Adversarial Bilateral Information Design

Gabriel Ziegler*

June 26, 2020

Abstract

Information provision is a significant component of business-to-business interaction. Furthermore, the provision of information is often conducted bilaterally. This precludes the possibility of commitment to a grand information structure if there are multiple receivers. Consequently, in a strategic situation, each receiver needs to reason about what information other receivers get. Since the information provider does not know this reasoning process, a motivation for a robustness requirement arises: the provider seeks an information structure that performs well no matter how the receivers actually reason. In this paper, I provide a general method to study how to optimally provide information under these constraints. The main result is a representation theorem, which makes the problem tractable by reformulating it as a linear program in belief space. Furthermore, I provide novel bounds on the dependence among receivers' beliefs, which provide even more tractability in some special cases.

Keywords: bilateral contracting, information design, robust design, adversarial design, belief manipulation, belief distributions, Bayes plausibility, Fréchet-Hoeffding bounds, dependence bounds.
JEL Classification: C72, D82, D83, L86.

---

*Northwestern University, Department of Economics; 2211 Campus Drive, Evanston, IL 60208, USA; ziegler@u.northwestern.edu.

I am deeply indebted to my advisor, Marciano Siniscalchi, for continuous support and endless discussions during various stages of the project. Furthermore, special thanks are due to Eddie Dekel, Alessandro Pavan, and Peio Zuazo-Garin for insightful comments which greatly improved the paper. Matteo Camboni, Henrique Brasiliense de Castro Pires, Laura Doval, Piotr Dworczak, Andres Espitia, Ryan Lee, Matthew Leisten, Kota Murayama, Lorenzo Stanca, Bruno Strulovici, and Quitze Valenzuela-Stookey provided useful thoughts and comments. I want to thank seminar participants at various places. The usual disclaimer applies.

1 INTRODUCTION

Information provision and bilateral contracting are ubiquitous in today's economy. For example, contract research organizations (CROs) provide information to downstream firms (called sponsors), which are typically pharmaceutical or biotechnology companies. Sponsors, such as pharmaceutical companies, engage with CROs to outsource part of the drug development. If an agreement is reached, the contract specifies which trials the CRO will conduct for the given sponsor, but not which trials are performed for other sponsors. This is a typical example of bilateral contracting: the contract is contingent only on events that can be verified by *both* of the involved parties. Furthermore, it is common for sponsors of the same CRO to be direct competitors.

Leaving aside details of specific industries, three considerations are crucial for any information provision organization determining what information to provide to clients. First, the provider effectively *commits* to deliver specific information to a given client in a contract. For example, a contract will specify exactly which medical tests will be conducted. Second, the *bilateral* nature of *contracting* excludes commitment to a grand information structure shared with all clients. That is, a contract will only state which tests will be conducted for a specific sponsor and will not state which tests will be performed for other sponsors.[1] Third, the receivers' use of the information is determined within an interactive setting. Therefore, a receiver faces *strategic uncertainty* and needs to reason about what information other receivers get. Crucially, the details of this reasoning process are usually unknown to the information provider. For example, the decision for one sponsor to conduct further research on a drug depends on whether the sponsor believes its competitors are also developing a competing drug and, if so, what information the sponsor believes its competitors are receiving.

---

[1] Contracts do not specify such details for several reasons. First, CROs have reputational concerns. If CROs disclose which trials they were conducting for a sponsor's competitor, the CRO might reveal the competitor's private information, undermining the CRO's relationship with the competitor. Second, a contract that is contingent on every trial conducted for every sponsor is complex and lacks enforceability. These reasons are broadly applicable and do not only affect CROs. In particular, the second point was raised by McAfee and Schwartz (1994) regarding any supplier that deals with multiple downstream firms.

In this paper, I provide a general, yet tractable, method for examining how an information provider determines which information to supply bilaterally to multiple receivers, taking into consideration each of the three aspects outlines above. In particular, motivated by the severity of strategic uncertainty, I take an adversarial approach which ensures robustness to details of receivers' strategic reasoning and is tractable. That is, the information provided to one receiver is required to be optimal for the designer no matter how that receiver thinks about the information other receivers may get. The adversarial approach adopted here ensures that the supplied information is optimal even if nature "chooses" the receiver's reasoning that is least advantageous to the provider.

First, I formalize the issue of robustness to the receivers' reasoning. From a CRO's point of view, I provide a precise answer to the following question: given that a pharmaceutical sponsor gets some information about their drug, how does the pharmaceutical sponsor decide whether to bring the drug to the market or, for example, drop the project altogether? As noted above, sponsors face strategic uncertainty because they do not know what information their competitors have access to. This section's primary contribution is to provide a solution concept that captures this kind of uncertainty. The key insight is that the reasoning about the competitors' information can be sidestepped: to form a best-reply the competitors' information is not relevant, but only the beliefs about the state of nature and the competitor's action matter. For this, a characterization of "rational" competitor's action for any information structure is needed: all *belief-free rationalizable* actions. Furthermore, I demonstrate that this solution concept depends only upon players' first-order beliefs about the payoff state. For a CRO, this means that the solution concept depends only on the information a sponsor receives about their own drug, but not on how a sponsor thinks about the information its competitors have.

Second, I contribute to the foundations of information design with multiple receivers. Mathevet et al. (2020, p.2) describe information design as "an exercise in belief manipulation;" therefore, it is crucial to characterize which beliefs can be induced by a designer. If there is only one receiver, it is well known that there is only one restriction on the distribution of beliefs about the state of nature. The average belief under this distribution is equal to the prior—a requirement deemed *Bayes plausibil-*

ity by Kamenica and Gentzkow (2011). This paper extends this characterization to multiple receivers (cf. Theorem 1). Furthermore, I provide necessary bounds on the dependence of beliefs if there are two receivers (cf. Proposition 2). These bounds are reminiscent of, but usually tighter than, the Fréchet-Hoeffding bounds known from copulas in probability theory and statistics.[2] Moreover, these bounds are novel not only for information-design and the economics literature more generally, but—to the best of my knowledge—to probability theory as well and provide tractability because they are related to the supermodular stochastic ordering. Even more tractability is gained when more assumptions are put on the primitives, which in particular include supermodular games. I illustrate this in a stylized version of the problem faced by a CRO.

The remainder of the paper is organized as follows: the next subsections elaborate on related literature and provide the setting for the stylized model of a CRO, which will be used as a running example throughout the paper. Section 2 develops the solution concept.[3] The main representation theorem for the general design problem is formalized in Section 3. Section 4 studies the case of oure persuasion, which includes the derivation of the belief-dependence bounds and the solution to the CRO model. In Section 5, I discuss some extensions and highlight issues related to interpretations of the model. All proofs can be found in the Appendix.

1.1 Related Literature

This paper is related to several strands of the literature: a solution concept capturing a notion of robustness, general information design, and adversarial and bilateral design. In this section, I discuss these three strands in detail.

1.1.1 Robust Solution Concepts

Harsanyi's (1968) theory of games with incomplete information is partially motivated by the possibility that players' information structures may not be common knowledge. The solution concept I develop in this paper is directly inspired by the literature on informational robustness which later formalized Harsanyi's insights about robustness.

---

[2] A standard reference for copulas including the Fréchet-Hoeffding bounds is Nelsen (2006).

[3] A foundation of the solution concept is given in Subsection B.1.

Early pioneers in this area include Aumann (1987), Brandenburger and Dekel (1987), and Forges (1993, 2006). Bergemann and Morris (2013, 2016) recently exploited the full power of informational robustness to provide robust predictions in economic environments with uncertainty.[4] Within this subset of the literature, my work is closest to that of Bergemann and Morris (2017). Their paper is concerned with robustness over all information structures from the perspective of an outside observer, while this paper instead focuses on the notion of robustness from a player's perspective. This allows sharper predictions because a player considers parts of the information structure that an outside observer does not know. In this vein, a solution concept similar to mine is used by Börgers and Li (2019) to define strategic simplicity. Like the solution concept in this paper, Börgers and Li's solution concept depends only on first-order beliefs.[5] However, these authors do not assume common belief in rationality and also do not provide a foundation for their solution concept.

As discussed in Subsection 5.2, my solution concept can be given an epistemic foundation by simply modifying the arguments introduced by Battigalli and Siniscalchi (2007) and developed further in Battigalli et al. (2011). In each of these papers players have symmetric knowledge about the information structure. Either the full information structure is commonly known, or no (common) knowledge about the information structure is assumed at all. In my case, there is no assumption about common knowledge of the information structure, but each player knows *her* own information structure.

1.1.2 Information Design

The literature on information design originated from contributions of Calzolari and Pavan (2006), Bergemann and Pesendorfer (2007), Brocas and Carrillo (2007), and Eső and Szentes (2007). Since then the literature has grown rapdily. The interested reader is referred to two recent reviews by Bergemann and Morris (2019) and Kamenica (2019). I highlight papers here that are more closely related to this one,

---

[4] Other papers dealing with related ideas about robustness include Battigalli and Siniscalchi (2003), Dekel et al. (2007), Liu (2015), Tang (2015), and Germano and Zuazo-Garin (2017).

[5] In robust mechanism design, Artemov et al. (2013) study robust mechanism design when the designer knows that the first-order beliefs belong to a specific set of beliefs. In contrast to my appraoch, the (sets of) first-order beliefs are common knowledge among the players in their setting. A similar approach was considered by Ollár and Penta (2017).

which provide general methods to analyze information design as this paper does. The seminal paper pertaining to a single receiver is Kamenica and Gentzkow (2011) which illustrates the usefulness of the concavification approach for information design. Regarding multiple receivers, Taneva (2019) uses a Myersonian approach, exploiting a version of the revelation principle, which can be interpreted as a akin to partial implementation known from mechanism design.

The closest work on information design is the upcoming article by Mathevet et al. (2020). Like Taneva (2019), Mathevet et al. consider information design in cases when the designer has the power to commit to the provision of a grand information structure. However, for a given grand information structure, they allow for the case of adversarial equilibrium selection. Thus, their approach is reminiscent of full implementation in mechanism design. They show that attaining robustness to equilibrium selection requires constructing the full *hierarchy of beliefs* for each receiver.[6] My approach is complementary to theirs. In my setting, strategic uncertainty arises from the bilateral contracting environment which excludes commitment to a grand information structure. Therefore, in my case the designer is not only concerned about equilibrium selection, but also about strategic uncertainty. My proposed solution concept reflects this more general robustness concern. In addition, I show that my robust solution concept depends only on induced first-order beliefs. Therefore, it is not necessary to induce a full hierarchy of beliefs, but it suffices to look at first-order beliefs only. Thus, the approach I propose is closer in spirit to Kamenica and Gentzkow (2011): since they consider a single receiver, by definition only first-order beliefs matter. However, in the present paper there are multiple receivers and therefore a new characterization in terms of distributions of first-order beliefs is needed. This is the main result of Section 3.

Recent and independent work by Arieli, Babichenko, Sandomirskiy, and Tamuz (2020) studies the question of which distributions over (first-order) beliefs can be induced by information structures in the case of binary states of nature.[7] They provide a full characterization of these distributions for two receivers and extend this

---

[6] Similar to the full implementation literature the revelation does not apply in Mathevet et al.'s (2020) setting either.

[7] Arieli et al. also consider applications to social persuasion, i.e. persuading multiple receivers in a non-strategic setting.

characterization to multiple receivers in spirit of No Trade Theorems. In Section 4, I provide bounds on the dependence structure across two receivers for these distributions, which are necessary, but not sufficient. Under more stringent assumptions (which include the CRO model of Subsection 1.2) my bounds become equivalent to the conditions of Arieli et al. (2020) and are therefore also sufficient.

1.1.3 Adversarial and Bilateral Design

A few recent studies employ an adversarial approach to information design:[8] Carroll (2016), Goldstein and Huang (2016), Inostroza and Pavan (2018), and Hoshino (2019).[9] All apply the adversarial selection for a solution concept that relies on a grand information structure. In this paper the adversarial selection is more severe because of the additional robustness coming from the bilateral contracting environment. Bilateral information design with or without adversarial robustness is, to the best of my knowledge, new to this paper.

In a recent review, Carroll (2019) discusses adversarial selection aspects in mechanism design. Bilateral contracting has a long history in economics and has been studied extensively in industrial organization.[10] The relevant paper from this body of literature is Dequiedt and Martimort (2015). Dequiedt and Martimort examine bilateral mechanism design when the designer cannot commit to a grand mechanism. My paper shares the motivation for analyzing a setting with limited commitment with Dequiedt and Martimort. They overcome the limited commitment by imposing appropriate ex-post incentive constraints on side of the principal. In equilibrium, these ex-post constraints determine all beliefs of the agents including how they think about other agents' contracts. My approach resolves the limited-commitment issue in a different way. In my model, the designer does not assume that all beliefs are in equilibrium and therefore needs consider the reasoning of the receivers. By taking an adversarial approach, the designer circumvents these issues and seeks an information structure that is robust to the reasoning of the receivers.[11]

---

[8] That is, in addition to Mathevet et al. (2020) as mentioned above.

[9] I thank Nageeb Ali for making me aware of Hoshino's paper.

[10] The interested reader is referred to two handbook chapters: Bresnahan and Levin (2012) and Segal and Whinston (2012).

[11] In this sense, the literature on mechanism design without or with limited commitment is also related.

1.2 Leading Example: A Stylized CRO Model[12]

Consider a situation where a CRO conducts medical trials for two pharmaceutical companies called Pfizr (P) and Novarty (N).[13] Both work on developing similar breast cancer drugs. For simplicity, suppose that each drug could be either effective, or ineffective, and one drug is effective if and only if the other drug is effective. Thus, there are two states of nature, i.e. $\Theta = \{0, 1\}$ representing an ineffective drug and an effective drug, respectively. Furthermore, there are two possible actions the pharmaceutical companies can take: either conduct further research (R), or drop the project (D). Profits (i.e. payoffs) are such that, if firms knew the effectiveness of the drug, they would like to conduct research if and only if the drug is actually effective. However, if a pharmaceutical company decides to conduct further research, its payoff will be lower if the competitor also conducts further research. The reduction in payoffs could be caused by lower expected profits in the future, because the competitor's drug is likely to be on the market. The following payoff tables represent such a situation.

		Novarty	
		R	D
Pfizr	R	$1 \quad 1$	$2 \quad 0$
	D	$0 \quad 2$	$0 \quad 0$

$\theta = 1$ (effective)

Novarty	
R	D
$-2 \; -2$	$-1 \quad 0$
$0 \quad -1$	$0 \quad 0$

$\theta = 0$ (ineffective)

For any belief (about the state of nature) that puts probability greater than $^2/_3$ on the state in which the drug is effective ($\theta = 1$),[14] R is the dominant action. Similarly, for any belief less than $^1/_3$, the dominant action becomes D. For intermediate beliefs about θ, the best action depends beliefs about competitors' actions. Formal analysis in this paper shows that these predictions are exactly those which are robust to the reasoning about the information of the competitor. For example, if Pfizr assigns probability close to one to $\theta = 1$, then it does not matter what information Novarty gets and Pfizr should conduct further research. However, if the probability of $\theta = 1$ is $^1/_2$, Novarty's information matters. To see this, consider the Novarty medical trials,

---

[12] For readers familiar with information design, this section can be skipped. However, in the main analysis I will refer back to this example to illustrate some of the results.

[13] These companies and names are purely fictional.

[14] Henceforth, I will always associate *belief* with the probability of the state being $\theta = 1$.

conducted by a CRO, that reveal with high probability that the drug is ineffective. In such a case, Novarty will drop the project with high probability too. This implies that Pfizr should conduct more research (given their belief about θ). On the other hand, if the medical trials for Novarty are such that there is a high likelihood of revealing that the drug is effective, then Novarty is likely conducting research and Pfizr should drop the project (again given their belief about θ). Thus, Pfizrs beliefs about Novartys information matter. Therefore, if robustness is a concern, the CRO should take both actions, R and D, into account.

By providing information to the pharmaceutical companies, the CRO can effectively influence the actions taken by the pharmaceutical companies. For example, a natural assumption is that the CRO prefers further research rather than dropping the project, because of the likelihood that further research will include subsequent trials for the CRO to conduct. The goal of this paper is to provide a tractable method for solving for the optimal provision of information in such settings. In the remainder of this subsection, I highlight some specific information structures that are part of the CRO's choice set.

Suppose that both pharmaceutical companies have a prior belief that assigns probability $1/3$ to the drugs being effective. A trivial choice of the CRO would be to provide no information. In this case and similar to the explanation above, $\{R, D\}$ is the robust prediction for both receivers. Thus, under adversarial selection, the CRO expects both companies to drop the project, which would be the worst possible outcome from the CRO's perspective. Another possibility would be for the CRO to provide full information to each pharmaceutical company. In this case, each company will conduct further research if and only if their drug is effective. Overall, there will be further research (by both firms) with probability equal to the prior, i.e. slightly above 33%. However, the CRO could increase the probability of further research by providing information that does not fully reveal the effectiveness of the drugs.

