probability one to the action $a_\ell = \sum_{\theta \in \Theta} \theta \cdot h_\ell^1(\theta) \cdot 1_N$, where $1_N \in \mathbb{R}^N$ is the N-dimensional vector of ones.

Fix an induced Bayesian game $\mathrm{BG}(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$ and let (δ_e, δ_ℓ) be the hierarchy mappings induced by β. Fix $T_\ell \in \mathcal{T}_\ell$ and observe that ℓ's payoff from choosing $a_\ell \in A_\ell$ at first-order belief h_ℓ^1 is $\sum_{\theta \in \Theta} -\|\theta \cdot 1 - a_\ell\|^2 \cdot h_\ell^1(\theta)$. Since this is a sum of squared Euclidean distances, the optimal action a_ℓ minimizes the expected squared distance. This is achieved when $a_\ell = \sum_{\theta \in \Theta} \theta \cdot h_\ell^1(\theta) \cdot 1_N$, which is the layman's expectation of the state applied uniformly across all tasks. Therefore, $\hat{\sigma}$ is a BE of $\mathrm{BG}(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$ if and only if for each $T_\ell \in \mathcal{T}_\ell$, $\hat{\sigma}_\ell(T_\ell) = \varsigma_\ell(\delta_\ell(T_\ell))$.

Fix $T_\ell \in \mathcal{T}_\ell$ and write $h_\ell = \delta_\ell(T_\ell)$. Note that for each $(h_e, h_\ell) \in H$, $\hat{\sigma}_\ell(h_\ell) =$

$f_\ell^1(h_e, h_\ell) \cdot 1_N$. Hence,

$$\begin{aligned}\Pi_\ell(T_\ell \mid \hat\sigma, \beta) &= \sum_{\theta\in\Theta} -\|\theta \cdot 1_N - \hat\sigma_\ell(T_\ell)\|^2 \cdot \operatorname{marg}_\Theta \beta_\ell(T_\ell)(\theta) \\ &= \int_{\Theta\times H_e} -\|\theta \cdot 1_N - f_\ell^1(h_e, h_\ell) \cdot 1_N\|^2 \; dh_\ell^\infty.\end{aligned}$$

Now fix $(\theta, T_e) \in \Theta \times \mathcal{T}_e$ and write $h_e = \delta_e(\theta, T_e)$. Observe that

$$\begin{aligned}\Pi_e(\theta, T_e \mid \hat\sigma, \beta) &= \omega \sum_{T_\ell\in\mathcal{T}_\ell} -\|b_1 + \theta \cdot b_2 - \hat\sigma_\ell(T_\ell)\|^2 \beta_e(\theta, T_e)(T_\ell) \\ &= \omega \int_{H_\ell} -\|b_1 + \theta \cdot b_2 - f_\ell^1(h_e, h_\ell) \cdot 1_N\|^2 \; dh_e^\infty.\end{aligned}$$

Therefore, $\mathrm{RF} = \{(\varsigma_e, \varsigma_\ell)\}$ is a reduced form for G. ■

Proof of Proposition 8.3. Fix an information structure. i.e. a set of public signals S and a mapping $\chi : \Theta \to \mathcal{S}$. The prior μ and the information structure defines a probability space over $\Theta \times \mathcal{S}$. Write $\mathcal{F}_\ell = \{\Theta \times \{s\} : s \in \mathcal{S}\}$ is ℓ's information partition. Write $\mathbf{\Theta} : \Theta \times \mathcal{S} \to \Theta$ 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. Write $\overline{\theta} = \mathbb{E}[\mathbf{\Theta}]$. To show the result we first show some identities. Note that

$$\mathbb{E}[\mathbf{F}] = \mathbb{E}\big[\mathbb{E}[\mathbf{\Theta} \mid \mathcal{F}_\ell]\big] = \mathbb{E}[\mathbf{\Theta}] = \overline{\theta}, \tag{11}$$

where the second equality follows from the law of iterated expectations. Since $\mathbf{F}$ is $\mathcal{F}_\ell$ measurable,

$$\mathbb{E}[\mathbf{F}\,\mathbf{\Theta}] = \mathbb{E}\big[\mathbb{E}[\mathbf{F}\,\mathbf{\Theta} \mid \mathcal{F}_\ell]\big] = \mathbb{E}\big[\mathbf{F}\,\mathbb{E}[\mathbf{\Theta} \mid \mathcal{F}_\ell]\big] = \mathbb{E}[\mathbf{F}^2], \tag{12}$$

