On Information Design in Games[*]

Laurent Mathevet
New York University

Jacopo Perego
Yale University

Ina Taneva[†]
University of Edinburgh

June 19, 2018

Abstract

Information provision in games influences behavior by affecting players' beliefs about the state, as well as their higher-order beliefs. We characterize the extent to which a designer can manipulate players' beliefs by disclosing information. Building on this, our next results describe the structure of optimal belief distributions, including a concave-envelope representation that subsumes the single-agent result of Kamenica and Gentzkow (2011). Our belief-based approach to information design applies to various equilibrium selection rules and solution concepts. We use it to compute the optimal information structure in an investment game under adversarial equilibrium selection.

Keywords: information design, disclosure, belief manipulation, belief distributions, extremal decomposition, concavification.

JEL Classification: C72, D82, D83.

[*]This work is a version of a paper that circulated under the title "Information Design: The Epistemic Approach," first posted on September 27, 2016.

[†]Special thanks are due to Anne-Katrin Roesler for her insightful comments as a discussant at the 2016 Cowles Foundation conference, and to Philippe Jehiel, Elliot Lipnowski, Stephen Morris, David Pearce, József Sákovics, and Siyang Xiong for their helpful comments and conversations. We are also grateful to Andrew Clausen, Olivier Compte, Jon Eguia, Willemien Kets, Matt Jackson, Qingmin Liu, Efe Ok, Ennio Stacchetti, and Max Stinchcombe, as well as to seminar audiences at Columbia University, Institut d'Anàlisi Econòmica, New York University, Michigan State University, Paris School of Economics, Stony Brook University, University of Cambridge, University of Edinburgh, UT Austin, the SIRE BIC Workshop, the 2016 Decentralization conference, the 2016 Canadian Economic Theory conference, the 2016 Cowles Foundation conference, the 2016 North American Summer Meeting of the Econometric Society, and SAET 2017. Taneva acknowledges financial support from ESRC grant ES/N00776X/1.

E-mail: `lmath@nyu.edu` and `jacopo.perego@yale.edu` and `ina.taneva@ed.ac.uk`

1. Introduction

Monetary incentives, such as taxes, fines, wages, and insurance are ways of manipulating agents' payoffs to incentivize a range of behaviors, from exerting effort to risk-taking. In incomplete information environments, strategic transmission of information may also be used as a tool to affect agents' behavior, in this case by manipulating their beliefs. Information design analyzes the latter, in a setting where a designer commits to disclosing information to a group of interacting agents.

Under incomplete information, agents' behavior depends, in part, on their beliefs about the uncertain state of the world. For example, an investor's belief about the quality of a new technology influences his decision whether to invest or not in the startup that launches it. However, his decision also depends on how likely he thinks other investors are to fund the startup, which in turn depends on their own beliefs about the state and other investors' decisions. Thus, an agent's beliefs about the other agents' beliefs about the state also affect his decision, as do his beliefs about their beliefs about his beliefs about the state, and so on. These higher-order beliefs are absent from the single-agent environment, but they are an important part of the information design problem with multiple interacting agents.

This paper contributes to the foundations of information design in three ways. First, we characterize the feasible distributions of agents' beliefs that a designer can induce through the choice of information structure. Information design is ultimately an exercise in belief manipulation, whether it is explicitly modeled as such or solved—if at all possible—by way of incentive compatible distributions over actions and states. However, an information designer cannot induce just any belief distribution she wishes to. In the single-agent case, the designer is constrained to distributions over beliefs about the state that on average equal to the prior, a condition known as Bayes plausibility (Kamenica and Gentzkow (2011)). In the multi-agent case, there is an additional requirement that agents' beliefs should be consistent with one another. We show that one way to satisfy this consistency condition is to choose a Bayes plausible distribution of the *designer's* first-order beliefs, and then garble it to produce the agents' belief distributions. This perspective is useful, as it allows us to satisfy both consistency and Bayes plausibility by working with first-order beliefs only.