For illustrative purposes, consider first a case where the CRO can actually commit to a grand information structure and therefore does not have to worry about what conjectures the receivers form about their competitor.[15] This problem can be ana-

---

[15] With commitment to a grand information structure, Pfizr would exactly know what information Novarty gets. That is, not the exact realization (i.e. the result of the trial), but the information structure overall (i.e. which trials will be conducted).

lyzed with tools provided by Bergemann and Morris (2016) and Taneva (2019) and the solution provides an upper bound for the CRO under the bilateral-contracting assumptions of interest.[16] Consider the following information structure, where both companies get one of two possible reports: either the trial reveals that the drug is ineffective (bad news, b) or the trial suggests the drug is effective but without fully proving the drugs efficacy (good news, g). The reports are generated according to the distribution shown in Table 1.[17]

Table 1: Optimal Information with Full Commitment.

		Report for Novarty			
		$\theta = 1$		$\theta = 0$	
		b	g	b	g
Report for Pfizr	b	0	0	0	$1/2$
	g	0	1	$1/2$	0

For example, when getting the good news, Pfizr will update its belief to get a posterior of $1/2$, but since the designer committed to the grand information structure Pfizr knows even more: Novarty will get bad news with probability $1/3$, which is higher than the ex-ante probability of bad news, equal to $1/6$. Furthermore, Pfizr also knows how the state describing the effectiveness of the drugs correlates with the Novarty reports. This reasoning about Novarty's reports is crucial because under these assumptions, a unique Bayes-Nash equilibrium exists,[18] where the receivers conduct further research if and only if they receive good news. Thus, with full commitment to a grand information structure the designer can ensure that at least one company will conduct further research with certainty, while both will conduct research with probability equal to the prior belief of $1/3$.

However, the CRO cannot actually commit to the grand information structure. Due to the bilateral-contracting assumption, the CRO can only commit to the marginal distributions and the receivers have to reason about the competitors' information. For

---

[16] Applying the more robust method akin to full implementation of Mathevet et al. (2020) yields the same result for this example.

[17] The information structure in Table 1 is optimal for a designer with symmetric, increasing, and submodular preferences, i.e. $v(R, D) = v(D, R)$, $v(R, \cdot) \geq v(D, \cdot)$, and $v(R, R) + v(D, D) \leq v(D, R) + v(R, D)$.

[18] The equilibrium action profile is also the unique interim-correlated rationalizable profile.

example, if the CRO adopts the above information structure, Pfizr could nevertheless conjecture that Novarty does not obtain any useful information from the CRO. For the information structure based on this conjecture, a Bayes-Nash equilibrium exists wherein Pfizr will drop the project given *either* report.[19] Novarty could reason similarly. If the CRO is concerned about adversarial selection, then the CRO's worst-case scenario results in both pharmaceutical companies dropping the project. The question then becomes, is there a way to get these companies to conduct further research given that only bilateral contracting is possible and the designer is concerned about adversarial selection?[20]

Table 2: Optimal Information for Adversarial Bilateral design.

		Report for Novarty			
		$\theta = 1$		$\theta = 0$	
		b	g	b	g
Report for Pfizr	b	0	0	$1/2$	$1/4$
	g	0	1	$1/4$	0

A positive answer is provided by the robust information structure described in Table 2.[21] This information structure reduces the overall probability of the good report from $2/3$ to $1/2$. Now, after receiving the good report the posterior is $2/3$, which makes R a dominant action. Thus, each report now has a unique dominant action[22] and the conjecture about the competitor's information no longer plays a role. The optimal information structure exactly balances the *trade-off* between inducing posteriors that are robust to receivers' conjecture about the information of their competitor and making further research as likely as possible. However, to achieve this, the proposed robust information structure reduces the probability of at least

---

[19] In this conjectured equilibrium, Novarty would conduct research, but this does not matter for the rest of the analysis.

[20] The arguments in this paragraph relate to a foundation I give in Subsection B.1 for the solution concept developed in Section 2.

[21] As before, this information structure is optimal for the same preferences as stated in Footnote 17.

[22] With the exact posterior of $2/3$ both actions are still undominated. Therefore, the induced posterior should be $2/3 + \varepsilon$ for some small $\varepsilon > 0$. This example ignores this tie-breaking issue here. The full theory presented below does account for this.

one receiver conducting further research to $2/3$.[23] Therefore, the CRO suffers a loss of about 33 percent that at least one company will conduct further research relative to the optimal full commitment information structure. This is the loss due to the constraints of bilateral contracting.

2 A Robust Solution Concept

This section develops a solution concept that delivers predictions that are robust in the sense that they depend on what information the player receives about the economic fundamental, but do not depend on how the player reasons about information other players might receive. I refer to these predictions as *individual robust predictions* and the corresponding solution concept is developed in two stages. The first stage builds on the concept of belief-free rationalizability (see Battigalli et al., 2011).[24] This version of rationalizability is robust to any information any player might get. Thus, this stage corresponds to robustness across information structure from an outside observer. For the purposes of this paper, this solution concept is too extreme since it does not take into account any information that a player gets about the state of nature, which describes, for example, the effectiveness of a drug. The second stage of the solution concept adds exactly this information, therefore refining belief-free rationalizability. I argue that this new solution concept reflects the robust prediction given that a player knows his/her information about the state of nature.

There are two players $i \in N := \{1, 2\}$, who will be also called receivers.[25] Each player has a finite set of actions A_i and as usual $A = A_1 \times A_2$ denotes the set of action profiles.[26] Uncertainty is modeled via a finite set of *states of nature* denoted by Θ. Each agent's preferences are represented by a utility function $u_i : A \times \Theta \to \mathbb{R}$.

---

[23] Even with this robust information structure both receivers will conduct further research with probability of $1/3$.

[24] Battigalli (2003) and Battigalli and Siniscalchi (2003) introduce a more general class of versions of rationalizability. One instance corresponds to belief-free rationalizability.

[25] This section is concerned only with the predictions of receivers' actions for the given information structure. The sender/designer does not play a role and will be introduced later.

[26] I follow the standard notation that for a fixed player i, A_{-i} denotes the set of actions for the other player $3 - i$. More generally, I use this notation for any player-specific sets.

All these components form an *economic environment* $\mathcal{E} = \langle \Theta, (A_i, u_i)_{i \in N} \rangle$,[27] which is assumed to be common knowledge.

Example 1. The economic environment for the CRO example is succinctly described by the two payoff tables specified in Subsection 1.2. ◇

The economic environment does not specify any information the players might have. Most solution concepts need a specification of the information structure. However, Battigalli et al. (2011) provide a solution concept—*belief-free rationalizability*—that depends only on the economic environment, capturing the exact behavioral implications of (correct) common belief in rationality.[28] This concept is defined inductively as follows: for $i \in N$, let $BFR_i^0 := A_i$ and for any $k \in \mathbb{N}$ inductively define,[29]

$$BFR_i^k := \Big\{ a_i \in A_i : \exists \mu_i \in \Delta(\Theta \times A_{-i}) \text{ s.t.} \tag{1}$$

$$(1)\, \operatorname{supp} \mu_i \subseteq \Theta \times BFR_{-i}^{k-1} \;\text{ and }\; (2)\, a_i \in \operatorname*{arg\,max}_{a_i' \in A_i} \sum_{\theta, a_{-i}} \mu_i(\theta, a_{-i}) u_i(a_i', a_{-i}, \theta) \Big\}.$$

Then define $BFR_i := \cap_{k \geq 0} BFR_i^k$. According to the usual arguments (e.g. Wald, 1949; Pearce, 1984), this procedure is the same as deleting ex-post dominated actions iteratively. An action $a_i \in A_i$ is ex-post dominated (relative to $X_{-i} \subseteq A_{-i}$), if there exists $\alpha_i \in \Delta(A_i)$ such that

$$\sum_{a_i'} \alpha(a_i') u_i(a_i', a_{-i}, \theta) > \; u_i(a_i, a_{-i}, \theta), \quad \text{for all } (a_{-i}, \theta) \in X_{-i} \times \Theta.$$

Example 2. In the CRO example from Subsection 1.2 it is easy to see that no action is ex-post dominated; hence $BFR_i = A_i$. ◇

---

[27] This is different from a basic game which is widely used in information design (see e.g. Bergemann and Morris, 2013; Mathevet et al., 2020). The difference is that a basic game also specifies a common prior on the states of nature.

[28] Bergemann and Morris (2017) also mention this solution concept, but they call it ex post rationalizability. They also define a notion of belief-free rationalizability, which is stronger than the version used here.

[29] As usual, for any set X, $\Delta(X)$ denotes the set of probability measures on X. If the underlying set X is infinite, I will differ slightly from the standard notation by denoting the set of finite support probability measures with $\Delta(X)$. For any $\mu \in \Delta(X)$, $\operatorname{supp} \mu$ denotes the support of μ.

As mentioned at the beginning of this section, belief-free rationalizability only takes the economic environment and rationality as primitive objects. In the current situation, a player has some information about the state of nature which affects his/her *individual robust predictions*.[30] Thus, Player 1 is assumed have a prior $\pi_1 \in \Delta(\Theta)$ and gets some information about the state of nature, which is described by a marginal information structure.[31]

Definition 1. *Fix an economic environment $\mathcal{E}$. A* marginal information structure *(*for $\mathcal{E}$*) is $I_1 = \langle S_1, \psi_1 \rangle$, where (i) S_1 is a finite set of signals and (ii) $\psi_1 : \Theta \to \Delta(S_1)$ is a conditional signal distribution.*

This marginal information structure does not specify any possible signals for the other player, nor does it it specify the signal distribution for the other player. Thus, this marginal information structure provides information *only* about the state of nature. The solution concept depends only on the marginal information structure.[32] This solution concept will be a set of pure strategies denoted by $R_1(I_1, \pi_1) \subseteq A_1^{S_1}$ and is formally defined as follows.

Each signal realization $s_1 \in S_1$ induces a posterior belief[33] $\mu_{s_1} \in \Delta(\Theta)$ by Bayesian updating. Since these signals only induce a belief about the state of nature θ, these beliefs are not rich enough to form a best-reply in an interactive setting. To form a best-reply, beliefs about the actions of the other player are also needed. A *rational-extended* belief incorporates this additional requirement by assigning positive probability only to the belief-free rationalizable actions of the other player.

Definition 2. *Fix an economic environment $\mathcal{E}$, a prior $\pi_1 \in \Delta(\Theta)$ and a marginal information structure I_1. A* rational-extended belief *for $s_1 \in S_1$ is a belief $\tilde{\mu}_1 \in \Delta(\Theta \times A_2)$ such that (i)* $\operatorname{marg}_\Theta \tilde{\mu}_1 = \mu_{s_1}$ *as given by Bayesian updating and (ii)* $\operatorname{supp} \tilde{\mu}_1 \subseteq \Theta \times BFR_2$*. Let $\mathcal{M}_1 : S_1 \rightrightarrows \Delta(\Theta \times A_2)$ denote the set of rational-extended*

---

[30] The remainder of this section describes the perspective of Player 1. To apply it to Player 2, switch the player indicies.

[31] The restriction to finite signals might not be without loss. However, an extension to countable signal spaces is straightforward and whether this is without loss remains an open question. In this section, I also assume that each signal realization $s_1 \in S_1$ has (ex-ante) positive probability. This can be relaxed at the cost of more cumbersome notation. See Subsection B.1.

[32] Similar to before, the solution concept also depends on the economic environment, but this dependence will be implicit.

[33] To save on notation, the player's index is kept implicit by using the signals' index.

beliefs for each $s_1 \in S_1$*, i.e.*

$$\mathcal{M}_1(s_1) = \{\tilde{\mu} \in \Delta(\Theta \times A_2) : \tilde{\mu} \textit{ is a rational-extended belief for } s_1\}.$$

Finally, these rational-extended beliefs allow me to define the individual robust prediction.

Definition 3. *Fix an economic environment* $\mathcal{E}$*, a prior* $\pi_1 \in \Delta(\Theta)$*, and a marginal information structure* I_1*. A pure strategy* $b : S_1 \to A_1$ *is* conceivable *for* (π_1, I_1) *if* b *is optimal for at least one selection of* $\mathcal{M}_1$*, i.e.* b *is optimal given* μ_1*, i.e. for each* $s_1 \in S_1$*, there exists* $\tilde{\mu}_1 \in \mathcal{M}_i(s_1)$ *such that* $b(s_1) \in \arg\max_{a_1' \in A_1} \sum_{\theta, a_2} \tilde{\mu}_1(\theta, a_2) u_1(a_1', a_2, \theta)$*. The* individual robust prediction *is the set of all conceivable strategies and is denoted by* $R_1(I_1, \pi_1)$*.*

A foundation in terms of explicit epistemic assumptions is discussed Subsection 5.2: the individual robust prediction corresponds to the behavioral implications of *common* knowledge of the economic environment, *common* belief in rationality, and *knowledge* of the marginal information structure. Thus, the prediction does not rely on implicit or explicit *common* knowledge assumptions about the marginal information structure. This is relevant for later questions about information design. The nature of bilateral contracting allows the designer to only commit to a marginal information structure. The receiver understands this marginal information, but needs to reason about what actions their opponent chooses. This reasoning process is not transparent to the designer. Thus, all actions the designer can rule out are exactly those strategies that are not part of the individual robust prediction. This is the essence of Definition 3.[34] Independently of the foundations, the robust predictions are often simple to calculate as the following example shows.

Example 3. Table 3 shows the marginal information for Pfizr induced by the full commitment optimal information structure described in Table 1. The bad report

---

[34] In Subsection B.1, I provide another foundation of this solution concept in terms of informational robustness and Bayes-Nash equilibirum analyis similar in spirit to Bergemann and Morris (2013, 2016, 2017). This foundation relies on a theory of how player's resolve uncertainty about the grand information structure: each player conjectures a grand information structure consistent with their marginal information structure. Given this conjecture, each player chooses a strategy as predicted by a Bayes-Nash equilibrium. The individual robust predictions correspond to the union across all such conjectures and all corresponding equilibria.

Table 3: Pfizr's marginal information derived from the information structure of Table 1.

		$\theta = 1$	$\theta = 0$
Report for Pfizer	b	0	$1/2$
	g	1	$1/2$

leads to a posterior[35] of zero, whereas the good report induces a posterior belief of $1/2$. Example 2 established that all actions are belief-free rationalizable. Thus, the sets of rational-extended beliefs for each signal are given by:

$$\mathcal{M}_P(b) = \{\tilde{\mu} \in \Delta(\Theta \times A_N) : \tilde{\mu}(1,R) + \tilde{\mu}(1,D) = 0\}\,, \text{ and}$$
$$\mathcal{M}_P(g) = \{\tilde{\mu} \in \Delta(\Theta \times A_N) : \tilde{\mu}(1,R) + \tilde{\mu}(1,D) = 1/2\}\,.$$

Since Research (R) is a dominated action if the drug is ineffective, R cannot be part of the individual robust prediction for the bad report. However, for the good report both actions are conceivable. For example, D is a best-reply to $\mu(1,R) = 1 - \mu(0,R) = 1/2$, whereas R is a best-reply $\mu(1,D) = 1 - \mu(0,D) = 1/2$. Both beliefs are valid rational-extended belief for the good signal. Thus, the individual robust prediction for Pfizr is $R_P(\text{Table 3}, 1/3) = \{(D,D),(D,R)\}$, where the first coordinate indicates the action after the bad report, and the second coordinate corresponds to the good report. ◇

Thus far the solution concept has been stated from an ex-ante perspective, which is relevant for later questions about information design question. However, it will also be useful to have the solution concept in an interim form. This is done by defining a correspondence $R_1(\cdot|I_1,\pi_1) : S_1 \rightrightarrows A_1$ as

$$R_1(s_1|I_1,\pi_1) := \{a_1 \in A_1 : \exists b \in R_1(I_1,\pi_1) \text{ s.t. } a_1 = b(s_1)\}\,.$$

The interim individual robust prediction relies only on the belief about the state of nature that is induced by the signal. Thus, the solution concept does not depend on the (marginal) information structure it is defined for, but only on the posteriors it generates. Moreover, the robust predictions can be strategically distinguished

---

[35] Recall that within this example beliefs correspond to the likelihood of the state of the drug being effective ($\theta = 1$).

by changing the economic environment. The following proposition formalizes these simple observations, which will be useful to address the information-design question.

Proposition 1. *Fix a set of states of nature* Θ. *Consider an economic environment* $\mathcal{E}$ *(with states of nature given by* Θ*), two priors* $\pi_1, \pi'_1 \in \Delta(\Theta)$ *and two marginal information structures* $I_1 = \langle S_1, \psi_1 \rangle$ *and* $I'_1 = \langle S'_1, \psi'_1 \rangle$. *For all* $(s_1, s'_1) \in S_1 \times S'_1$, *if* $\mu_{s_1} = \mu_{s'_1}$, *then* $R_1(s_1|I_1, \pi_1) = R_1(s'_1|I'_1, \pi'_1)$.

Conversely, consider two priors $\pi_1, \pi'_1 \in \Delta(\Theta)$ *and two marginal information structures* $I_1 = \langle S_1, \psi_1 \rangle$ *and* $I'_1 = \langle S'_1, \psi'_1 \rangle$. *If there exists* $(s_1, s'_1) \in S_1 \times S'_1$ *and* $\theta \in \Theta$ *such that* $\mu_{s_1}(\theta) \neq \mu_{s'_1}(\theta)$ *then there exists a (finite) economic environment (holding* Θ *fixed) such that* $R_1(s_1|I_1, \pi_1) \cap R_1(s'_1|I'_1, \pi'_1) = \emptyset$.