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 $\overline{\theta}^2 \le \mathbb{E}[\mathbf{F}^2] \le \mathbb{E}[\mathbf{\Theta}^2]$. where the first inequality follows from Equation (11) and the second from the fact that

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

Given the payoff $\pi_\ell(\theta, a_\ell) = -\|a_\ell - \theta \cdot 1^N\|^2$, the layman's expected payoffs are

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

Given $\pi_e(\theta, a_\ell) = -\omega \cdot \|a_\ell - b_1 - \theta \cdot b_2\|^2$, the expert's expected payoffs are

$$\begin{aligned}\mathbb{E}[-\omega\|\mathbf{F} \cdot 1^N - b_1 - \mathbf{\Theta} \cdot b_2\|^2] &= \mathbb{E}[-\omega\|\mathbf{F} \cdot 1^N - b_1 - \mathbf{\Theta} \cdot b_2\|^2] \\ &= \mathbb{E}[-\omega(\|\mathbf{F} \cdot 1^N\|^2 - 2\langle \mathbf{F} \cdot 1^N, b_1 + \mathbf{\Theta} \cdot b_2\rangle + \|b_1 + \mathbf{\Theta} \cdot b_2\|^2)] \\ &= \mathbb{E}[-\omega\|\mathbf{F} \cdot 1^N\|^2 + 2\omega\langle \mathbf{F} \cdot 1^N, b_1\rangle + 2\omega\langle \mathbf{F} \cdot 1^N, \mathbf{\Theta} \cdot b_2\rangle - \omega\|b_1 + \mathbf{\Theta} \cdot b_2\|^2] \\ &= \mathbb{E}[-\omega N\mathbf{F}^2 + 2\omega\langle 1^N, b_1\rangle\mathbf{F} + 2\omega\langle 1^N, b_2\rangle\mathbf{F}\mathbf{\Theta} - \omega\|b_1 + \mathbf{\Theta} \cdot b_2\|^2].\end{aligned}$$

So, adding these terms, the ex ante total welfare is given by

$$\begin{aligned}\mathrm{TW} &= \mathbb{E}[-N\mathbf{\Theta}^2 + 2N\mathbf{\Theta}\mathbf{F} - N\mathbf{F}^2] + \mathbb{E}[-\omega N\mathbf{F}^2 + 2\,\omega\langle 1^N, b_1\rangle\mathbf{F} + 2\omega\langle 1^N, b_2\rangle\mathbf{F}\mathbf{\Theta} - \omega\|b_1 + \mathbf{\Theta} \cdot b_2\|^2] \\ &= \mathbb{E}[-N\mathbf{\Theta}^2 + 2N\mathbf{F}^2 - N\mathbf{F}^2 - \omega N\mathbf{F}^2 + 2\omega\langle 1^N, b_1\rangle\overline{\theta} + 2\omega\langle 1^N, b_2\rangle\mathbf{F}^2 - \omega\|b_1 + \mathbf{\Theta} \cdot b_2\|^2] \\ &= 2\,\omega\langle 1^N, b_1\rangle\,\overline{\theta} + \mathbb{E}[-N\mathbf{\Theta}^2 - \omega\|b_1 + \mathbf{\Theta} \cdot b_2\|^2] + \big(N - \omega N + 2\,\omega\langle 1^N, b_2\rangle\big) \cdot \mathbb{E}[\mathbf{F}^2].\end{aligned}$$

where the second equality follows from Equations (11) and (12). Notice that the first term $2\omega N\langle 1^N, b_1\rangle\overline{\theta} + \mathbb{E}[-N\mathbf{\Theta}^2 - \omega\|b_1 + \mathbf{\Theta} \cdot b_2\|^2]$ does not depend on the information structure used. Hence, the welfare maximizing information structure maximizes $\big(N - \omega\, N + 2\,\omega\langle 1^N, b_2\rangle\big) \cdot \mathbb{E}[\mathbf{F}^2]$.