Second, we elicit the structure of belief distributions and characterize the designer's problem in terms of it. We show that any belief distribution that the designer can induce, and hence any information structure, is a combination of basic communication schemes. From this follows a two-step approach to the information design problem: the first step is maximization within, which selects an optimal basic component for every distribution of states, and the second is maximization between, which optimally combines those through public randomization subject to Bayes plausibility. The second step corresponds to the concavification of a value function coming from the first step, and subsumes the single-agent result of Kamenica and Gentzkow (2011). This representation has a dynamic programming feature, which can be helpful for computing a solution. In addition, there is a special feature to our decomposition into basic communication schemes: it is maximal, as these components are obtained by filtering out all public information. Therefore, they do not only correspond to private information, but to *purely* private information, while the public component is given the largest possible weight. In terms of our solution technique, this implies that concavification takes on the largest possible role in the optimization.

Third, the above results apply to a variety of solution concepts and equilibrium selection rules. The choice of solution concept can address many problems in information design, in much the same way that it does in mechanism design. For example, a designer may be concerned that agents only have bounded depths of reasoning,[1] that they can deviate in coalitions, or that they can communicate. A designer may also want an information structure that performs well in all possible scenarios under the chosen solution concept. This robustness requirement can be achieved by maximizing the designer's minimum payoff, that is, by using an adversarial selection rule, which assumes agents will play the designer's least favorite outcome under the solution concept. Current methods have focused on Bayes Nash equilibrium (BNE) as a solution concept and designer-preferred equilibrium selection.

We apply our approach to an investment game where the coordination motive is a source of multiple equilibria under incomplete information, and the designer would like to maximize investment. In this problem, as in other similar ones, the possibility that agents coordinate on the equilibrium least preferred by the designer is a serious issue. In response, the designer should choose the information structure that maximizes investment probabilities in the worst equilibrium. Information design under such adversarial equilibrium selection is outside the scope of existing methods. Under the optimal information structure, computed by applying our approach, every agent either receives a private message that makes investing uniquely optimal due to a combination of first- and higher-order beliefs, or receives a public message that makes it common knowledge that the state is low, hence not investing is uniquely optimal. The private messages create contagion à la Rubinstein (1989), which we refer to as the bandwagon effect: one message induces a first-order belief

[1] In some contexts, it might be reasonable to assume that players can formulate beliefs up to some order but remain agnostic beyond. Earlier versions of this paper included an application with bounded rationalizability as a solution concept.

that is high enough to incentivize investment by itself, while all other messages aim to cause investment by an induction argument that uses beliefs of incrementally higher order.

The single-agent problem has been a rich subject of study since the influential work of Kamenica and Gentzkow (2011) (e.g., Ely, Frankel, and Kamenica (2015), Lipnowski and Mathevet (2018), Kolotilin et al. (2017)). The standard problem is mathematically analogous to Aumann and Maschler (1995), which studies a repeated game with an informed player who exploits his knowledge against an uninformed opponent. Brocas and Carrillo (2007) and Rayo and Segal (2010) proposed models of single-agent information design, with binary state space and sequential binary signals for the former, and optimal advertising about product attributes for the latter. In related work, Benoît and Dubra (2011) study the extent to which a population can hold unrealistically confident beliefs, even if agents process information in a Bayesian way. By contrast, the theory of information design in games is not as well understood. Bergemann and Morris (2016) and Taneva (2014) formulate the Myersonian approach to Bayes Nash information design. This approach is based on a notion of correlated equilibrium under incomplete information, Bayes correlated equilibrium (BCE), which characterizes all possible BNE outcomes that could arise under all information structures. The BCE approach elegantly avoids the explicit modeling of belief hierarchies, and proves useful for solving information design problems by means of linear programming. However, it fundamentally relies on BNE as a solution concept and on selecting the designer-preferred equilibrium in case of multiplicity. In contrast, our results develop the belief-based approach to information design, which can be viewed as the multi-player analogue to Kamenica and Gentzkow (2011)'s single-agent formulation. Finally, optimal information structures have been previously derived in specific strategic environments, as in Vives (1988), Morris and Shin (2002), and Angeletos and Pavan (2007). More recent works study information design in voting games (Alonso and Câmara (2016), Chan et al. (2016)); dynamic bank runs (Ely (2017)); stress testing (Inostroza and Pavan (2017)); auctions (Bergemann, Brooks, and Morris (2017)); contests (Zhang and Zhou (2016)); or focus on public information in games (Laclau and Renou (2016)).