With Proposition 1 in mind,[36] I abuse notation for the interim version of the solution concept and write it as a correspondence defined on belief space, i.e. $R_1 : \Delta(\Theta) \rightrightarrows A_1$. Thus, R_1 denotes the ex-ante version, whereas $R_1(\mu_1)$ denotes the interim version. The interim notion is illustrated by applying it to the CRO example.

Example 4. Due to the binary state space, the interim individual robust predictions (defined on belief space) can be illustrated by means of a simple diagram. Figure 1 shows these predictions for both companies, where, a belief corresponds to the probability of the drug being effective. It was already argued in the introduction, that for beliefs greater than $2/3$ R is uniquely undominated, whereas for beliefs lower than $1/3$ D is the only dominant action. For all intermediate beliefs, a similar argument as in the previous example can establish that both actions are the individual robust prediction. ◇

3 ADVERSARIAL BILATERAL INFORMATION DESIGN

The previous section prepared the stage to address the question of information design with bilateral contracting. Due to the nature of bilateral contracting, receivers' behavior is not uniquely predicted and the information designer is concerned about

---

[36] As stated the proposition requires that every signal happens with positive probability. If any signals have zero ex-ante probability, then the proposition needs to be adjusted to condition on positive probability signals only.

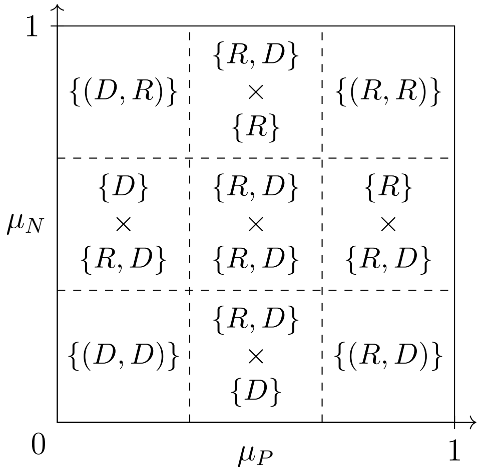

Figure 1: Individual robust predictions of the CRO game.

robustness to this uncertainty. For this, the previous section introduced a solution concept that captures robust predictions of receivers' actions. Crucially, this solution concept depends only on the receiver's belief about the states of nature. This feature produces a general representation theorem for information design with an adversarial and bilateral aspect.

To formally address the design question, the economic environment $\mathcal{E}$ needs to be appended with the preferences of the designer (she) $v : A \times \Theta \to \mathbb{R}$, which describes the utility she gets if the receivers take actions $a = (a_1, a_2)$. Furthermore, I assume that she knows the receivers' priors, and that these priors are the same as her prior, i.e. $\pi_1 = \pi_2 = \pi \in \Delta(\Theta)$.[37] Given this assumption, it is without loss to assume that the prior has full support. Together these components form a *design environment* $\mathcal{D} = \langle \mathcal{E}, \pi, v \rangle$. In such an environment, the designer chooses (grand) information structures, which specifies signals and distributions over signals for both receivers:

---

[37] The assumption says the designer knows the prior of the receivers, which happens to be the same prior. It does not state that players know the prior of their opponent, i.e. there is no *common* prior. Relaxing the assumption of the designer knowing the receivers' priors is active research even for the single receiver case. See, for example, Beauchêne et al. (2019), Kosterina (2020), and Pahlke (2020). Heterogeneous priors with the same support can be incorporated along the lines of Alonso and Câmara (2016). If priors with different supports are allowed, an extension is not straightforward. Galperti (2019) addresses some of the subsequent issues in the case of a single receiver. Applying Galperti's approach to the multiple receivers setting of this paper seems interesting for future research.

Definition 4. *Fix an economic environment $\mathcal{E}$. A (grand)* information structure (for $\mathcal{E}$) *is $I = \langle (S_1, S_2), \Psi \rangle$, where for each player $i \in N$, (i) S_i is a finite set of signals and (ii) $\Psi_i : \Theta \to \Delta(S_1 \times S_2)$ is a conditional signal distribution. Let $\mathcal{I}$ denote the set of information structures (for $\mathcal{E}$).*

As before, I assume that each signal happens with positive probability.[38] Additionally, a given information structure I induces a marginal information structure, denoted by $\operatorname{marg}_i I$ (or sometimes just I_i—no confusion should arise), by marginalization. That is, $\psi_i(\cdot|\theta) = \operatorname{marg}_{S_i} \Psi(\cdot|\theta)$, for all $\theta \in \Theta$, which justifies the naming.

The timeline of the overall design game is as follows:

Step 1: Designer chooses an information structure $I \in \mathcal{I}$.

Step 2: Receivers learn their respective marginal information structure I_i.

Step 3: The state of nature θ realizes and signals (s_1, s_2) are sent according to $\Psi(\cdot|\theta)$.

Step 4: For each signal (s_1, s_2), Nature recommends a conceivable action for each receiver to minimize the payoff of the designer.

Step 5: Each receiver plays as recommended by Nature.

Step 6: Payoffs are realized.

The bilateral contracting assumption is reflected in Step 2: a contract only specifies the marginal information structure for each player. Step 4 corresponds to the adversarial selection of the receivers' actions. Due to bilateral contracts, there might be multiple conceivable actions for each receiver, giving rise to uncertainty as to which actions will be played. Here, the designer is assumed to be very sensitive to this uncertainty and she considers a worst-case scenario.

3.1 THE GENERAL PROBLEM AND ITS REPRESENTATION

With this timing in mind, the information-design problem can be stated formally as $\sup_{I \in \mathcal{I}} V(I)$, where

$$V(I) := \sum_{\theta \in \Theta, s \in S} \pi(\theta)\psi(s|\theta) \min_{(a_i \in R_i(s_i|I_i,\pi))_{i \in N}} v(a_1, a_2, \theta), \tag{2}$$

---

[38] This is without loss in this section.

and recall that I_i is the marginal information structure derived from I.[39] If a maximizer exists,[40] then the resulting information structure captures robustness in the following sense: the optimal information structure performs well no matter how Nature chooses and coordinates the receivers' conceivable actions.

Given the structure of the problem, a natural approach would be to try to use a version of the revelation principle. However, the standard revelation principle argument á la Myerson (1982) does not apply here: this approach requires tie-breaking in favor of the designer. Instead, adversarial selection, by definition, selects actions that are incentive-compatible for the agents and bad for the principal. The following example illustrates that such an approach is bound to fail and shows that the problem is even more subtle than the tie-breaking issue.[41]

Example 5. Let $\Theta = \{0, 1\}$ and consider an economic environment, where player 2 has two actions (x and y) and is indifferent between them. Thus, $R_2(\mu_2) = \{x, y\} = A_2$ for any $\mu_2 \in \Delta(\Theta)$. Player 1 has three actions a, b, c and payoffs are given by Table 4.

Table 4: Payoffs for Player 1.

		Player 2's action			
		$\theta = 1$		$\theta = 0$	
		x	y	x	y
Player 1's action	a	2	0	0	2
	b	3	0	1	0
	c	0	1	0	3

First, b is conceivable for any belief: b is a best-reply if Player 1 is certain that player 2 chooses x. Similarly, c is also always conceivable. For beliefs close to certainty of either state, a is dominated by a mixture of b and c (e.g. in state $\theta = 1$ almost all the weight of the mixture will be on b). However, beliefs around $1/2$ about θ makes a conceivable. For example, suppose the belief about θ is exactly $1/2$, then consider

---

[39] See the discussion after Definition 1.

[40] In general, a maximizer might not exist. The adversarial approach includes tie-breaking against the designer's favor. This can lead to a failure of upper semicontinuity of the objective function.

[41] I am indebted to Marciano Siniscalchi for providing this simple, yet elucidative, example. Inostroza and Pavan (2018, Example 1) illustrate a similar issue when the designer has full commitment.

the following rational-extended belief: $\tilde{\mu}(1,x) = \tilde{\mu}(0,y) = 1/2$. For this belief, a is a best-reply. It can be verified that for any belief $\mu \in \Delta(\Theta)$ such that $\mu \in [1/4, 3/4]$ a is conceivable.

Now, consider a designer who only cares about Player 1's action. In particular, assume her (state-independent) preferences are given by $a \prec b \prec c$. Figure 2 shows the robust predictions for Player 1 in belief space and the implied worst-case selection for the designer.

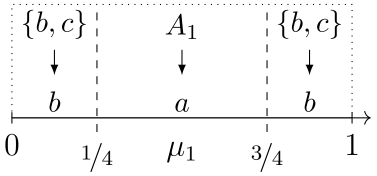

Figure 2: Robust Predictions for Player 1 and implied designer's worst-case choice.

For any prior $\pi \in \Delta(\Theta)$ the designer can get her (constrained) best outcome (b) by fully revealing the state. This optimal payoff cannot be attained with recommendation in general. For example, consider a prior belief of $\pi = 1/2$. A recommendation would send b with certainty. However, this signal does not provide information beyond the prior and therefore the worst-case prediction will be a rather than b as recommended.

The crucial failure is that a revelation principle with some sort of recommendations usually works by pooling signals together. This gives rise to a posterior that is a convex combination of the posteriors derived from each of the pooled signals. However, it is not true that a best-reply to the convex combination is also a best-reply to one of the original posteriors. For example, here, a is a best-reply to a convex combination of beliefs that are certain about a state. For each of these extreme beliefs, a is dominated by either b or c. ◇

Example 5 illustrates that there is no obvious simplification in signal space available that does not use some specific structure of the underlying economic environment. Since the individual robust prediction R_i depends only on the belief induced by the signal (see Proposition 1), the problem can be simplified by working with beliefs directly similarly to the single-receiver case of Kamenica and Gentzkow (2011). However, for multiple receivers, their approach does not readily extend itself because the designer has to address the full hierarchy of beliefs. This approach has been studied by Mathevet et al. (2020).

In the present paper the information designer can only commit to the marginal information structures because of the bilateral contracting assumption. In this setting, the players know what information they will receiver about the state of nature, but they do not know what information their opponent receives. The individual robust prediction corresponds to such an environment. Thus, the current setting raises a question about which distribution over beliefs can be induced by an information structure.[42]

Bayesian updating gives rise to receiver i's posterior belief about the state of nature $\mu_{s_i} \in \Delta(\Theta)$.[43] Furthermore, the information structure gives rise to a distribution over beliefs and the state of nature, i.e. an element of $\Delta(\Delta(\Theta) \times \Delta(\Theta) \times \Theta)$. Formally, this distribution τ is given by

$$\lambda(\mu_1, \mu_2, \theta) = \sum_{i \in N} \sum_{s_i : \mu_{s_i} = \mu_i} \pi(\theta) \Psi(s_1, s_2 | \theta). \tag{3}$$

Say a distribution over beliefs τ is induced by some information structure, if there exists an information structure such that τ can be derived from the information structure by Bayesian updating and Equation 3. Using Proposition 1, the objective from Equation 2 can be rewritten as follows:

$$\begin{aligned} V(I) &= \sum_{\theta \in \Theta, s \in S} \pi(\theta) \psi(s|\theta) \min_{\left(a_i \in R_i(\mu_{s_i})\right)_{i \in N}} v(a_1, a_2, \theta) \\ &= \sum_{\mu_1, \mu_2, \theta} \lambda(\mu_1, \mu_2, \theta) \min_{(a_i \in R_i(\mu_i))_{i \in N}} v(a_1, a_2, \theta), \end{aligned}$$

where λ corresponds to the distribution over beliefs induced by I. Now, the objective is stated purely in terms of beliefs and the actual information structure no longer plays a role. However, a simplification of the design problem calls for a characterization of

---

[42] Indeed, this is an open question in the literature. Ely (2017, p. 47) raises this concern quite directly by stating that "[...] there is no useful generalization for the multi-agent case".

[43] Mechanically, Bayesian updating gives rise to a belief about the other receiver's signals as well. However, as mentioned above only beliefs about the states matter.

a subset of $\Delta(\Delta(\Theta) \times \Delta(\Theta) \times \Theta)$ so that every element of this subset is induced by some information structure.

Obviously, consistency with the prior π requires the marginal of λ on the state space to coincide with π, i.e. $\operatorname{marg}_\Theta \lambda = \pi$. Furthermore, it is well known that another requirement that needs to be satisfied for any distribution over beliefs is that the belief of each player averages out to the prior, i.e. for each $i \in N$

$$\sum_{\mu_1, \mu_2, \theta} \mu_i \lambda(\mu_1, \mu_2, \theta) = \pi. \tag{4}$$

Kamenica and Gentzkow (2011) show that this condition is also sufficient to characterize the marginal distribution over beliefs for each player. However, these martingale properties on the marginals are not enough to characterize the possible joint distributions. Intuitively, what is missing are constraints linking together co-movement of beliefs across players.

Table 5: λ not induced by any information structure.

		μ_2			
λ		$\theta = 1$		$\theta = 0$	
		0	1	0	1
μ_1	0	0	0	0	$1/2$
	1	$1/2$	0	0	0

Example 6. Let $\Theta = \{0, 1\}$ and consider a uniform prior. Table 5 states a candidate distribution λ, which satisfies consistency with the prior and satisfies the martingale property for each player. However, no information structure induces such a distribution over beliefs. Intuitively, why no information structure can give rise to such a posterior distribution is easily seen: the extreme posteriors reflect the idea that the information structure fully reveals the state to the receivers. But if this is the case, there is no way to reveal one state to Player 1 and, at the same time, reveal the other state to Player 2. ◇

The following representation theorem takes care of the restrictions across players and follows from a direct-revelation argument in belief space:

Theorem 1 (Representation Theorem)**.** *Fix a design environment* $\mathcal{D}$ *and define* $\nu(\mu_1,\mu_2,\theta) := \min_{(a_i \in R_i(\mu_i))_{i\in N}} v(a_1,a_2,\theta)$*. The designer's problem can be represented as*

$$\sup_{I\in\mathcal{I}} V(I) = \sup_{\lambda\in\Delta(\Delta(\Theta)^2\times\theta)} \sum_{\mu_1,\mu_2,\theta} \lambda(\mu_1,\mu_2,\theta)\nu(\mu_1,\mu_2,\theta)$$

$$\text{s.t.} \quad (1)\ \operatorname*{marg}_{\Theta} \lambda = \pi,$$

$$(2)\ \frac{\sum_{\mu_{-i}} \lambda(\mu_i,\mu_{-i},\cdot)}{\sum_{\mu_{-i},\theta} \lambda(\mu_i,\mu_{-i},\theta)} = \mu_i, \quad \textit{for every } \mu_i \in \operatorname{supp}\lambda.$$

Furthermore, this restated problem is is a linear program.

Example 6 (continuing from p. 22)**.** The proposed distribution of Table 5 does not satisfy condition (2) of the program states in Theorem 1. To see this, consider the case where Player 2 is certain of $\theta = 0$ (i.e. $\mu_2 = 0$), so that

$$\frac{\sum_{\mu_1} \lambda(\mu_1,0,0)}{\sum_{\mu_1,\theta} \lambda(\mu_1,0,\theta)} = 0 \neq \mu_2(0) = 1-\mu_2 = 1.$$

This formally illustrates that λ is not induced by any information structure as was intuitively explained before. ◇

The characterization of the distributions over belief in the representation theorem does not make use of the martingale properties (Equation 4). Indeed, the two conditions in the theorem imply the martingale condition, because

$$\sum_{\mu_i,\mu_{-i},\theta} \tau(\mu_i,\mu_{-i},\theta)\mu_i = \sum_{\mu_i} \mu_i \sum_{\mu_{-i},\theta} \tau(\mu_i,\mu_{-i},\theta) \overset{(2)}{=} \sum_{\mu_i}\sum_{\mu_{-i}} \tau(\mu_i,\mu_{-i},\cdot) \overset{(1)}{=} \pi. \tag{5}$$

Furthermore, this characterization is a direct extension of the single-receiver characterization of Kamenica and Gentzkow (2011), i.e. if one receiver does not get any information the martingale condition (Equation 4) remains the only constraint for the other receiver.[44]

---

[44] Henceforth, for a given set X and any $x \in X$, $\delta_x \in \Delta(X)$ denotes the Dirac measure concentrated at x.

Corollary 1 (Kamenica and Gentzkow, 2011)**.** *Fix an economic environment* $\mathcal{E}$ *and a full-support prior* $\pi \in \Delta(\Theta)$*. Consider* $\lambda \in \Delta(\Delta(\Theta) \times \Delta(\Theta) \times \Theta)$ *with marginals (on* $\Delta(\Theta)$*)* τ_1 *and* τ_2 *and suppose that (i)* $\tau_2 = \delta_\pi$ *and (ii)* $\operatorname{marg}_\Theta \lambda = \pi$*. Then,* λ *is induced by an information structure if and only if* $\sum_{\mu_1} \tau_1(\mu_1)\mu_1 = \pi$*.*

4 THE CASE OF PURE PERSUASION

So far the general problem allowed for the designer having intrinsic preference on how the information is provided to the receivers. This section will address the case where the designer only cares about the receivers' actions, but does not care about the state of nature herself. Henceforth, with a slight abuse of notation, the designer's preferences are given by $v : A \to \mathbb{R}$. Thus the objective of the designer becomes to maximize $\sum_{\theta\in\Theta, s\in S} \pi(\theta)\psi(s|\theta) \min_{(a_i \in R_i(s_i|I_i,\pi))_{i\in N}} v(a_1, a_2)$. In this case the belief-space approach simplifies the problem even further, because now the distribution over two beliefs of the receivers are a sufficient to calculate the expected value for the designer.