Recall that $\mathbb{E}[\mathbf{\Theta}]^2 \leq \mathbb{E}[\mathbf{F}^2] \leq \mathbb{E}[\mathbf{\Theta}^2]$. Hence, if $1 \leq \omega(1 - \frac{2}{N}\langle 1_N, b_2\rangle)$, the information structure that maximizes welfare satisfies $\mathbb{E}[\mathbf{F}^2] = \mathbb{E}[\mathbf{\Theta}]^2$. Observe, revealing no information implies $\mathbf{F} = \mathbb{E}[\mathbf{\Theta}]$ almost surely. So, revealing no information maximizes welfare. By contrast, if $1 \geq \omega(1 - \frac{2}{N}\langle 1_N, b_2\rangle)$ the information structure that maximizes welfare satisfies $\mathbb{E}[\mathbf{F}^2] = \mathbb{E}[\mathbf{\Theta}^2]$. As a result, full revelation of the state maximizes total welfare. ■

B.6 Proofs of Section 8.2

Proof of Proposition 8.4. Fix an induced Bayesian game $\mathrm{BG}(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$. Strict concavity of π_i implies that each best response is single valued and there is no loss to analyze only pure strategy profiles [Zimper, 2006]. Let $(\hat{\sigma}_e, \hat{\sigma}_\ell)$ be a (pure) strategy profile of $\mathrm{BG}(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$. For notational convenience, write $\hat{\sigma}_i(\cdot) \to A_i$ for the pure strategy of i. By concavity, $(\hat{\sigma}_e, \hat{\sigma}_\ell)$ is a Bayesian equilibrium if and only if it satisfies the following first-order conditions:

$$\hat{\sigma}_e(\theta, T_e) = \theta + \alpha \cdot \sum_{T'_\ell \in \mathcal{T}_\ell} \hat{\sigma}_\ell(T'_\ell) \cdot \beta_e(\theta, T_e)(T'_\ell), \quad \text{and} \tag{13}$$

$$\hat{\sigma}_\ell(T_\ell) = \sum_{(\theta', T'_e) \in \Theta \times \mathcal{T}_e} (\theta' + \alpha \cdot \hat{\sigma}_e(\theta', T'_e)) \cdot \beta_\ell(T_\ell)(\theta', T'_e). \tag{14}$$

Write $\varsigma_e : \Theta \times H_e \to A_e$ and $\varsigma_\ell : H_\ell \to A_\ell$ as

$$\varsigma_e(\theta, h_e) := \theta + (1+\alpha) \cdot \int_{H_\ell} \left(\sum_{k=1}^{\infty} \alpha^{2k-1} f_\ell^{2k-1}(h_e, h_\ell) \right) dh_e^\infty, \text{ and}$$

$$\varsigma_\ell(h_\ell) := (1+\alpha) \int_{\Theta \times H_e} \left(\theta + \sum_{k=1}^{\infty} \alpha^{2k} f_e^{2k}(h_e, h_\ell) \right) dh_\ell^\infty.$$

Write $(\sigma_e^*, \sigma_\ell^*)$ for the strategy profile in $\mathrm{BG}(\mathcal{M}, u_e, u_\ell)$ such that $\sigma_e(\theta, T_e) = \varsigma_e(\theta, \delta_e(\theta, T_e))$ and $\sigma_\ell(T_\ell) = \varsigma_\ell(\delta_\ell(T_\ell))$ We show that $(\sigma_e^*, \sigma_\ell^*)$ is a BE of $\mathrm{BG}(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$. To see this, note that