2. The Information Design Problem

A set $N = \{1, \ldots, n\}$ of players interact in an uncertain environment. The variable $\theta \in \Theta$ describes the uncertain state of the world, where the set Θ is finite. Every player $i \in N$ has a finite action set A_i and utility function $u_i : A \times \Theta \to \mathbb{R}$, where $A = \prod_i A_i$. A priori, the players only know that θ is distributed according to

$\mu_0 \in \Delta\Theta$, which is common knowledge. We refer to $G = (\Theta, \mu_0, N, \{A_i\}, \{u_i\})$ as the **base game**.

A designer commits to disclosing information to the players about the payoff relevant state θ. This is modeled by an **information structure** (S, π), where S_i is the finite set of messages that player i can receive, $S = \prod_i S_i$ is the set of message profiles, and $\pi : \Theta \to \Delta S$ is the information map. In any state θ, the message profile $s = (s_i)$ is drawn according to $\pi(s|\theta)$ and player i privately observes s_i. An information structure can be thought of as an experiment concerning the state, such as an audit, a stress test or a medical analysis. As is standard in information design, this model assumes that the designer commits to the information structure at a time when she does not know the realized state, but only knows its distribution μ_0.[2] The information designer's preferences are summarized by the payoff function $v : A \times \Theta \to \mathbb{R}$. Her objective is to maximize her expected payoff through the choice of information structure.

The combination of information structure and base game, $\mathcal{G} = \langle G, (S, \pi)\rangle$, constitutes a **Bayesian game** in which agents play according to a **solution concept** $\Sigma(\mathcal{G}) \subseteq \{\sigma = (\sigma_i) \,|\, \sigma_i : S_i \to \Delta A_i \text{ for all } i\}$. The resulting outcomes are distributions over action profiles and states, represented by

$$O_\Sigma(\mathcal{G}) = \{\gamma \in \Delta(A \times \Theta) : \text{there exists } \sigma \in \Sigma(\mathcal{G}) \text{ such that}$$
$$\gamma(a, \theta) = \textstyle\sum_s \sigma(a|s)\pi(s|\theta)\mu_0(\theta) \text{ for all } (a, \theta)\}.$$

Assume that O_Σ is non-empty and compact-valued. Given that G is finite, this holds when Σ is BNE, for example. For a fixed base game, we just write $O_\Sigma(S, \pi)$. When O_Σ contains multiple outcomes, the designer expects that *one* of them will happen, which is described by a **selection rule** $g : D \subseteq \Delta(\Theta \times A) \mapsto g(D) \in D$. The worst and the best outcomes are natural selection criteria. A pessimistic designer, or one interested in robust information design, expects the worst outcome:

$$g(D) \in \operatorname*{argmin}_{\gamma \in D} \sum_{a,\theta} \gamma(a, \theta) v(a, \theta) \tag{1}$$

for all compact $D \subseteq \Delta(\Theta \times A)$. An optimistic designer would instead expect the best equilibrium, with argmax instead of argmin in (1). Other criteria, such as random choice rules, could also be considered. Letting $g^{(S,\pi)} := g(O_\Sigma(S, \pi))$, the designer's ex-ante expected payoff is given by

$$V(S, \pi) := \sum_{a,\theta} g^{(S,\pi)}(a, \theta)\, v(a, \theta). \tag{2}$$

[2] While commitment can be a strong assumption in some situations, it holds implicitly in repeated environments wherein a sender makes announcements periodically and wants to be trusted in the future (Best and Quigley (2018) and Mathevet, Pearce, and Stacchetti (2018)). As such, the commitment assumption provides a useful benchmark in various contexts.

Finally, the information design problem is $\sup_{(S,\pi)} V(S,\pi)$.