For a given a information structure I and the induced distribution over beliefs *and* states $\lambda \in \Delta(\Delta(\Theta) \times \Delta(\Theta) \times \Theta)$, consider the marginal distribution over beliefs alone, i.e. $\operatorname{marg}_{1,2} \lambda =: \tau \in \Delta(\Delta(\Theta) \times \Delta(\Theta))$. Similar to above this allows to exploit Proposition 1 to rewrite the objective in belief space as $\sum_{\mu_1,\mu_2} \tau(\mu_1, \mu_2)\nu(\mu_1, \mu_2)$, where $\nu(\mu_1, \mu_2) := \min_{(a_i \in R_i(\mu_i))_{i\in I}} v(a_1, a_2)$. However, working in belief space requires again a characterization of the choice set of the designer; that is a characterization of distributions over receivers' beliefs that can be induced by information structures. However, this is an open problem in the literature. Very recently and independently from my work, Arieli et al. (2020) provide such a characterization for the case of binary states. Their characterization requires a quantification over all subsets of the support of the distributions over beliefs and therefore does not readily yield a simplification of the design problem. Instead, I follow a different route by providing bounds on how dependent the beliefs can be across the two receivers. Although, these bounds turn out to be only a necessary condition for distributions over beliefs to be induced by any information structure, they are tractable and work for any (finite) number of states of nature.[45] Furthermore, these bounds are sufficient under more assumptions about the design environment. One set of such assumptions will be presented later.

4.1 MEASURING DEPENDENCE OF RANDOM VARIABLES

A bit more notation is needed to introduce the relevant measure of dependence for random variables that is also relevant when realizations are beliefs. Let $\mathbf{X}$ and $\mathbf{Y}$ be real-valued random variables.[46] distributed according to cumulative distribution functions (CDFs) F_X and F_Y, respectively. Then the Fréchet class $\mathcal{F}(F_X, F_Y)$ is the set of all joint CDFs with marginals given by F_X and F_Y.

Definition 5[47] (Joe, 1997, Section 2.2.1)**.** *Fix two univariate CDFs F_1 and F_2. Consider $F, F' \in \mathcal{F}(F_1, F_2)$. F' is said to be* more concordant than F *(denoted by $F \precsim F'$) if $F(x,y) \leq F'(x,y)$, for all $(x,y) \in \mathbb{R}^2$,*

Intuitively, this stochastic ordering formalizes the idea that large values happen more often together (across both dimensions) under F' than under F. Furthermore, the Fréchet class $\mathcal{F}$ can be bounded according to this stochastic ordering. That is, for given univariate CDFs F_1 and F_2, for every $F \in \mathcal{F}(F_1, F_1)$, $\underline{F} \precsim F \precsim \overline{F}$, where

$$\underline{F}(x,y) := \max\{0, F_1(x) + F_2(y) - 1\}, \text{ and} \tag{6}$$

$$\overline{F}(x,y) := \min\{F_1(x), F_2(y)\}. \tag{7}$$

These bounds are often called Fréchet-Hoeffding bounds[48] and they correspond to extremal dependence across the two dimensions. The lower bound corresponds to countermonotonic random variables (i.e. low realizations in one dimension happen only with high realizations in the other dimension), whereas the upper bound describes comonotonic random variables (i.e. perfect positive dependence). These bounds also describe the extremal dependence for information structures.[49]

---

[45] Arieli et al. (2020, Appendix B) discuss why my bounds are only necessary. An earlier version of this paper erroneously claimed that these bounds are sufficient in general as well. I thank Omer Tamuz for pointing this out to me.

[46] The definition readily extends to random variables taking values in a totally ordered set.

[47] This stochastic order is also known as the positive quadrant dependent (PQD) ordering. See, e.g., Shaked and Shanthikumar (2007, Chapter 9).

[48] They play an important role in Copula theory. For more see, for example, Nelsen (2006).

[49] For this, the set of individual signals needs be endowed with *any* total order. Recall that information structures are distributions over signals conditional on the state of nature, see Definition 4.

For distributions over beliefs more restrictive bounds can be established. In general, the Fréchet-Hoeffding bounds are too wide for distributions over beliefs. **??** shows a belief distribution that attains the lower Fréchet-Hoeffding bounds. However, this belief distributions cannot be induced by an information structure, meaning that the usual Fréchet-Hoeffding bounds can be tightened to bound the distributions of beliefs induced by any information structure.[50]

Thus, the usual Fréchet-Hoeffding bounds can be tightened to provide necessary conditions for distributions over beliefs induced by information structures. In this section, I introduce and discuss such bounds that are useful for the information design question at hand. Since these bounds concern CDFs defined on beliefs, the space of beliefs needs to be ordered. Although the proposed bounds hold for *any* total order, it is convenient to take a linear extension of the first-order stochastic dominance order. To do this, endow the state of nature Θ with a total order, i.e. $\Theta = \{\theta_1, \ldots, \theta_K\}$ for some finite $K < \infty$ and the order corresponds to the indexing set. Then endow $\Delta(\Theta)$ with a completion of first-order stochastic dominance giving rise to a lattice structure. Given $\mu, \mu' \in \Delta(\Theta)$, a sufficient condition for $\mu \geq \mu'$ is μ first-order stochastic dominating μ', i.e. for every $L = 1, \ldots, K$, $\sum_{k=1}^{L} \mu(\theta_k) \leq \sum_{k=1}^{L} \mu'(\theta_k)$.

Given this order, define CDFs over beliefs analogously to the case of CDFs of real-valued random variables. That is, for a given distribution $\tau \in \Delta(\Delta(\Theta))$, define the associated CDF by $T(\mu) = \sum_{\mu' \leq \mu} \tau(\mu')$. Similarly, $\Delta(\Theta) \times \Delta(\Theta)$ is endowed with the product order derived from the order on each dimension. Then, for any joint distribution $\tau \in \Delta(\Delta(\Theta) \times \Delta(\Theta))$ the associated (joint) CDF is given by

$$T(\mu) = T(\mu_1, \mu_2) = \sum_{\mu'_1 \leq \mu_1, \mu'_2 \leq \mu_2} \tau(\mu'_1, \mu'_2).$$

With these definitions in hand, the belief-dependence bounds can be defined. Similar to the Fréchet-Hoeffding bounds, these bounds are defined for given marginal distributions.

If all conditional distributions are equal to their (upper or lower) Fréchet-Hoeffding bound (fixing the conditional marginal distributions), then I say the information structures attains its bound.

[50] Conversely, although the information structure from Table 1 attains the lower Fréchet-Hoeffding bound, the induced belief distribution does not attain the Fréchet-Hoeffding bound.

Definition 6. *Fix two univariate distributions over beliefs* $\tau_1, \tau_2 \in \Delta(\Delta(\Theta))$ *and a prior* $\pi \in \Delta(\Theta)$. *The* lower belief-dependence bound *is defined as*

$$\underline{T}(\mu_1, \mu_2) = \max_{0 \leq L \leq K} \max \left\{ \underline{T}_1(\mu_1, \mu_2; L), \underline{T}_2(\mu_1, \mu_2; L) \right\}, \tag{8}$$

where for each[51] $L = 0, \dots, K$,

$$\begin{aligned} \underline{T}_1(\mu_1, \mu_2; L) &= \sum_{\mu_1' \leq \mu_1} \tau_1(\mu_1') \sum_{k=1}^{L} \mu_1'(\theta_k) + \sum_{\mu_2' \leq \mu_2} \tau_2(\mu_2') \sum_{k=1}^{L} \mu_2'(\theta_k) - \sum_{k=1}^{L} \pi(\theta_k), \\ &\textit{and} \\ \underline{T}_2(\mu_1, \mu_2; L) &= \sum_{\mu_1' \leq \mu_1} \tau_1(\mu_1') \sum_{k=L+1}^{K} \mu_1'(\theta_k) + \sum_{\mu_2' \leq \mu_2} \tau_2(\mu_2') \sum_{k=L+1}^{K} \mu_2'(\theta_k) - \sum_{k=L+1}^{K} \pi(\theta_k). \end{aligned} \tag{9}$$

The upper belief-dependence bound *is defined as*

$$\overline{T}(\mu_1, \mu_2) = \min_{1 \leq L \leq K} \min \left\{ \overline{T}_1(\mu_1, \mu_2; L), \overline{T}_2(\mu_1, \mu_2; L) \right\}, \tag{10}$$

where for each[52] $L = 0, \dots, K$,

$$\begin{aligned} \overline{T}_1(\mu_1, \mu_2; L) &= \sum_{\mu_1' \leq \mu_1} \tau_1(\mu_1') \sum_{k=1}^{L} \mu_1'(\theta_k) + \sum_{\mu_2' \leq \mu_2} \tau_2(\mu_2') \sum_{k=L+1}^{K} \mu_2'(\theta_k), \\ &\textit{and} \\ \overline{T}_2(\mu_1, \mu_2; L) &= \sum_{\mu_1' \leq \mu_1} \tau_1(\mu_1') \sum_{k=L+1}^{K} \mu_1'(\theta_k) + \sum_{\mu_2' \leq \mu_2} \tau_2(\mu_2') \sum_{k=1}^{L} \mu_2'(\theta_k). \end{aligned} \tag{11}$$

A few observation are in order. First, as argued in the previous section, the goal is to tighten the usual Fréchet-Hoeffding bounds using the restrictions imposed by the actual information structures and Bayesian updating. Thus, the belief-dependence bounds should be tighter, which is indeed the case. Formally, for the lower bound we have that $\underline{F}(\mu_1, \mu_2) \leq \underline{T}(\mu_1, \mu_2)$ since $\underline{F}(\mu_1, \mu_2) = \max_{L \in \{0,K\}} \underline{T}(\mu_1, \mu_2; L) \leq \underline{T}(\mu_1, \mu_2)$. For the upper bound the reversed inequality, $\overline{F}(\mu_1, \mu_2) \geq \overline{T}(\mu_1, \mu_2)$, holds

---

[51] By convention, empty sums are defined to be zero.

[52] Again, by convention, empty sums are defined to be zero.

because $\overline{F}(\mu_1, \mu_2) = \min\left\{\overline{T}_1(\mu_1, \mu_2; K), \overline{T}_2(\mu_1, \mu_2; K)\right\} \geq \overline{T}(\mu_1, \mu_2)$. Second, if the marginal distributions are equal, i.e. $\tau_1 = \tau_2$, then the upper belief-dependence bound is actually the same as the upper Fréchet-Hoeffding bound. Formally:

Lemma 1. *Fix an economic environment $\mathcal{E}$ and a full-support prior $\pi \in \Delta(\Theta)$. Consider two univariate distributions $\tau_1, \tau_2 \in \Delta(\Delta(\Theta))$ such that $\tau_1 = \tau_2$ and suppose that $\mathbb{E}_{\tau_1}[\mu_1] = \pi$. Then, the upper belief-dependence bound is the usual upper Fréchet-Hoeffding bound, i.e. $\overline{T} = \overline{F}$.*

Furthermore, these bounds are indeed necessary conditions for distributions over beliefs to be induced by any information structure.

Proposition 2. *Fix an economic environment $\mathcal{E}$ and a full-support prior $\pi \in \Delta(\Theta)$. $\tau \in \Delta(\Delta(\Theta) \times \Delta(\Theta))$ is induced by an information structure only if*[53]

1. $\sum_{\mu_1,\mu_2} \tau(\mu_1, \mu_2)\mu_1 = \sum_{\mu_1,\mu_2} \tau(\mu_1, \mu_2)\mu_2 = \pi$, *and*
2. $\underline{T} \precsim T \precsim \overline{T}$.

Thus, these bounds gives rise to a problem for finding an upper bound of the pure-persuasion design problem:

Corollary 2. *Fix a design environment $\mathcal{D}$. Then,*

$$\begin{aligned}\sup_{I \in \mathcal{I}} V(I) \leq \overline{V}(\pi) := \sup_{\tau \in \Delta(\Delta(\Theta)^2)} & \sum_{\mu_1,\mu_2} \tau(\mu_1, \mu_2)\nu(\mu_1, \mu_2) \\ \textit{s.t.} & \sum_{\mu_1,\mu_2} \tau(\mu_1, \mu_2)\mu_1 = \sum_{\mu_1,\mu_2} \tau(\mu_1, \mu_2)\mu_2 = \pi, \\ \textit{and} & \;\; \underline{T} \precsim T \precsim \overline{T}.\end{aligned}$$

This corollary shows that the designer solves the (relaxed) problem as if she chooses marginal belief distributions for each receiver subject to the familiar Bayes plausibility conditions. Moreover, the beliefs across the two receivers cannot be too dependent so that the joint distribution satisfies the belief-dependence bounds. The

---

[53] Here, a slight abuse of notation appears: the belief bounds are formally only defined for two marginal beliefs. In the statement there is only the joint distribution τ. The belief bounds correspond to the bounds defined by using the two marginals distributions derived from τ.

constraints on the distributions of beliefs are tractable, especially if the designer utility ν (as a function on belief space) has special properties.

For two-dimensional real-vectors it is well known[54] that the stochastic order $\precsim$ (recall Definition 5) has a dual characterization in terms of utility functions. In particular, $F \precsim G \iff \mathbb{E}_F[w(x,y)] \leq \mathbb{E}_G[w(x,y)]$, for all Bernoulli utility functions $w : \mathbb{R}^2 \to \mathbb{R}$ that are supermodular. Meyer and Strulovici (2015) extend this result to distribution over a finite, n-dimensional lattice. Since the order on beliefs was assumed to be a total order, Meyer and Strulovici's results apply to the setting of this paper. Thus, if ν in Corollary 2 is supermodular, then the pure persuasion design problem can be simplified by first solving

$$\begin{aligned} \sup_{\tau_1,\tau_2 \in \Delta(\Delta(\Theta))} & \sum_{\mu_1,\mu_2} \tau(\mu_1,\mu_2)\nu(\mu_1,\mu_2) \\ \text{s.t. } & \sum_{\mu_1} \tau_1(\mu_1)\mu_1 = \sum_{\mu_2} \tau_2(\mu_2)\mu_2 = \pi, \\ & \text{and } T = \overline{T}, \end{aligned}$$

and then verifying whether the resulting τ is induced by an information structure. Symmetrically, if ν is submodular the last constraint would be replaced by $T = \underline{T}$. In either case, the problem is simplified because the choice set contains only marginal distributions.[55]

Since the utility function ν is an object derived from the primitive objects stated in a design environment $\mathcal{D}$, I introduce a broad class of environments which provides easy verifiable sufficient conditions on primitives to ensure that the derived object ν satisfies sub- or supermodularity whenever the primitive function v satisfies these properties. In addition, a subclass of these environments allows me to provide an upper bound on the cardinality of the signal space (see Example 5).

Definition 7. *An economic environment* $\mathcal{E} = \langle \Theta, (A_i, u_i)_{i \in N} \rangle$ *is* monotone *if*

1. *the states of nature* Θ *are endowed with an total order,*

---

[54] In probability theory, this is known at least since Cambanis et al. (1976) and Tchen (1980).

[55] The concavification approach of Kamenica and Gentzkow (2011) can be useful here as well as it gives an (even more) relaxed version of the actual designer's problem. It suggests solving the concavification approach and then checking whether the resulting distribution is induced by an information structure (and therefore has to satisfy the belief bounds).

2. *for each player $i \in N$, the set of actions A_i is endowed with an total order, and*

3. *for each player $i \in N$, the utility function has increasing differences in (a_i, θ), i.e. for all $(a_i, \theta), (a'_i, \theta') \in A_i \times \Theta$ and all $a_{-i} \in A_{-i}$,*

$$a'_i \geq a_i \text{ and } \theta' \geq \theta \implies u_i(a'_i, a_{-i}, \theta') + u_i(a_i, a_{-i}, \theta) \geq u_i(a'_i, a_{-i}, \theta) + u_i(a_i, a_{-i}, \theta').$$

A design environment $\mathcal{D} = \langle \mathcal{E}, \pi, v \rangle$ is monotone *if*

1. *the economic environment $\mathcal{E}$ is monotone, and*

2. *the designer's utility function $v : A \to \mathbb{R}$ is increasing*[56] *with respect to the product order induced by the orders on the set of actions A_i, i.e. for all $(a_1, a_2) \in A$, $a'_i \geq a_i$, for all $i = 1, 2 \implies v(a'_1, a'_2) \geq v(a_1, a_2)$.*

This class of environments here is quite general,[57] but specific enough to translate the preference for complementarities from action space to belief space as formally stated in the next proposition. This proposition, therefore, provides a simple way to check the primitives to ensure that the derived Bernoulli utility in in belief space is either sub- or supermodular.