$$\begin{aligned}
\varsigma_e(\theta, h_e) &= \theta + (1+\alpha) \cdot \int_{H_\ell} \left(\sum_{k=1}^{\infty} \alpha^{2k-1} f_\ell^{2k-1}(h_e, h_\ell) \right) dh_e^\infty \\
&= \theta + \alpha \cdot \int_{H_\ell} \left((1-\alpha) \cdot \sum_{k=1}^{\infty} \alpha^{2k} f_e^{2k}(h_e, h_\ell) \right) dh_e^\infty \\
&= \theta + \alpha \cdot \int_{H_\ell} \left((1-\alpha) \int_{\Theta \times H_e} \left(\theta + \sum_{k=1}^{\infty} \alpha^{2k-2} f_\ell^{2k-1}(h_e, h_\ell) \right) dh_\ell^\infty \right) dh_e^\infty \\
&= \theta + \alpha \int_{H_\ell} \varsigma_\ell(h_\ell) \; dh_e^\infty.
\end{aligned}$$

Hence, if $h_e = \delta_e(\theta, T_e)$, then,

$$\sigma_e^*(\theta, T_e) = \varsigma_e(\theta, h_e) = \theta + \alpha \int_{H_\ell} \varsigma_\ell(h_\ell) \; dh_e^\infty = \theta + \alpha \sum_{T'_\ell \in \mathcal{T}_\ell} \sigma_\ell^*(T'_\ell) \cdot \beta_e(\theta, T_e)(T'_\ell).$$

So, $(\sigma_e^*, \sigma_\ell^*)$ satisfies Equation (13). Moreover,

$$\begin{aligned}
\varsigma_\ell(h_\ell) &= (1+\alpha)\int_{\Theta\times H_e}\left(\theta + \sum_{k=1}^{\infty}\alpha^{2k}\, f_e^{2k}(h_e,h_\ell)\right) dh_\ell^\infty \\
&= \int_{\Theta\times H_e}\left((1+\alpha)\theta + (1+\alpha)\sum_{k=1}^{\infty}\alpha^{2k}\, f_e^{2k}(h_e,h_\ell)\right) dh_\ell^\infty \\
&= \int_{\Theta\times H_e}\left(\theta + \alpha\left(\theta + (1+\alpha)\sum_{k=1}^{\infty}\alpha^{2k-1}\, f_e^{2k}(h_e,h_\ell)\right)\right) dh_\ell^\infty \\
&= \int_{\Theta\times H_e}\theta + \alpha\cdot\varsigma_e(\theta,h_e)\; dh_\ell^\infty.
\end{aligned}$$

Hence, if $h_\ell = \delta_\ell(T_\ell)$, then

$$\sigma_\ell^*(T_\ell) = \varsigma_\ell(h_\ell) = \int_{\Theta\times H_e}\theta + \alpha\cdot\varsigma_e(\theta,h_e)\; dh_\ell^\infty = \sum_{(\theta',T_e')\in\Theta\times\mathcal{T}_e}\big(\theta' + \alpha\cdot\sigma_e^*(\theta',T_e')\big)\cdot\beta_\ell(T_\ell)(\theta',T_e').$$

So, Equation (14) holds. Thus, $(\sigma_e^*, \sigma_\ell^*)$ is a BE of $\mathrm{BG}(\mathcal{T}_e,\mathcal{T}_\ell,\beta)$.

We now show that $(\sigma_e^*, \sigma_\ell^*)$ is the unique Bayesian equilibrium of $\mathrm{BG}(\mathcal{T}_e,\mathcal{T}_\ell,\beta)$. Fix a Bayesian equilibrium $(\hat\sigma_e,\hat\sigma_\ell)$ of the $\mathrm{BG}(\mathcal{T}_e,\mathcal{T}_\ell,\beta)$. Observe, since $(\hat\sigma_e,\hat\sigma_\ell)$ and $(\sigma_e^*, \sigma_\ell^*)$ satisfy Equation (13),

$$\begin{aligned}
\sup_{\theta,T_e}|\hat\sigma_e(\theta,T_e) - \sigma_e^*(\theta,T_e)| &= \left|\alpha\sum_{T_\ell\in\mathcal{T}_\ell}\big(\hat\sigma_\ell(T_\ell) - \sigma_\ell^*(T_\ell)\big)\cdot\beta_e(\theta,T_e)(T_\ell)\right| \\
&\le |\alpha|\cdot\sup_{T_\ell\in\mathcal{T}_\ell}|\hat\sigma_\ell(T_\ell) - \sigma_\ell^*(T_\ell)|.
\end{aligned}$$

Moreover, an analogous argument shows that

$$\sup_{T_\ell}|\hat\sigma_\ell(T_\ell) - \sigma_e^*(T_\ell)| \le |\alpha|\cdot\sup_{(\theta,T_e)\in\Theta\times\mathcal{T}_e}|\hat\sigma_e(\theta,T_\ell) - \sigma_e^*(\theta,T_\ell)|.$$

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

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

Thus, $(\sigma_e^*, \sigma_\ell^*) = (\hat\sigma_e,\hat\sigma_\ell)$. Therefore, $(\sigma_e^*, \sigma_\ell^*)$ is the unique BE of $\mathrm{BG}(\mathcal{T}_e,\mathcal{T}_\ell,\beta)$.

To show that $\{(u_e,u_\ell)\}$ is a reduced form representation for G, it suffices to show that $(\sigma_e^*, \sigma_\ell^*)$ induces the belief-dependent utilities (u_e,u_ℓ). Fix $(\theta,T_e)\in\Theta\times\mathcal{T}_e$. If

$h_e = \delta_e(\theta, T_e)$, then

$$\begin{aligned}
\varsigma_e(\theta, h_e) &= \theta + (1+\alpha) \int_{H_\ell} \left(\sum_{k=1}^{\infty} \alpha^{2k-1}\, f_\ell^{2k-1}(h_e, h_\ell) \right) dh_e^\infty \\
&= \theta + \alpha \int_{H_\ell} (1+\alpha) \left(\sum_{k=1}^{\infty} \alpha^{2k-1}\, f_\ell^{2k-1}(h_e, h_\ell) \right)\, dh_e^\infty \\
&= \theta + \alpha \int_{H_\ell} f^\alpha(h_e, h_\ell)\, dh_e^\infty .
\end{aligned}$$

Thus, $u_e(\theta, h_e) = \frac{1}{2}\varsigma_e(\theta, h_e)^2$. So, by Equation (13)

$$\begin{aligned}
\Pi_e(\theta, T_e \mid \sigma^*, \beta) &= \sigma_e^*(\theta, T_e) \left(\theta - \tfrac{1}{2}\sigma_e^*(\theta, T_e) + \alpha \textstyle\sum_{T_\ell \in \mathcal{T}_\ell} \sigma_\ell^*(T_\ell) \cdot \beta_e(\theta, T_e)(T_\ell) \right) \\
&= \sigma_e^*(\theta, T_e) \left(\tfrac{1}{2}\sigma_e^*(\theta, T_e) \right) \\
&= \tfrac{1}{2}\varsigma_e(\theta, h_e)^2 \qquad (15)
\end{aligned}$$

Therefore, $\Pi_e(\theta, T_e \mid \sigma^*, \beta) = u_e(\theta, h_e)$. Now fix $T_\ell \in \mathcal{T}_\ell$ and write $h_\ell = \delta_\ell(T_\ell)$. Recall that f_ℓ^{2k+1} does not depend on h_e. Hence, for each $h_\ell \in H_\ell$,

$$\begin{aligned}
\varsigma_\ell(h_\ell) &= (1+\alpha) \int_{\Theta \times H_e} \left(\theta + \sum_{k=1}^{\infty} \alpha^{2k}\, f_e^{2k}(h_e', h_\ell) \right) dh_\ell^\infty \\
&= (1+\alpha) \left(f_\ell^1(h_e, h_\ell) + \sum_{k=1}^{\infty} \alpha^{2k} f_\ell^{2k+1}(h_e, h_\ell) \right).
\end{aligned}$$

Observe, this implies that

$$\varsigma_\ell(h_\ell) = (1+\alpha) \int_{\Theta \times H_e} \left(\sum_{k=1}^{\infty} \alpha^{2k}\, f_\ell^{2k-1}(h_e, h_\ell) \right) dh_\ell^\infty = (1+\alpha) \int_{\Theta \times H_e} f^\alpha(h_e, h_\ell)\, dh_\ell^\infty .$$

So, by Equation (14),

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

Therefore, $\Pi_\ell(T_\ell \mid \tilde{\sigma}, \beta) = u_\ell(h_\ell)$, as desired. ■

Proposition B.1. *Consider the game G of Section 8.2. 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 8.5 to compute the agents' ex ante utility under full revelation of the state and no information sharing.