Proposition 3. *Consider a monotone design environment $\mathcal{D}$. Suppose the designer's utility $v : A \to \mathbb{R}$ is supermodular then the derived utility $\nu : \Delta(\Theta) \times \Delta(\Theta) \to \mathbb{R}$ on belief space (endowed with the first-order stochastic dominance order) is supermodular, where $\nu(\mu_1, \mu_2) := \min_{(a_i \in R_i(\mu_i))_{i \in N}} v(a_1, a_2)$. Similarly, if v is submodular, then ν is submodular as well.*

In the general problem, Example 5 illustrates that using recommendations similar to the usual revelation principle does not work. For monotone design environments with a restriction on information structures, action recommendations provide a rich enough signal space. Action recommendations turn out to be useful even when working in belief space as will be illustrated in Subsection 4.2.[58] For this, say that an

---

[56] Only monotonicity of v is needed for all of the following analysis. The definition uses increasingness to simplify the notation.

[57] Supermodular games usually have an underlying economic environment that is monotone. However, the class of monotone environments is more general since it does not specify increasing differences in (a_i, a_{-i}), which is assumed to transform an economic environment to supermodular game.

[58] Whether this revelation principle argument is useful for working directly in signal space is an open question.

information structure I is *direct* if for every $i \in N$, $S_i \subseteq A_i$ and for every signal $a = (a_1, a_2)$, it holds that $\min_{\left(a'_i \in R_i(s_i|I_i,\pi)\right)_{i\in N}} v(a'_1, a'_2) = v(a)$. Then, the following proposition is akin to a standard revelation principle.

Proposition 4 (Revelation Principle)**.** *Suppose the design environment $\mathcal{D}$ is monotone. Restrict the choice of information structures to information structures that give rise to posteriors that are totally ordered by first-order stochastic dominance for each player.*[59] *Then, there exists an information structure I with value $V(I)$ if and only if there exists a direct information structure $\hat{I}$ such that $v(I) = v(\hat{I})$.*

This result is interpreted slighlty differenlty the usual interpretation of the revelation principle as in Myerson (1982) or Kamenica and Gentzkow (2011). Here, the designer sends action recommendations to the receivers like in the usual version, but the receivers do not have to be obedient and follow the recommendation. Instead, whatever action the receiver chooses, for the designer the action will be at least as good as if the receiver had followed the recommendation.

4.2 THE PROBLEM OF A CRO SOLVED

Now, the problem of the CRO introduced Subsection 1.2 can be solved. Recall that the economic environment $\mathcal{E}$ can be summarized by the two game tables in Subsection 1.2. This economic environment is actually a monotone one. Furthermore, the prior of both pharmaceutical companies was specified as $\pi = {}^1\!/_3$, thus it remains to specify the preferences for the designer (i.e. the CRO) to get a design environment. For now, assume that preferences are such that the CRO prefers further research over dropping the project for both companies, i.e. $v(R, \cdot) > v(D, \cdot)$ $v(\cdot, R) > v(\cdot, D)$, which makes the design environment monotone as well. Using Figure 1, it is easy to obtain the CRO utility function defined on belief space, as shown in Figure 3.

Given this derived utility function ν, the optimal information and the corresponding value can be obtained by applying Theorem 1. For an explicit illustration, I will assume the CRO preferences are symmetric and submodular.[60] That is, $v(R, R) + v(D, D) \leq v(R, D) + v(D, R)$, which implies that ν is submodular. Here,

---

[59] For example, if the state space is binary, then this assumption is without loss of generality.

[60] In case of supermodular preferences, the concavification approach as explained above can be applied to derive the optimal information structure directly. Details are available upon request.

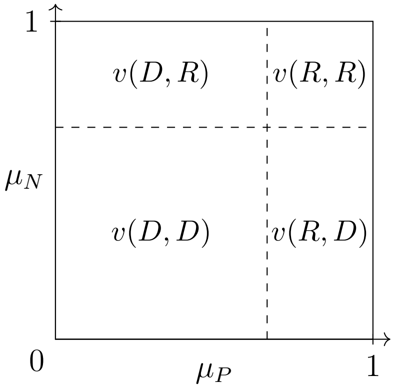

Figure 3: ν, CRO utility function defined on belief space.

the concavification approach is not useful since it would yield a belief distribution (see Table 6) which cannot be induced by any information structure. This can be verified by checking that this distribution violates the lower belief-dependence bound. Thus, a different approach is needed for this case. By Proposition 4, it is sufficient to consider marginal belief distributions with binary support only: one supported belief leads to actions D in the worst-case and the other leads to action R in the worst-case. Therefore, for each receiver we need to consider beliefs $(\mu_i^D, \mu_i^R) \in [0, {}^2\!/\!_3) \times [{}^2\!/\!_3, 1]$ only.[61] Moreover, it is easy to see that distributions leading to both actions with positive probability are better than just sticking to the prior (on each dimension). Thus, $(\mu_i^D, \mu_i^R) \in [0, {}^1\!/\!_2) \times [{}^2\!/\!_3, 1]$ by Bayes plausibility. Using Theorem 1 the solution is readily available computationally. However, in this case the problem can be solved directly using Corollary 2. As binary signals suffice and the problem has only states of nature, the bounds of Proposition 2 coincide with the characterization of Arieli

Table 6: Result from concavification approach for submodular preferences.

		Belief of Novarty	
		0	2/3
Belief of Pfizr	0	0	1/2
	2/3	1/2	0

---

[61] As before, we change the tie-breaking assumption here, which simplifies the notation, but does not change the essence of the argument.

et al. (2020).[62] First, the lower belief-dependence bound[63] has to be binding due to submodularity. Furthermore, it has to be strictly tighter at some point than the usual Fréchet-Hoeffding lower bound, otherwise Table 6 would be the solution. Given the binary signals per receiver and the possible values for these, the only point where the bound is binding is at (μ_1^D, μ_2^D). For the other cases the Fréchet-Hoeffding bound is the same as the belief-dependence bound. Thus, letting τ_i denote the marginal distributions, $\tau(\mu_1^D, \mu_2^D) = \tau_i(\mu_1^D)\left(1-\mu_1^D\right)+\tau_2(\mu_2^D)\left(1-\mu_2^D\right)-(1-\pi)$, has to hold for any possible joint distribution. This allows me to simplify the program as stated in Theorem 1 by making the problem separable between the two agents.[64] The reformulated program becomes

$$\sup_{\tau_1,\tau_2\in\Delta(\Delta(\Theta))} \sum_{\mu_1} \tau_1(\mu_1) f(\mu_1) + \sum_{\mu_2} \tau_2(\mu_2) f(\mu_2) \ \text{ s.t. } \sum_{\mu_1} \tau_1(\mu_1)\mu_1 = \sum_{\mu_2} \tau_2(\mu_2)\mu_2 = \pi,$$

where $f(\mu) := \mathbf{1}\left[\mu < {}^2\!/\!_3\right](2v+\mu-1) + \mathbf{1}\left[\mu \geq {}^2\!/\!_3\right](1-\mu)$ using a normalization on the payoffs for the CRO.[65] The solution to this program determines the optimal marginal distributions, which are then combined to a joint distribution via the lower belief-dependence bound. Due to the established separability, the reformulation can be solved with the concavification technique from Kamenica and Gentzkow (2011) yielding $\mu_i^{D,*} = 0$ and $\mu_i^{R,*} = {}^2\!/\!_3$. By Bayes plausibility this gives the same marginal distribution as in Table 6, but these marginals must be put together with the lower belief-dependence bounds. This yields the optimal information structure as foreshadowed in the introduction and stated in Table 2.

5 Discussion

In this section, I discuss some extensions of the model and highlight some conceptual aspects.

---

[62] Arieli et al. (2020) discuss details in their Appendix B.

[63] For the binary state case first-order stochastic dominance is a total order. By Lemma 2, only $\underline{T}_1$ has to be considered for the lower bound.

[64] Derivations are shown in Subsection B.2.

[65] In particular, $v(D,D) = -1$, $v(R,R) = 0$, and $v(R,D) = v(D,R) =: v \in [-{}^1\!/\!_2, 0]$. This is without loss of generality.

5.1 EXTENSION TO MULTIPLE RECEIVERS

In this paper, I have focused only on two players only. This simplifies the notation significantly. The solution concept introduced in Section 2 readily extends to any finite number of players if the definitions of belief-free rationalizability (Equation 1) and rational-extended beliefs (Definition 2) are adapted to allow for general correlated beliefs about the opponents' actions. Moreover, the general design problem discussed in Section 3 can be adjusted accordingly to multiple receivers. However, the bounds in Section 4 do not extend to multiple players without adaption. Of course, the functional form of the belief bounds is specific to two receivers, but a similar approach as in the proof of Proposition 2 can be adapted. To derive these bounds an extension of Joe (1997, Theorem 3.11) to higher dimensions is necessary. Deriving these bounds explicitly and studying their properties is left for future research.

5.2 COMMON BELIEF OF RATIONALITY

Throughout this paper, I operated from the assumption that the economic environment is common knowledge among the players. In the examples this did not matter too much, but it this knowledge is crucial for the solution concept, which also requires common knowledge of rationality. A slight adaption of Battigalli et al. (2011, Section 3.1–3.2, see also Section 4.2) shows that the individual robust prediction corresponds to the behavioral implications of common belief of the economic environment *and* rationality, as well as *knowledge* of the marginal information structure. For certain economic environments, this has important consequences for the design of information structures. To see this, consider the following economic environment:

		Novarty	
		R	D
Pfizr	R	$1 \quad -1$	$2 \quad 0$
	D	$0 \quad -1$	$0 \quad 0$

$\theta = 1$

Novarty	
R	D
$-2 \quad -1$	$-1 \quad 0$
$0 \quad -1$	$0 \quad 0$

$\theta = 2$

The payoffs for Pfizr are the same as in the CRO example, however Novarty now has an (ex-post) dominated action: R is always worse than D. For the same prior as before ($\pi = 5/9$) the robust prediction without any information would be $\{R\}$ for Pfizr (and, of course, D for Novarty). Thus, without providing any information the designer gets the best possible outcome. Suppose now, that the designer does not assume common knowledge of rationality among the receivers but still assumes rationality and knowledge of the marginal information structure for each receiver. The corresponding (even more) robust predication can be obtained by dropping part (2) in the definition of Definition 2. In this example, this version of robust prediction (interpreted as a function of first-order beliefs) for Pfizr yields the same as in the running example in the main text of this article. Therefore, if the designer is concerned about robustness under these less restrictive assumptions, she will engage in Bayesian Persuasion á la Kamenica and Gentzkow (2011) with Pfizr. This means that the designer will optimally reveal the state of the drug being ineffective sometimes, which implies that Pfizr will drop the project occasionally. This is in contrast to the behavior under the assumption of common knowledge of rationality, where Pfizr will conduct further research with certainty. What is the *right* optimal information structure for the designer? This depends on the assumptions the designer wants to make. In this paper, the designer imposes common knowledge of rationality.

5.3 Robust Information Design

The key aspect of robust mechanism design as initiated by Bergemann and Morris (2005), and the Wilson (1987)-doctrine more generally, is relaxing the implicit common knowledge assumption to obtain more realistic models. Given the discussion in the previous subsection, the model presented here can be interpreted likewise, but in the realm of information design. In robust mechanism design, the implicit assumptions are relaxed by considering a sufficiently rich Harsanyi-type space. In contrast, in information design the Harsanyi-type space is the actual designed information structure. Mathevet et al. (2020) provide a method to study this design problem. My model can be interpreted as relaxing the common knowledge assumption about the designed information structure. But to remain in the realm of information design, the players still know their designed marginal information structure. The solution concept proposed in this paper captures these assumptions exactly as explained in the previous section. In addition, the adversarial selection assumption reflects the robustness aspect.

5.4 LIMITED COMMITMENT AND ADVERSARIAL ROBUSTNESS

Due to the bilateral arrangement the information designer is effectively limited to commit only to the marginals of the (grand) information structure. This limited commitment might open up the possibility of communicating more than just the marginal information structure to the respective receivers. For example, receivers might engage in forward-induction upon seeing the proposed marginal information structure. If this additional communication is explicitly account for then several solution concepts are available for the resulting communication game. Picking one such solution concept over the others is problematic in the view of robustness. However, any implied receiver's behavior under any solution concept that maintains common belief of rationality and knowledge of the marginal information structure must be within the individual robust prediction. Thus, the approach in this paper is robust to the selection across these solution concepts as well. Furthermore, in any such communication game the assumptions about receiver's knowledge of the designer's preferences becomes crucial as well.[66] Again, the proposed design approach sidesteps this issues using the adversarial approach in combination with the robustness incorporated in the individual robust prediction.

Thus, allowing for this extraneous communication and making this appropriate assumptions explicit is in contrast to the robust design considered in the main text. However, making these stronger assumptions might be of interest for particular applications. Since this is beyond the scope if this paper, it is left for future research.

A PROOFS

Proof of Proposition 1. The statement is trivial if $|\Theta| = 1$, so suppose $|\Theta| > 1$.

The first part follows directly from the definition, since BFR_i depends only on the economic environment and the rational-extended beliefs exactly capture only the beliefs about the states of nature, which are the same by assumption.

For the second part, fix $\theta' \in \Theta$ such that

$$\mu := \frac{\psi_1(s_1|\theta')\pi_1(\theta')}{\sum_\theta \psi_1(s_1|\theta)\pi_1(\theta)} \neq \frac{\psi_1'(s_1'|\theta')\pi_1'(\theta')}{\sum_\theta \psi_1'(s_1'|\theta)\pi_1'(\theta)} =: \mu'.$$

---

[66] This, of course, raises the next level question, of whether this assumption is common knowledge and whether the designer's rationality is (common) knowledge.

Consider the following economic environment: $A_i = \{\mu, \mu'\}$ and payoffs are given by $u_i(a_i, a_{-i}, \theta) = (a_i - 1[\theta = \theta'])^2$. By construction only the belief about the state matters for best-replies, so the difference between the induced belief on Θ and an rational-extended belief does not matter. Now, note that μ (as action) is the unique best-reply to μ (as belief). Then, by construction $R_1(s_1|I_1, \pi_1) = \{\mu\}$ and $R_1(s_1'|I_1', \pi_1') = \{\mu'\}$ and the conclusion follows. ■

Proof of Theorem 1. I only proof (1)+(2). The rest is obvious or follows from the previous discussion.

Fix an information structure $I \in \mathcal{I}$ and let $\lambda \in \Delta(\Delta(\Theta) \times \Delta(\Theta) \times \Theta)$ be the induced distribution. Then (1) is satisfied, because

$$\sum_{\mu_1,\mu_2} \lambda(\mu_1, \mu_2, \theta) = \sum_{\mu_1,\mu_2} \left[\sum_{i \in N} \sum_{s_i : \mu_{s_i} = \mu_i} \pi(\theta)\Psi(s_1, s_2|\theta) \right] = \pi(\theta) \sum_{s_1,s_2} \Psi(s_1, s_2|\theta) = \pi(\theta).$$

For (2), consider $\mu_1 \in \Delta(\Theta)$. Then,

$$\sum_{\mu_2} \lambda(\mu_1, \mu_2, \theta) = \sum_{\mu_2} \left[\sum_{i \in N} \sum_{s_i : \mu_{s_i} = \mu_i} \pi(\theta)\Psi(s_1, s_2|\theta) \right] = \sum_{s_1 : \mu_{s_1} = \mu_1} \sum_{s_2} \pi(\theta)\Psi(s_1, s_2|\theta)$$

$$= \sum_{s_1 : \mu_{s_1} = \mu_1} \left[\mu_{s_1}(\theta) \sum_{s_2, \theta'} \pi(\theta')\Psi(s_1, s_2|\theta') \right] = \mu_1(\theta) \sum_{s_1 : \mu_{s_1} = \mu_1} \left[\sum_{s_2, \theta'} \pi(\theta')\Psi(s_1, s_2|\theta') \right]$$

$$= \mu_1(\theta) \sum_{\mu_2, \theta'} \lambda(\mu_1, \mu_2, \theta').$$

The argument for player 2 is the same.