Write $\overline{\theta} = \mathbb{E}[\mathbf{\Theta}]$ and $d = (1+\alpha)\sum_{k=1}^{\infty} \alpha^{2k} = (1+\alpha)^{-1}$. Note that $1 + \alpha d = d$. So, if h has degenerate belief for θ, then $f(h) = d\theta$. Moreover, $f(\widetilde{h}) = d\overline{\theta}$. Write FR_i (resp. NI_i) for the expected utility under full revelation (resp. no information) for agent i. Observe, the ex ante expert's benefit from information sharing is

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

where the second equality follows from the fact that $\big(\mathbb{E}[\mathbf{\Theta} + \alpha d \mathbf{\Theta}]\big)^2 = \big(\mathbb{E}[\mathbf{\Theta} + \alpha d \overline{\theta}]\big)^2$, 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 $\alpha \geq 0$.

Likewise the ex ante layman's benefit from information sharing is

$$\mathrm{FR}_\ell - \mathrm{NI}_\ell = \mathbb{E}[(d\mathbf{\Theta})^2] - \mathbb{E}[(d\overline{\theta})^2] = \mathrm{Var}[(d\mathbf{\Theta})^2] - \mathrm{Var}[(d\overline{\theta})^2] = d^2 \mathrm{Var}[\mathbf{\Theta}^2],$$

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

Notice that this value is always positive. Hence, regardless of the value α, the layman gets positive benefits from information sharing. Note, revealing the state 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 $\alpha \in (1-\sqrt{2}, 1)$. ∎

B.7 Proofs of Section 8.3

Proof of Proposition 8.6. First, we show (i). Notice that u_e is supermodular at common degenerate beliefs if and only if g satisfies weakly increasing differences on $\Theta \times \{0, 1\}$. Hence, the result follows from Theorem 5.1.

Now we show (ii). By Theorem 6.1, it suffices to show that (u_e, u_ℓ) is statistically submodular. Fix a feasible probability space $\mathcal{P}$ that impacts (u_e, u_ℓ). Note f^1_ℓ is acute (See Lemma B.7). Moreover, Since g has strict decreasing differences, the statistic $\hat{f}(h) := g\big(0, f^1_\ell(h)\big) - g\big(1, f^1_\ell(h)\big)$ is acute. (See Lemma B.8.) Write $\mathbf{F}^1_\ell$ (resp. $\hat{\mathbf{F}}$) for the random variable associated with f^1_ℓ (resp. $\hat{f}$). Observe, for each θ,

$$\mathbb{E}\big[u_e(1, \mathbf{H}_e) - u_e(0, \mathbf{H}_e)|\mathbf{\Theta} = \theta\big] = \mathbb{E}\big[g(1, \mathbf{F}^1_\ell) - g(0, \mathbf{F}^1_\ell)|\mathbf{\Theta} = \theta\big] = -\mathbb{E}\big[\hat{\mathbf{F}} \mid \mathbf{\Theta} = \theta\big]. \quad (16)$$