Conversely, suppose there exists λ with conditions (1) and (2), I will construct an information structure which induces λ. For this let $S_1 = \operatorname{supp} \operatorname{marg}_1 \tau$ and $S_2 = \operatorname{supp} \operatorname{marg}_2 \tau$ and define the conditional signal distribution as $\Psi(\mu_1, \mu_2|\theta) = \frac{\lambda(\mu_1, \mu_2, \theta)}{\pi(\theta)}$. Note that condition (1) implies that Ψ gives rises to valid distributions. Furthermore, for signals which happen with positive probability condition (2) gives

$$\mu_{\mu_i}(\theta) = \frac{\sum_{\mu_{-i}} \lambda(\mu_i, \mu_{-i}, \theta)}{\sum_{\mu_{-i}, \tilde{\theta}} \lambda(\mu_i, \mu_{-i}, \theta)} = \mu_i(\theta).$$

Hence, $\sum_{i \in N} \sum_{s_i : \mu_{s_i} = \mu_i} \pi(\theta)\Psi(s_1, s_2|\theta) = \pi(\theta)\Psi(\mu_1, \mu_2|\theta) = \lambda(\mu_1, \mu_2, \theta)$ so that the constructed information structure induces λ. ■

For the following consider the following definitions. Given $\tau_1, \tau_2 \in \Delta(\Delta(\Theta))$ and a prior $\pi \in \Delta(\theta)$ define

$$\Pi(L) = \sum_{k=1}^{L} \pi(\theta_k) \quad \text{and} \quad T(\mu_1, \mu_2) = \sum_{\mu_1' \leq \mu_1} \sum_{\mu_2' \leq \mu_2} \tau(\mu_1', \mu_2')$$

$$T_1(\mu_1) = \sum_{\mu_1' \leq \mu_1} \tau_1(\mu_1') \quad \text{and} \quad T_2(\mu_2) = \sum_{\mu_2' \leq \mu_2} \tau_2(\mu_2')$$

$$M_1(\mu_1, L) = \sum_{\mu_1' \leq \mu_1} \tau_1(\mu_1') \sum_{k=1}^{L} \mu_1'(\theta_k) \quad \text{and} \quad M_2(\mu_2, L) = \sum_{\mu_2' \leq \mu_2} \tau_2(\mu_2') \sum_{k=1}^{L} \mu_2'(\theta_k).$$

With these definitions, the elementary functions of the belief-dependence bounds can be restated as

$$\begin{aligned}
\underline{T}_1(\mu_1, \mu_2; L) &= M_1(\mu_1, L) + M_2(\mu_2, L) - \Pi(L) \\
\underline{T}_2(\mu_1, \mu_2; L) &= T_1(\mu_1) - M_1(\mu_1, L) + T_2(\mu_2) - M_2(\mu_2, L) - [1 - \Pi(L)] \\
\overline{T}_1(\mu_1, \mu_2; L) &= M_1(\mu_1, L) + T_2(\mu_2) - M_2(\mu_2, L) \\
\overline{T}_2(\mu_1, \mu_2; L) &= T_1(\mu_2) - M_1(\mu_2, L) + M_2(\mu_1, L),
\end{aligned}$$

for every $L = 0, \ldots, K$.[67]

Proof of Lemma 1. By Symmetry $T_1 = T_2$ and similar for M_i. Thus, I will drop the indices. Without loss say $\mu_1 \leq \mu_2$, then $T(\mu_1) \leq T(\mu_2)$. Fix any L and then $\overline{T}_1(\mu_1, \mu_2; L) = M(\mu_1) + T(\mu_2) - M(\mu_2) \geq M(\mu_1) + T(\mu_1) - M(\mu_1) = T(\mu_1)$. Similarly, $\overline{T}_2(\mu_1, \mu_2; L) = T(\mu_1) - M(\mu_1) + M(\mu_2) \geq T(\mu_1) - M(\mu_1) + M(\mu_1) = T(\mu_1)$. ■

Lemma 2. *Fix two univariate belief-distributions $\tau_1, \tau_2 \in \Delta(\Theta)$ and a full-support prior $\pi \in \Delta(\Theta)$. Suppose that (i) $\mathbb{E}_{\tau_i}[\mu_i] = \pi$, and (ii) $\operatorname{supp}_i \tau_i$ is totally ordered by first-order stochastic dominance, then for every $L = 0, \ldots, K$*

$$T_i(\mu_i) - M_i(\mu_i, L) \leq T_i(\mu_i) \left[1 - \Pi(L)\right]. \tag{12}$$

Furthermore, $\underline{T}_2(\mu_1, \mu_2; L) \leq \max_L \underline{T}_1(\mu_1, \mu_2; L)$.

Proof. Using the total order and (i), for every L and every $\mu_i \in \operatorname{supp} \tau_i$

$$\mathbb{E}_{\tau_i}[\sum_{k \leq L} \mu_i'(\theta_k) | \mu_i' \leq \mu_i] \mathbb{P}(\mu_i' \leq \mu_i) + \mathbb{E}_{\tau_i}[\sum_{k \leq L} \mu_i'(\theta_k) | \mu_i' > \mu_i] \mathbb{P}(\mu_i' > \mu_i) = \sum_{k \leq L} \mathbb{E}_{\tau_i}[\mu_i'(\theta_k)] = \Pi(L),$$

---

[67] Recall that a summation over an empty set is zero.

and by first-order stochastic dominance we also know that

$$\mathbb{E}_{\tau_i}[\sum_{k\leq L}\mu_i'(\theta_k)|\mu_i'\leq\mu_i]\geq\sum_{k\leq L}\mu_i(\theta_k)\geq\mathbb{E}_{\tau_i}[\sum_{k\leq L}\mu_i'(\theta_k)|\mu_i'>\mu_i].$$

Thus, $\Pi(L)\leq\mathbb{E}_{\tau_i}[\sum_{k\leq L}\mu_i'(\theta_k)|\mu_i'\leq\mu_i]={}^{M_i(\mu_i,L)}/_{T_i(\mu_i)}$, which implies the first inequality in Equation 12.

For the second part, the inequality Equation 12 gives

$$\begin{aligned}\underline{T}_2(\mu_1,\mu_2;L)&=T_1(\mu_1)-M_1(\mu_1,L)+T_2(\mu_1)-M_2(\mu_1,L)-[1-\Pi(L)]\\&\leq T_1(\mu_1)\,[1-\Pi(L)]+T_2(\mu_2)\,[1-\Pi(L)]-[1-\Pi(L)]\\&\leq T_1(\mu_1)+T_2(\mu_2)-1\leq\max_L\underline{T}_1(\mu_1,\mu_2;L).\end{aligned}$$

■

Proof of Proposition 2. Consider τ that is induced by an information structure. Since it is induced by an information structure, there exists $\lambda\in\Delta(\Delta(\Theta)\times\Delta(\Theta)\times\Theta)$ with marginal distribution on $\Delta(\Theta)\times\Delta(\Theta)$ given by τ and properties (1) and (2) as stated in Theorem 1. By Equation 5 the marginal conditions (1) of Proposition 2 are satisfied. Now, define

$$\lambda_1(\mu_1,\theta)=\mu_1(\theta)\sum_{\mu_2}\tau(\mu_1,\mu_2),\quad\text{and}\quad\lambda_2(\mu_2,\theta)=\mu_2(\theta)\sum_{\mu_1}\tau(\mu_1,\mu_2).$$

Since τ is a (bivariate) marginal of λ and due to (2) of Theorem 1, λ_1 and λ_2 are the two other bivariate marginals of λ.

Now, by Joe (1997, Theorem 3.11), λ with the given bivariate marginals exists only if for every $L=0,\dots,K$ and every $\mu_1,\mu_2\in\Delta(\Theta)$,

$$\begin{aligned}\max\{&0,T(\mu_1,\mu_2)-[T_1(\mu_1)-M_1(\mu_1,L)]\,,T(\mu_1,\mu_2)-[T_2(\mu_2)-M_2(\mu_2,L)]\,,\\&M_1(\mu_1,L)+M_2(\mu_2,L)-\Pi(L)\}\\&\leq\\\min\{&T(\mu_1,\mu_2),M_1(\mu_1,L),M_2(\mu_2,L),\\&T(\mu_1,\mu_2)+[1-\Pi(L)]-[T_1(\mu_1)-M_1(\mu_1,L)]-[T_2(\mu_2)-M_2(\mu_2,L)]\}\,.\end{aligned}\tag{13}$$

Now, it remains to prove that existence of λ and Equation 13 imply the bounds $\overline{T}\succsim T\succsim\underline{T}$. By way of contradiction, suppose the bounds are violated, then we have (at least) one of the following cases for some $L=0,\dots,K$ and some $\mu_1,\mu_2\in\Delta(\Theta)$

- If $T(\mu_1,\mu_2)>\overline{T}_1(\mu_1,\mu_2;L)=M_1(\mu_1,L)+T_2(\mu_2)-M_2(\mu_2,L)$, then

$$T(\mu_1,\mu_2)-[T_2(\mu_1)-M_2(\mu_1,L)]>M_1(\mu_2,L).$$

- If $T(\mu_1, \mu_2) > \overline{T}_2(\mu_1, \mu_2; L) = T_1(\mu_2) - M_1(\mu_2, L) + M_2(\mu_1, L)$, then

$$T(\mu_1, \mu_2) - [T_1(\mu_1) - M_1(\mu_1, L)] > M_2(\mu_2, L).$$

In either case, Equation 13 is violated. Similarly, if $T(\mu_1, \mu_2) < \underline{T}_1(\mu_1, \mu_2; L) = M_1(\mu_1, L) + M_2(\mu_2, L) - \Pi(L)$ or $T(\mu_1, \mu_2) < \underline{T}_2(\mu_1, \mu_2; L) = T_1(\mu_1) - M_1(\mu_1, L) + T_2(\mu_2) - M_2(\mu_2, L) - [1 - \Pi(L)]$ then Equation 13 is violated.

Thus, if the bounds are not satisfied at any point, Equation 13 is violated. This means that there is no trivariate distribution with the marginals given by τ, λ_1, and λ_2. However, this is in contradiction with the existence of λ. ■

Proof of Proposition 3. I will only prove the case of supermodularity. Consider $\mu_i \in \Delta(\Theta)$ and $\eta : \Theta \to \Delta(A_{-i})$ such that $\operatorname{supp} \nu(\cdot|\theta) \subseteq BFR_{-i}$ for all $\theta \in \Theta$. Since supermodularity is preserved under summation (i.e. expectation) , the best-reply is increasing (in the strong set-order) in first-order beliefs μ_i (holding η fixed), see van Zandt and Vives (2007). Thus the robust prediction correspondence is increasing (in the strong-set order). Now, let $b_i(\mu_i) = \min\{a_i \in R_i(\mu_i)\}$, which is increasing in μ_i. Because v is increasing, $\nu(\mu_1, \mu_2) = v(b_1(\mu_1), b_2(\mu_2))$. If μ_1 first-order stochastic dominates μ_1', then $b_1(\mu_1) \geq b_1(\mu_1')$. Thus,

$$\nu(\mu_1, \mu_2) - \nu(\mu_1', \mu_2) = v(b_1(\mu_1), a_2(\mu_2)) - v(b_1(\mu_1'), a_2(\mu_2)),$$

is increasing in μ_2 because $b_2(\cdot)$ is and v is supermodular. ■

Proof of Proposition 4. One direction is obvious. For the other fix an information structure I. Then, define[68] $S_i^1 = \{s_i \in S_i : a_i^1 \in R_i(s_i)\}$ and for $1 < k \leq J_i$

$$S_i^k = \left\{s_i \in S_i : a_i^k \in R_i(s_i) \text{ and } a_i^l \notin R_i(s_i) \text{ for all } l < k\right\}.$$

Now, let $\hat{S}_i = \left\{a_i^j \in A_i : S_i^j \neq \emptyset\right\} \subseteq A_i$ and set the signal distribution to $\hat{\Psi}(a_1^{j_1}, a_2^{j_2}|\theta) = \sum_i \sum_{s_i \in S_i^{j_i}} \Psi(s_1, s_2|\theta)$. Now, for a given $a_i^j \in \hat{S}_i$, the induced first-order belief (call it μ) will be a convex combination of beliefs (i.e μ_{s_i} for $s_i \in S_i^j$). Since these beliefs are totally ordered, one of these beliefs is the lowest according to first-order stochastic dominance; call it $\underline{\mu}$. Thus, the convex combination (i.e. μ) is also greater than $\underline{\mu}$. As shown in Equation A, the robust-prediction correspondence is increasing. Thus, $R_i(\underline{\mu}) \leq R_i(\mu)$ in the strong set order.

By construction, we have $a_i^m \notin R_i(\underline{\mu})$ for $m < j$ implying that $a_i^m \notin R_i(\mu)$. Furthermore, we know that a_i^j is conceivable for each μ_{s_i} for $s_i \in S_i^j$. That is, for each such $s_i \in S_i^j$ there exists $\eta_{s_i}(\cdot|\cdot) : \Theta \to \Delta(A_{-i})$ such that $a_i^j \in BR_i(\mu_{s_i} \circ \eta_{s_i})$. Consider[69] $\tilde{\mu} = \sum_{s_i \in \hat{S}_i^{\mathcal{A}_i}} \lambda_{s_i} \mu_{s_i} \circ \eta_{s_i}$, which has marginal μ by construction. And

---

[68] The superscripts refer to the indexing set of the actions, i.e. $A_i = \{a_i^1, \ldots, a_i^{J_i}\}$.

[69] Let λ_{s_i} denote the coefficients of the convex combination.

since a_i^j is a best-reply to each belief separately, it's also a best-reply to the convex combination. Proving $a_i^j \in R_i(\mu)$.

So we established $a_i^j \in R_i(a_i^j)$ and $a_i^m \notin R_i(a_i^j)$, for all $m < j$. Thus, by Definition 7 for any $(a_1, a_2) \in \hat{S}_1 \times \hat{S}_2$ $\min_{a_i' \in R_i(a_i)} v(a_1', a_2') = v(a_1, a_2)$. Proving that the information structure is direct. That the values are the same follows trivially from the construction. ■

References

Alonso, R. and O. Câmara (2016): "Bayesian Persuasion with Heterogeneous Priors," *Journal of Economic Theory*, 165, 672–706.

Arieli, I., Y. Babichenko, F. Sandomirskiy, and O. Tamuz (2020): "Feasible Joint Posterior Beliefs," Mimeo.

Artemov, G., T. Kunimoto, and R. Serrano (2013): "Robust Virtual Implementation: Toward a Reinterpretation of the Wilson Doctrine," *Journal of Economic Theory*, 148, 424 – 447.

Aumann, R. J. (1987): "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, 55, 1–18.

Battigalli, P. (2003): "Rationalizability in infinite, dynamic games with incomplete information," *Research in Economics*, 57, 1–38.

Battigalli, P., A. Di Tillio, E. Grillo, and A. Penta (2011): "Interactive Epistemology and Solution Concepts for Games with Asymmetric Information," *The B.E. Journal of Theoretical Economics*, 11.

Battigalli, P. and M. Siniscalchi (2003): "Rationalization and Incomplete Information," *Advances in Theoretical Economics*, 3.

——— (2007): "Interactive Epistemology in Games with Payoff Uncertainty," *Research in Economics*, 61, 165–184.

Beauchêne, D., J. Li, and M. Li (2019): "Ambiguous Persuasion," *Journal of Economic Theory*, 179, 312–365.

Bergemann, D. and S. Morris (2005): "Robust Mechanism Design," *Econometrica*, 73, 1771–1813.

——— (2013): "Robust Predictions in Games With Incomplete Information," *Econometrica*, 81, 1251–1308.

——— (2016): "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games," *Theoretical Economics*, 11, 487–522.

——— (2017): "Belief-free Rationalizability and Informational Robustness," *Games and Economic Behavior*, 104, 744–759.

——— (2019): "Information Design: A Unified Perspective," *Journal of Economic Literature*, 57, 44–95.

Bergemann, D. and M. Pesendorfer (2007): "Information Structures in Optimal Auctions," *Journal of Economic Theory*, 137, 580–609.

Börgers, T. and J. Li (2019): "Strategically Simple Mechanisms," *Econometrica*, 87, 2003–2035.

Brandenburger, A. and E. Dekel (1987): "Rationalizability and Correlated Equilibria," *Econometrica*, 55, 1391–1402.

Bresnahan, T. and J. Levin (2012): "Vertical Integration and Market Structure," in *The Handbook of Organizational Economics*, ed. by R. Gibbons and J. Roberts, Princeton University Press, 853–890.

Brocas, I. and J. D. Carrillo (2007): "Influence Through Ignorance," *The RAND Journal of Economics*, 38, 931–947.

Calzolari, G. and A. Pavan (2006): "On the Optimality of Privacy in Sequential Contracting," *Journal of Economic Theory*, 130, 168–204.

Cambanis, S., G. Simons, and W. Stout (1976): "Inequalities for $\mathbb{E}k(X, Y)$ when the Marginals are Fixed," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 36, 285–294.

Carroll, G. (2016): "Informationally Robust Trade and Limits to Contagion," *Journal of Economic Theory*, 166, 334–361.

——— (2019): "Robustness in Mechanism Design and Contracting," *Annual Review of Economics*, 11, 139–166.

Dekel, E., D. Fudenberg, and S. Morris (2007): "Interim Correlated Rationalizability," *Theoretical Economics*, 2, 15–40.

Dequiedt, V. and D. Martimort (2015): "Vertical Contracting with Informational Opportunism," *American Economic Review*, 105, 2141–2182.

Ely, J. C. (2017): "Beeps," *American Economic Review*, 107, 31–53.

Eső, P. and B. Szentes (2007): "Optimal Information Disclosure in Auctions and the Handicap Auction," *The Review of Economic Studies*, 74, 705–731.

Forges, F. (1993): "Five Legitimate Definitions of Correlated Equilibrium in Games with Incomplete Information," *Theory and Decision*, 35, 277–310.

——— (2006): "Correlated Equilibrium in Games with Incomplete Information Revisited," *Theory and Decision*, 61, 329–344.

Galperti, S. (2019): "Persuasion: The Art of Changing Worldviews," *American Economic Review*, 109, 996–1031.

Germano, F. and P. Zuazo-Garin (2017): "Bounded Rationality and Correlated Equilibria," *International Journal of Game Theory*, 46, 595–629.

Goldstein, I. and C. Huang (2016): "Bayesian Persuasion in Coordination Games," *American Economic Review*, 106, 592–596.

Harsanyi, J. C. (1968): "Games with Incomplete Information Played by 'Bayesian' Players, I-III. Part I. The Basic Model," *Management Science*, 14, 320–334.

Hoshino, T. (2019): "Multi-Agent Persuasion: Leveraging Strategic Uncertainty," Mimeo.

Inostroza, N. and A. Pavan (2018): "Persuasion in Global Games with Application to Stress Testing," Mimeo.

Joe, H. (1997): *Multivariate Models and Multivariate Dependence Concepts*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis.

Kamenica, E. (2019): "Bayesian Persuasion and Information Design," *Annual Review of Economics*, 11, 249–272.

Kamenica, E. and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101, 2590–2615.

Kosterina, S. (2020): "Persuasion with Unknown Beliefs," Mimeo.

Liu, Q. (2015): "Correlation and Common Priors in Games with Incomplete Information," *Journal of Economic Theory*, 157, 49–75.

Mathevet, L., J. Perego, and I. Taneva (2020): "On Information Design in Games," *Journal of Political Economy*, 128, 1370–1404.

McAfee, R. P. and M. Schwartz (1994): "Opportunism in Multilateral Vertical Contracting: Nondiscrimination, Exclusivity, and Uniformity," *American Economic Review*, 84, 210–230.

Meyer, M. and B. Strulovici (2015): "Beyond Correlation: Measuring Interdependence Through Complementarities," Mimeo.

Myerson, R. B. (1982): "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems," *Journal of Mathematical Economics*, 10, 67–81.

Nelsen, R. B. (2006): *An Introduction to Copulas*, Springer Series in Statistics, New York, USA: Springer-Verlag.

Ollár, M. and A. Penta (2017): "Full Implementation and Belief Restrictions," *American Economic Review*, 107, 2243–77.

Pahlke, M. (2020): "A Note on Dynamic Consistency in Ambiguous Persuasion," Center for Mathematical Economics Working Papers 611, Center for Mathematical Economics, Bielefeld University.

Pearce, D. G. (1984): "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, 52, 1029–1050.

Piermont, E. and P. Zuazo-Garin (2020): "Heterogeneously Perceived Incentives in Dynamic Environments: Rationalization, Robustness and Unique Selections," Mimeo.

Segal, I. and M. D. Whinston (2012): "Property Rights," in *The Handbook of Organizational Economics*, ed. by R. Gibbons and J. Roberts, Princeton University Press, 100–158.

Shaked, M. and J. Shanthikumar (2007): *Stochastic Orders*, Springer Series in Statistics, Springer New York.

Taneva, I. (2019): "Information Design," *American Economic Journal: Microeconomics*, 11, 151–185.

Tang, Q. (2015): "Hierarchies of Beliefs and the Belief-invariant Bayesian Solution," *Journal of Mathematical Economics*, 59, 111–116.

Tchen, A. H. (1980): "Inequalities for Distributions with Given Marginals," *Annals of Probability*, 8, 814–827.

van Zandt, T. and X. Vives (2007): "Monotone Equilibria in Bayesian Games of Strategic Complementarities," *Journal of Economic Theory*, 134, 339–360.

Wald, A. (1949): "Statistical Decision Functions," *Ann. Math. Statist.*, 20, 165–205.

Wilson, R. (1987): "Game-Theoretic Analyses of Trading Processes," in *Advances in Economic Theory: Fifth World Congress*, ed. by T. Bewley, Cambridge University Press, Econometric Society Monographs, 33–70.

B ONLINE APPENDIX

B.1 A Foundation for the Individual Robust Predictions

In this section, I provide a foundation for the individual robust predictions in a similar spirit as the literature on informational robustness and Bayes Correlated Equilibrium. First, I need a result relating BFR to robustness across all information structures and across all Bayes-Nash equilibria (Proposition 5). This part is closet to the literature on informational robustness in the sense it takes the perspective of an outside observer. Second, I will give a foundation for the individual robust-predictions by adding back the marginal information structure of Player 1 (Theorem 2). Thus, this can be seen as a robustness from the player's perspective because he knows his marginal information structure. Since these foundations rely on non-common priors, I also need to take care of zero probability events. This is in contrast to the analysis in the main text and requires different definitions. Whenever zero probability events can be ruled out, all the following definitions reduce to the definitions of Section 2.

B.1.1 Robustness for an Outside Observer

Starting with an economic environment, a Bayesian game is obtained by adding priors for each player $\pi_i \in \Delta(\Theta)$ and specifying a (grand) information structure with possible heterogeneous signal functions.

Definition 8. *Fix an economic environment $\mathcal{E}$. A (grand)* generalized information structure *(for $\mathcal{E}$) is $I = \langle (S_i, \Psi_i)_{i \in N} \rangle$, where for each player $i \in N$,*

1. *S_i is a finite set of signals, and*
2. *$\Psi_i : \Theta \to \Delta(S_1 \times S_2)$ is a conditional signal distribution.*

A Bayesian game *$G = \langle \mathcal{E}, I, (\pi_i)_{i \in N} \rangle$ is given by (i) an economic environment $\mathcal{E}$, (ii) a generalized information structure I, and (iii) a prior $\pi_i \in \Delta(\Theta)$ for each player $i \in N$.*

A generlized information structure together with the two priors gives rises to a standard type space á la Harsanyi (1968) but without a common prior.[70] Without common priors and signal distributions the definition of equilibrium needs to account for zero probability events. For complete information games, Brandenburger and Dekel (1987) introduced *a posteriori equilibrium* to rule out the play of dominated actions after a zero probability events. The definition of equilibrium in this paper will be an extension to incorporate uncertainty about the states of nature. But first, we need to introduce a tool to define beliefs even in case of zero probability events.

---

[70] Recently, Piermont and Zuazo-Garin (2020) allow for even more disagreement by allowing for lack of common knowledge of the Harsanyi type space and the states of nature.

Definition 9. *Fix an economic environment* $\mathcal{E}$*, a player* i*, a prior* $\pi_i \in \Delta(\Theta)$ *and a generalized information structure* I*. A* conditional probability system (CPS) *for* (π_i, I) *is a mapping* $\mu_i : S_i \to \Delta(\Theta \times S_{-i})$ *such that for every* $(\theta, s_i, s_{-i}) \in \Theta \times S_1 \times S_2$*,*

$$\mu_i(\theta, s_{-i}|s_i) \left[\sum_{\theta', s'_{-i}} \pi_i(\theta')\Psi_i(s_i, s'_{-i}|\theta') \right] = \pi_i(\theta)\Psi_i(s_i, s_{-i}|\theta).$$

That is, a CPS defines beliefs about the state of nature and the opponent's signal realization for every signal relation of the given player. In addition, the beliefs have to be updated via Bayes' rule whenever possible. To formally state the appropriate version of equilibrium, it only remains to define strategies. A (behavioral) strategy for player i in a Bayesian Game G is a mapping $\beta_i : S_i \to \Delta(A_i)$.

Definition 10. *Fix an economic environment* $\mathcal{E}$*, priors* $\pi_i \in \Delta(\Theta)$ *for each player* $i \in N$*, and an information structure* I*. A* Bayes-Nash equilibrium (BNE) *for* (π_1, π_2, I) *is a tuple* (β_i, μ_i) *for each player* $i \in I$ *such that*

1. β_i *is a strategy,*
2. μ_i *is a CPS for* (π_i, I)*, and*
3. β_i *is optimal (given* μ_i *and* β_{-i}*), i.e. for each* $s_i \in S_i$

$$a_i \in \operatorname{supp} \beta_i(\cdot|s_i) \implies a_i \in \arg\max_{a'_i} \sum_{\theta, s_{-i}, a_{-i}} \mu_i(\theta, s_{-i}|s_i)\beta_{-i}(a_{-i}|s_{-i})u_i(a'_i, a_{-i}, \theta).$$

Let $BNE(\pi_1, \pi_2, I)$ *be the set of all BNEs for* (π_1, π_2, I)*.*[71]

Now, the first result states that belief-free rationalizability characterizes all actions that can be played in any Bayes-Nash equilibrium for any information structure (and any prior beliefs). Thus, without making any assumptions about the information structure an outside observer can not make any prediction that is a refinement of belief-free rationalizability. In this sense, belief-free rationalizability is robust to the specification of the (generalized) information structure.[72]

Proposition 5. *Fix an economic environment* $\mathcal{E}$*. For every player* i*,* $a_i \in BFR_i$ *iff there exists priors* (π_1, π_2)*, an information structure* I *and a signal* $s_i \in S_i$ *such that* $a_i \in \operatorname{supp} \beta_i(\cdot|s_i)$ *for some* $\beta_i \in BNE_i(\pi_1, \pi_2, I)$*.*

---

[71] The dependence on the economic environment is suppressed in this notation since it will be fixed throughout. Furthermore, I will slightly abuse notation and write $\beta = (\beta_1, \beta_2) \in BNE(\pi_1, \pi_2, I)$ if there exists CPS' $\mu = (\mu_1, \mu_2)$ such that $(\beta, \mu) \in BNE(\pi_1, \pi_2, I)$. Similarly, we will write $\beta_i \in BNE_i(\pi_1, \pi_2, I)$ if there exists β_{-i} and μ such that $(\beta_1, \mu_1, \beta_2, \mu_2) \in BNE(\pi_1, \pi_2, I)$.

[72] Bergemann and Morris (2017, Section 4.5) informally mention a result along these lines. Battigalli and Siniscalchi (2003, Proposition 4.2 and 4.3) prove a similar result in a slightly different setting.

B.1.2 Robustness from the Player's Perspective[73]

Now, we add back the marginal information structure of Player 1 (see Definition 1). Here as well, we need to take care of zero probability events and therefore rational-extended beliefs are not appropriate anymore. A version of a conditional probability system is needed again. Although, now it should only capture beliefs about the state of nature.

Definition 11. *Fix an economic environment $\mathcal{E}$, a prior $\pi_1 \in \Delta(\Theta)$ and a marginal information structure I_1. A* marginal conditional probability system (mCPS) *for (π_1, I_1) is a mapping $\mu_1 : S_1 \to \Delta(\Theta)$ such that for every $(\theta, s_1) \in \Theta \times S_1$,*

$$\mu_1(\theta|s_1)\left[\sum_{\theta'} \pi_1(\theta')\psi_1(s_1|\theta')\right] = \pi_1(\theta)\psi_1(s_1|\theta).$$

Similar to rational-extended beliefs, mCPS need to be extended as well.

Definition 12. *Fix an economic environment $\mathcal{E}$, a prior $\pi_1 \in \Delta(\Theta)$ and a marginal information structure I_1. A* rational-extended conditional probability system (rCPS) *for (π_1, I_1) is a mapping $\mu_1 : S_1 \to \Delta(\Theta \times A_2)$ such that*

1. *$\tilde{\mu}_1 = (\mu_1(\cdot|s_1))_{s_1 \in S_1}$ is a mCPS for (π_1, I_1), where $\tilde{\mu}_1(\cdot|s_1) = \operatorname{marg}_\Theta \mu_1(\cdot|s_1)$ for all $s_1 \in S_1$, and*
2. *for all $s_1 \in S_1$, $\operatorname{supp} \mu_1(\cdot|s_1) \subseteq \Theta \times BFR_2$.*

Finally, these rCPS' allow to define the individual robust prediction even with zero probability events.

Definition 13. *Fix an economic environment $\mathcal{E}$, a prior $\pi_1 \in \Delta(\Theta)$, and a marginal information structure I_1. A pure strategy $b : S_1 \to A_1$ is* conceivable *for (π_1, I_1) if there exists a rCPS μ_1 for (π_1, I_1) such that b is optimal given μ_1, i.e. for each $s_1 \in S_1$,*

$$b(s_1) \in \arg\max_{a_1'} \sum_{\theta, a_2} \mu_1(\theta, a_2|s_1) u_1(a_1', a_2, \theta).$$

The individual robust prediction *is the set of all conceivable strategies and is denoted by $R_1(I_1, \pi_1)$.*

The goal of this section is to provide a foundation of the individual robust predictions. That is, it should capture they idea of informational robustness across all information structures of the opponent (fixing the marginal information structure of the player). This leads to the idea of an extended information structure.

---

[73] This subsection is described from the perspective of player 1. It applies verbatim to player 2 by switching the player indices.

Definition 14. *Fix an economic environment* $\mathcal{E}$ *and a marginal information structure* $I_1 = \langle S_1, \psi_1 \rangle$*. An* extended information structure *(for* I_1*) is* $I = \langle (\hat{S}_i, \Psi_i)_{i \in N} \rangle$ *such that*

1. I *is a generalized information structure,*
2. $S_1 \subseteq \hat{S}_1$*, and*
3. $\operatorname{marg}_{S_1} \Psi_1(\cdot|\theta) = \psi_1(\cdot|\theta)$*, for all* $\theta \in \Theta$*.*

Let $\mathcal{I}(I_1)$ *be the set of extending information structures for* I_1*.*

Condition (1) ensures that an extended information structure is indeed a generalized information structure, whereas conditions (2) and (3) make sure that the extended information structure incorporates the marginal information structure of Player 1. A natural interpretation of this definition is that Player 1 conjectures a grand information structure for given economic environment so that she can analyze the resulting Bayesian game. However, since she knows exactly what information she gets about the state of nature, she uses this knowledge to rule out information structures which do not align with her marginal information structure. Indeed, the individual robust prediction correspond to all strategies that are conceivable across all such conjectures. This means that for each conceivable strategy there is an extending information structure (and a conjectured prior for the opponent)[74] and a corresponding Bayes-Nash equilibrium where this strategy is played.

Theorem 2. *Fix an economic environment* $\mathcal{E}$*, prior* $\pi_1 \in \Delta(\Theta)$*, and a marginal information structure* I_1*.* $b \in R_1(I_1, \pi_1)$ *iff there exists an extending information structure* $I \in \mathcal{I}(I_1)$*, a prior* $\pi_2 \in \Delta(\Theta)$*, and a corresponding BNE* β_i *such that* $b(s_i) \in \operatorname{supp} \beta_i(\cdot|s_i)$ *for all* $s_i \in S_i$*.*

Theorem 2 constitutes the main result of this section, because it provides an informational robustness foundation for the individual robust predictions.

B.1.3 Proofs for Subsection B.1

Since actions are finite it is immediate that the BFR procedure as stated in Equation 1 needs to stop at a finite number of iterations, which directly gives the usual, but convenient, fixed-point definition of belief-free rationalizability:

$$BFR_i = \Bigg\{ a_i \in A_i : \exists \mu_i \in \Delta(\Theta \times A_{-i}) \text{ s.t.}$$
$$(1) \operatorname{supp} \mu_i \subseteq \Theta \times BFR_{-i}, \tag{14}$$
$$(2)\, a_i \in \arg\max_{a_i' \in A_i} \sum_{\theta, a_{-i}} \mu_i(\theta, a_{-i}) u_i(a_i', a_{-i}, \theta) \Bigg\}.$$

---

[74] Recall that the economic environment does not specify priors.

Proposition 5. *Fix an economic environment $\mathcal{E}$. For every player i, $a_i \in BFR_i$ iff there exists priors (π_1, π_2), an information structure I and a signal $s_i \in S_i$ such that $a_i \in \operatorname{supp} \beta_i(\cdot|s_i)$ for some $\beta_i \in BNE_i(\pi_1, \pi_2, I)$.*

Proof. For given priors (π_1, π_2), information structure I, consider a signal s_i such that $a_i \in \operatorname{supp} \beta_i(\cdot|s_i)$ for some $(\beta_i, \hat{\mu}_i, \beta_{-i}, \hat{\mu}_{-i}) \in BNE(\pi_1, \pi_2, I)$. We show that $a_i \in BFR_i$ by induction, i.e. $a_i \in BFR_i^n$ for every n. The statement is trivial for $n = 0$. So assume the statement is true for $n \geq 0$. Consider the following belief $\mu_i \in \Delta(\Theta \times S_{-i} \times A_{-i})$ defined by

$$\mu_i(\theta, s_{-i}, a_{-i}) = \hat{\mu}_i(\theta, s_{-i}|s_i)\beta_{-i}(a_{-i}|s_{-i}),$$

Note that a_i is a best-reply to μ_i by the definition of BNE.

Let $m_i = \operatorname{marg}_{\Theta \times A_{-i}} \mu_i$, then we have

$$m_i(\theta, a_{-i}) > 0 \implies \mu_i(\theta, s_{-i}, a_{-i}) > 0 \text{ for some } s_{-i} \text{ such that } \beta_{-i}(a_{-i}|s_{-i}) > 0,$$

and by the induction hypothesis $a_{-i} \in BFR_{-i}^n$. Hence, $\operatorname{supp} \mu_i \subseteq \Theta \times BFR_{-i}^n$. Since, a_i is a best-reply to μ_i, $a_i \in BFR_i^{n+1}$.

Conversely, for every $a_i \in BFR_i$, there is a justifying belief $\mu_i^{a_i}$ satisfying (1) and (2) from BFR.[75] Then define a prior by

$$\pi_i(\theta) = \sum_{a_i \in BFR_i} \frac{\sum_{a_{-i}} \mu_i^{a_i}(\theta, a_{-i})}{|BFR_i|}$$

and consider the following information structure: $S_i = BFR_i$ and

$$\Psi_i(a_i, a_{-i}|\theta) = \frac{\mu_i^{a_i}(\theta, a_{-i})}{\pi_i(\theta)} |BFR_i|^{-1},$$

if $\pi_i(\theta) > 0$ and arbitrary otherwise. Note that for every $a_i \in BFR_i$, we have

$$\sum_{a_{-i}, \theta} \pi_i(\theta)\Psi_i(a_i, a_{-i}|\theta) = |BFR_i|^{-1} > 0,$$

so that the CPS is entirely determined by Bayesian updating.