Note, Since $\mathcal{P}$ impacts (u_e, u_ℓ) and $\hat{f}$ is essential for (u_e, u_ℓ), it follows that $\mathbb{P}\big[\hat{\mathbf{F}} = \hat{f}(\tilde{h}_e, \tilde{h}_\ell)\big] < 1$. Since $\hat{f}$ is acute, it follows that $\mathrm{Cov}\big[\mathbf{\Theta}, \hat{\mathbf{F}}\big] > 0$. This implies that $\mathbb{E}[\hat{\mathbf{F}} \mid \mathbf{\Theta} = 1] > \mathbb{E}[\hat{\mathbf{F}} \mid \mathbf{\Theta} = 0]$. So, by Equation (16),

$$\mathbb{E}\big[u_e(1, \mathbf{H}_e) - u_e(0, \mathbf{H}_e) \mid \mathbf{\Theta} = 1\big] < \mathbb{E}\big[u_e\big(1, \mathbf{H}_e\big) - u_e\big(0, \mathbf{H}_e\big) \mid \mathbf{\Theta} = 0\big].$$

Thus, (u_e, u_ℓ) is statistically submodular. ∎

B.8 Proofs of Section 9

Proof of Theorem 9.1. *If.* Fix $(u_e, u_\ell) \in \mathrm{RF}$ so that $g(\theta, \theta') = u_e(\theta, \eta_e(\theta'))$ satisfies cyclical monotonicity. It suffices to show that u_e is fully revealing. Observe, by Lemma B.2, there is some mapping $z : \Theta \to \mathbb{R}$ such that $g(\theta, \theta) + z(\theta) \geq g(\theta, \theta') + z(\theta')$ for each $\theta, \theta' \in \Theta$. By using an analogous argument as the proof of Theorem 5.1, it follows that u_e is fully 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 $(u_e, u_\ell) \in \mathrm{RF}$, a mechanism $\mathcal{M}'$, and a BE (ρ, β') of the psychological game $(\mathcal{M}', (u_e, u_\ell))$ such that (ρ, β) is equivalent to (σ, β).

Write $g(\theta, \theta') := u_e(\theta, \eta_e(\theta'))$. We show that g satisfies cyclical monotonicity. Let $\mathcal{M}^d$ be the extended direct mechanism induced by $\mathcal{M}'$ and (ρ, β). Write m for the protocol of $\mathcal{M}^d$ and write $z(\theta) := \sum_{y_e \in Y_e} y_e \mathrm{marg}_{\,Y_e} m(\theta)(y_e)$ for the expected transfers given state θ. Observe that $\mathcal{V}_e(\theta, \theta'|\mathcal{M}^d) = g(\theta, \theta') + z(\theta')$ and $\mathcal{M}^d$ is BIC. (See Theorem 2 in Rivera Mora [2024].) Hence, for each $\theta, \theta' \in \Theta$, $g(\theta, \theta) + z(\theta) \geq g(\theta, \theta') + z(\theta')$. Therefore, g satisfies cyclical monotonicity. (See Lemma B.2.) ∎

Lemma B.13. *The game G from Example 9.3 is not perfectly revealing or concealing.*

Proof. Analogously to Proposition 8.2, the game G has a reduced form representation $\mathrm{RF} = \{(u_e, u_\ell)\}$ such that

$$u_e(\theta, h_e) = -\int\limits_{h_\ell \in H_\ell} \left(\mathrm{b}(\theta) - f_\ell^1(h_e, h_\ell)\right)^2 dh_e^\infty, \quad \text{and}$$
$$u_\ell(h_\ell) = -\int_{\Theta \times H_e} \left(\theta - f_\ell^1(h_e, h_\ell)\right)^2 dh_\ell^\infty.$$

Write $g(\theta, \theta') = u_e(\theta, h_e^{\theta'}) = -(\mathrm{b}(\theta) - \theta')^2$. Observe, if $\boldsymbol{\theta}$ is the cycle $(2, 3, 2)$, then, $\mathcal{L}(g, \boldsymbol{\theta}) = g(2, 2) - g(2, 3) + g(3, 3) - g(2, 3) = -2 < 0$. Thus, u_e does not satisfies cyclical monotonicity on degenerate beliefs. Thus, G is not perfectly revealing. (See Theorem 9.1.)

We now show that G is not concealing. Consider a mechanism $\mathcal{M}$ in which the expert selects one of two (public) cheap talk messages $\underline{m}$ and $\overline{m}$. Consider the strategy profile σ of $(\mathcal{M}, G)$ in which 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. Thus G is not concealing. ■