---

[75] Here, the equivalent fixed-point definition of BFR stated in Equation 14 is used.

Now, fix $a_i \in BFR$ and consider the obedient strategies, i.e. $\beta_i(a_i|s_i) = 1[s_i = a_i]$. Then,

$$\begin{aligned} a_i &\in \arg\max_{a_i' \in A_i} \sum_{\theta, a_{-i}} \mu_i^{a_i}(\theta, a_{-i}) u_i(a_i', a_{-i}, \theta) \\ &\in \arg\max_{a_i' \in A_i} \sum_{\theta, a_{-i}} \Psi_i(a_i, a_{-i}|\theta)\pi_i(\theta) u_i(a_i', a_{-i}, \theta) \\ &\in \arg\max_{a_i' \in A_i} \sum_{\theta, a_{-i}, s_{-i}} \pi_i(\theta)\Psi_i(a_i, s_{-i}|\theta)\beta_{-i}(a_{-i}|s_{-i}) u_i(a_i', a_{-i}, \theta), \end{aligned}$$

so that the obedient strategy of i is indeed a best-reply to the obedient strategy of the other player (given the information structure). That is, β (and the CPS derived from Bayesian updating) constitute a BNE. ■

Proof. Fix a marginal information structure I_i and prior π_1.

For a given extending information structure $I \in \mathcal{I}(I_1)$, a prior π_2, and a corresponding BNE $(\beta, \hat{\mu})$ consider any selection $b(s_1) \in \operatorname{supp} \beta_i(\cdot|s_1)$ for all $s_1 \in S_1$. For every $s_2 \in \hat{S}_2$ and every $a_2 \in \operatorname{supp} \beta_2(\cdot|s_2)$, $a_2 \in BFR_2$ by Proposition 5. For each $s_1 \in S_1$ consider beliefs $\mu_1(\cdot|s_1) \in \Delta(\Theta \times A_2)$ defined by

$$\mu_1(\theta, a_2|s_1) = \sum_{\hat{s}_2} \hat{\mu}_1(\theta, \hat{s}_2|s_1)\beta_2(a_2|\hat{s}_2).$$

Then $\mu_1(\theta, a_2|s_1) > 0$ implies that there exists $s_2 \in \hat{S}_2$ such that $\beta_2(a_2|s_2) > 0$, which implies that $a_2 \in BFR_2$. Hence, $\operatorname{supp} \mu_1(\cdot|s_1) \subseteq \Theta \times BFR_2$ for every $s_1 \in S_1$. Furthermore, let $\tilde{\mu}_1(\cdot|s_1) = \sum_{a_2} \mu_1(\cdot, a_2|s_1)$ for every $s_1 \in S_1$, then

$$\begin{aligned} \tilde{\mu}_1(\theta|s_1) \left[\sum_{\theta'} \pi_1(\theta')\psi_1(s_1|\theta')\right] &= \sum_{a_2, \hat{s}_2} \hat{\mu}_1(\theta, \hat{s}_2|s_1)\beta_2(a_2|\hat{s}_2) \left[\sum_{\theta'} \pi_1(\theta')\psi_1(s_1|\theta')\right] \\ &= \sum_{\hat{s}_2} \hat{\mu}_1(\theta, \hat{s}_2|s_1) \left[\sum_{\theta'} \pi_1(\theta')\psi_1(s_1|\theta')\right] \\ &= \sum_{\hat{s}_2} \hat{\mu}_1(\theta, \hat{s}_2|s_1) \left[\sum_{\theta', \hat{s}_2'} \pi_1(\theta')\Psi_1(s_1, \hat{s}_2'|\theta')\right] \\ &= \sum_{\hat{s}_2} \pi_1(\theta)\Psi_1(s_1, \hat{s}_2|\theta) = \pi_1(\theta)\psi_1(s_1|\theta), \end{aligned}$$

where the third and last equality use property 3 of an extending information structure (Definition 14). The fourth equality follows from $\hat{\mu}_1$ being a CPS for (π_1, I) (see Definition 9). Thus, μ_1 is a rCPS and by construction $b(s_i)$ is a best-reply to $\mu_1(\cdot|s_1)$ for each $s_1 \in S_1$. This proves that b is conceivable.

Conversely, consider $b \in R_1(I_1, \pi_1)$. By definition of R_1 there exists a rCPS μ_1 such that b is optimal given μ_1. Define $BFR_1^- = BFR_1 \setminus \cup_{s_1 \in S_1}\{b(s_1)\}$ and set $\hat{S}_1 = S_1 \cup BFR_1^-$ and $\hat{S}_2 = BFR_2$.

For player 1, define a conditional signal distribution as follows.

$$\Psi_1(s_1, \hat{s}_2|\theta) = \frac{\mu_1(\hat{s}_2, \theta|s_1)}{\pi_1(\theta)} \sum_{\tilde{\theta}} \pi_1(\tilde{\theta})\psi_1(s_1|\tilde{\theta}), \quad \text{for all } s_1 \in S_1, \text{ and}$$

$$\Psi_1(a_1, \hat{s}_2|\theta) = 0 \quad \text{for all } a_1 \in BFR_1^-,$$

if $\pi_i(\theta) > 0$ and arbitrary otherwise. Since the marginal of μ_1 on θ is a mCPS, we have that $\operatorname{marg}_{S_1} \Psi_1 = \psi_1$.

Since $\hat{S}_2 \subseteq BFR_2$, there is a belief $\mu_2^{a_2}$ satisfying (1) and (2) from BFR[76] for each $a_2 \in \hat{S}_2$. Then define a prior by

$$\pi_2(\theta) = \sum_{a_2 \in \hat{S}_2} \frac{\sum_{a_1} \mu_2^{a_2}(\theta, a_1)}{|\hat{S}_2|}$$

and consider the following conditional signal distribution for player 2.

$$\Psi_2(s_1, a_2|\theta) = \frac{1}{|\hat{S}_2|} \frac{1}{|b^{-1}(b(s_1))|} \frac{\mu_2^{a_2}(b(s_1), \theta)}{\pi_2(\theta)}, \quad \text{for all } s_1 \in S_1, \text{ and}$$

$$\Psi_2(a_1, a_2|\theta) = \frac{1}{|\hat{S}_2|} \frac{\mu_2^{a_2}(a_1, \theta)}{\pi_2(\theta)}, \quad \text{for all } a_1 \in BFR_1^-,$$

if $\pi_2(\theta) > 0$ and arbitrary otherwise.

Since $\sum_{\hat{s}_1, \theta} \Psi_2(\hat{s}_1, s_2|\theta)\pi_2(\theta) = |\hat{S}_2|^{-1} > 0$ for all $s_2 \in \hat{S}_2$, the CPS for player 2 is determined by Bayesian updating. For player 1, consider the CPS that is defined by Bayesian updating if $\sum_{\tilde{\theta}, \hat{s}_2} \pi_1(\tilde{\theta})\Psi_1(s_1, \hat{s}_2|\tilde{\theta}) = \sum_{\tilde{\theta}} \pi_1(\tilde{\theta})\psi_1(s_1|\tilde{\theta}) > 0$ and in the other case for $s_1 \in S_1$ define

$$\hat{\mu}_1(\theta, \hat{s}_2|s_1) = \sum_{a_2} \mu_1(\theta, a_2|s_1) 1[a_2 = \hat{s}_2].$$

For $a_1 \in BFR_1^-$ there exists a justifying BFR belief $\mu_1^{a_1} \in \Delta(\Theta \times A_2)$, so take as a CPS belief[77]

$$\hat{\mu}_1(\theta, \hat{s}_2|s_1) = \sum_{a_2} \mu_1^{a_1}(\theta, a_2) 1[a_2 = \hat{s}_2].$$

---

[76] Again, the equivalent fixed-point definition of BFR stated in Equation 14 is used.

[77] By construction, these a_1 have zero probability under the signal distributions of player 1.

Now, consider the obedient strategies $\beta_1(b(s_1)|s_1) = 1$ if $s_1 \in S_1$, $\beta_1(a_1|a_1) = 1$ if $a_1 \in BFR_1^-$, and $\beta_2(a_2|a_2) = 1$ for every $a_2 \in \hat{S}_2$. It remains to verify that these strategies are optimal given the CPS (and the strategy of the opponent).

Player 1 For every $s_1 \in S_1$ with $\sum_{\tilde{\theta},\hat{s}_2} \pi_1(\tilde{\theta})\Psi_1(s_1, \hat{s}_2|\tilde{\theta}) > 0$ we have

$$\begin{aligned}
b(s_1) &\in \arg\max_{a_1' \in A_1} \sum_{\theta,a_2} \mu_1(\theta, a_2|s_1)u_1(a_1', a_2, \theta) \\
&\in \arg\max_{a_i' \in A_i} \sum_{\theta,a_2} \Psi_1(s_1, a_2|\theta)\pi_1(\theta)u_1(a_1', a_2, \theta) \\
&\in \arg\max_{a_i' \in A_i} \sum_{\theta,a_2,\hat{s}_2} \pi_1(\theta)\Psi_1(s_1, \hat{s}_2|\theta)\beta_2(a_2|\hat{s}_2)u_1(a_1', a_2, \theta),
\end{aligned}$$

where the second line uses the definition of the signal distribution and the belief in the last line is (equivalent to) the updated belief together with belief in the strategy of the other player.

For every $s_i \in S_i$ with $\sum_{\tilde{\theta},\hat{s}_2} \pi_1(\tilde{\theta})\Psi_1(s_1, \hat{s}_2|\tilde{\theta}) = 0$,

$$\begin{aligned}
b(s_1) &\in \arg\max_{a_1' \in A_1} \sum_{\theta,a_2} \mu_1(\theta, a_2|s_1)u_1(a_1', a_2, \theta) \\
&\in \arg\max_{a_i' \in A_i} \sum_{\theta,a_2,\hat{s}_2} \mu_1(\theta, a_2|s_1)1[a_2 = \hat{s}_2]u_1(a_1', \hat{s}_2, \theta) \\
&\in \arg\max_{a_i' \in A_i} \sum_{\theta,\hat{s}_2,a_2} \hat{\mu}_1(\theta, \hat{s}_2|s_1)\beta_2(a_2|\hat{s}_2)u_1(a_1', a_2, \theta).
\end{aligned}$$

Like in the last case, for every $a_1 \in BFR_1^-$ a_1 is a best-reply to $\hat{\mu}_1$ and β_2.

Player 2 For every $a_2 \in \hat{S}_2$ we have

$$\begin{aligned}
a_2 &\in \arg\max_{a_2' \in A_2} \sum_{\theta,a_1} \mu_2^{a_2}(\theta, a_1)u_2(a_1, a_2', \theta) \\
&\in \arg\max_{a_2' \in A_2} \sum_{\theta} \left[\sum_{a_1 \in \{b(s_1)\}_{s_1}} \mu_2^{a_2}(\theta, a_1)u_2(a_1, a_2', \theta) + \sum_{a_1 \in BFR_1^-} \mu_2^{a_2}(\theta, a_1)u_i(a_1, a_2', \theta) \right] \\
&\in \arg\max_{a_2' \in A_2} \sum_{\theta} \left[\sum_{s_1 \in S_1} \frac{\mu_2^{a_2}(\theta, b(s_1))}{|b^{-1}(b(s_1))|} u_2(b(s_1), a_2', \theta) + \sum_{a_1 \in BFR_1^-} \mu_2^{a_2}(\theta, a_1)u_2(a_1, a_2', \theta) \right] \\
&\in \arg\max_{a_2' \in A_2} \sum_{\theta} \sum_{\hat{s}_1 \in \hat{S}_1,a_1} \pi_2(\theta)\psi_2(\hat{s}_1, a_2|\theta)\beta_1(a_1|\hat{s}_1)u_2(a_1, a_2', \theta).
\end{aligned}$$

So that β (together with the constructed CPS) is indeed a BNE. ■

B.2 Detailed calculations for Subsection 4.2

To simplify notation let $\tau_{DD} := \tau(\mu_1^D, \mu_2^D)$ and similar for the other three cases and let $\tau_i := \tau_i(\mu_i^D)$. With this notation,

$$\tau_{DD} = \tau_1(1-\mu_1^D) + \tau_2(1-\mu_2^D) - (1-\pi).$$

Since marginal distribution average out to the prior: $(1-\tau_i)(1-\mu_i^R) + \tau_i(1-\mu_i^D) = 1-\pi$. Hence,

$$\begin{aligned}\tau_{DD} &= \tau_1(1-\mu_1^D) + \tau_2(1-\mu_2^D) - (1-\pi) - \sum_i \frac{(1-\tau_i)(1-\mu_i^R) + \tau_i(1-\mu_i^D)}{2} \\ &= \frac{1}{2}\left[\tau_1(1-\mu_1^D) - (1-\tau_1)(1-\mu_1^R) + \tau_2(1-\mu_2^D) - (1-\tau_2)(1-\mu_2^R)\right].\end{aligned}$$

Given the normalization on the utility of the designer, the objective becomes $-\tau_{DD} + v(\tau_{DR} + \tau_{RD})$. Furthermore, the following equalities hold:

$$\begin{aligned}\tau_{DR} = \tau_1 - \tau_{DD} &= \frac{1}{2}\left[\left(\tau_1\mu_1^D + 1 - \mu_1^R + \tau_1\mu_1^R\right) - \left(\tau_2(1-\mu_2^D) - (1-\tau_2)(1-\mu_2^R)\right)\right] \\ \tau_{RD} = \tau_2 - \tau_{DD} &= \frac{1}{2}\left[\left(\tau_2\mu_2^D + 1 - \mu_2^R + \tau_2\mu_2^R\right) - \left(\tau_1(1-\mu_1^D) - (1-\tau_1)(1-\mu_1^R)\right)\right].\end{aligned}$$

Plugging into the objective (ignoring the $1/2$ scaling):

$$\begin{aligned}& v\left[\left(\tau_1\mu_1^D + 1 - \mu_1^R + \tau_1\mu_1^R\right) - \left(\tau_1(1-\mu_1^D) - (1-\tau_1)(1-\mu_1^R)\right)\right] - \left(\tau_1(1-\mu_1^D) - (1-\tau_1)(1-\mu_1^R)\right) \\ &+ v\left[\left(\tau_2\mu_2^D + 1 - \mu_2^R + \tau_2\mu_2^R\right) - \left(\tau_2(1-\mu_2^D) - (1-\tau_2)(1-\mu_2^R)\right)\right] - \left(\tau_2(1-\mu_2^D) - (1-\tau_2)(1-\mu_2^R)\right) \\ &= \left[2v\tau_1 - \tau_1(1-\mu_1^D) + (1-\tau_1)(1-\mu_1^R)\right] + \left[2v\tau_2 - \tau_2(1-\mu_2^D) + (1-\tau_2)(1-\mu_2^R)\right] \\ &= \tau_1\left(2v - (1-\mu_1^D)\right) + (1-\tau_1)(1-\mu_1^R) + \tau_2\left(2v - (1-\mu_2^D)\right) + (1-\tau_2)(1-\mu_2^R),\end{aligned}$$

so that the objective is separable. From the main text, we know that $\mu_i^D < 2/3$ and $\mu_i^R \geq 2/3$. Thus, we can rewrite the objective as stated in the main text[78] with $f(\mu) := \mathbf{1}[\mu < 2/3](2v + \mu - 1) + \mathbf{1}[\mu \geq 2/3](1-\mu)$. Figure 4 plots this function and its concavification. Since $v \in [-1/2, 0]$ shifts f only vertically, it will not change the maximizer resulting from the concavification.

---

[78] To be precise, the values of f for $\mu \in [1/2, 2/3)$ can be set arbitrary as long as they are strictly below the resulting concavification. This can be done because from the analysis in the main text it is known that μ in this range cannot be optimal.

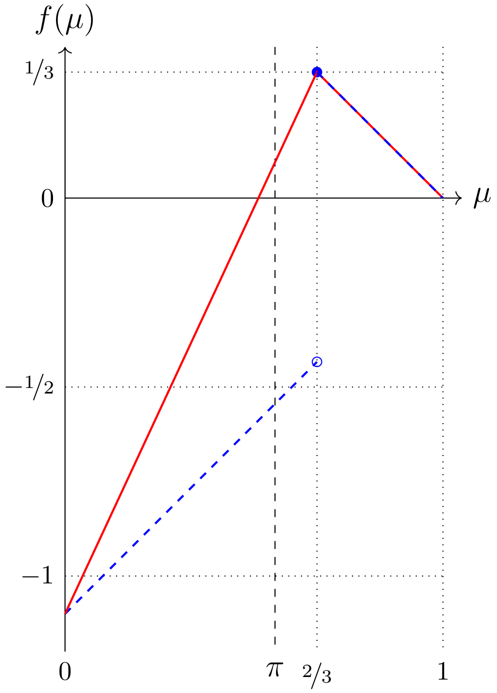

Figure 4: $f(\mu)$ in dashed blue (with $v = -0.05$) and concavification thereof in red.