3. Information Design as Belief Manipulation

We reformulate information design into a belief manipulation problem, analogously to Kamenica and Gentzkow (2011)'s approach to the single-agent case. Choosing an information structure is equivalent to choosing a distribution over belief hierarchies. However, only a special class of belief (hierarchy) distributions can be induced by information structures. Kamenica and Gentzkow (2011) established that choosing an information structure is equivalent to choosing a *Bayes plausible* distribution over first-order beliefs. We show that a similar equivalence holds in games, provided that players' beliefs are, in addition, consistent with each other.

3.1. Belief Distributions

A **belief hierarchy** t_i for player i is an infinite sequence $(t_i^1, t_i^2, \dots)$ whose components are coherent[3] beliefs of all orders: $t_i^1 \in \Delta\Theta$ is i's first-order belief; $t_i^2 \in \Delta(\Theta \times (\Delta\Theta)^{n-1})$ is i's second-order belief (i.e., a belief about θ and every j's first-order beliefs); and so on. Whereas a player's belief hierarchies are coherent, they may assign positive probability to other players' belief hierarchies that are not coherent. Brandenburger and Dekel (1993) show that we can construct a set of coherent belief hierarchies T_i for every i such that there exists a homeomorphism $\beta_i^* : T_i \to \Delta(\Theta \times T_{-i})$ for all i.[4] This map describes i's beliefs about (θ, t_{-i}) given t_i, and shows that there are sets of coherent belief hierarchies across players that only put positive probabilities on each other, making coherency common knowledge. Let $T := \prod_i T_i$ be the space of hierarchy profiles.

Given an information structure (S,π) and μ_0, player i receiving message s_i uses Bayes' rule to formulate beliefs $\mu_i(s_i) \in \Delta(\Theta \times S_{-i})$ about the state and other players' messages in $S_{-i} := \prod_{j \neq i} S_j$. In particular, $\mu_i^1(s_i) := \text{marg}_\Theta \mu_i(s_i)$ describes i's belief about the state given s_i, called *first*-order belief. Since every j has a first-order belief $\mu_j^1(s_j)$ for every message s_j, i's belief about s_j (given s_i) gives i a belief about $\mu_j^1(s_j)$. This belief about j's belief about the state is i's second-order belief $\mu_i^2(s_i)$ given s_i.[5] Since every j has a second-order belief $\mu_j^2(s_j)$ for every s_j, i can formulate a third-order belief given s_i and so on. By induction, every s_i induces a

[3] A hierarchy t is coherent if any belief t_i^k coincides with all beliefs of lower order, $\{t_i^n\}_{n=1}^{k-1}$, on lower order events: $\text{marg}_{X_{k-1}} t_i^k = t_i^{k-1}$ for all $k \geq 1$ where $X_{k-1} = \text{supp}\, t_i^{k-1}$.

[4] We often write $\beta_i^*(t_{-i}|t_i)$ and $\beta_i^*(\theta|t_i)$ to refer to the corresponding marginals.

[5] Technically speaking, a second-order belief also includes a first-order belief.

$\pi(\cdot\|0)$	s_1	s_2
s_1	1	0
s_2	0	0

$\pi(\cdot\|1)$	s_1	s_2
s_1	$\frac{1}{2}$	0
s_2	0	$\frac{1}{2}$

Table 1: A (Public) Information Structure

belief hierarchy $h_i(s_i) \in T_i$ for player i, and every message profile s induces a profile of belief hierarchies $h(s) := (h_i(s_i))_{i\in N}$.

Definition 1. An information structure (S, π) induces a distribution $\tau \in \Delta T$ over profiles of belief hierarchies, called a **belief(-hierarchy) distribution**, if

$$\tau(t) = \sum_\theta \pi\big(\{s : h(s) = t\}|\theta\big)\mu_0(\theta) \tag{3}$$

for all t.

For example, the information structure in Table 1 induces $\tau = \frac{3}{4}t_{1/3} + \frac{1}{4}t_1$ when $\mu_0 := \mu_0(\theta = 1) = \frac{1}{2}$, where t_μ is the hierarchy profile in which $\mu := \mu(\theta = 1)$ is commonly believed.[6]

We categorize belief distributions into public and private. This distinction is closely linked to the nature of information that induces those distributions.

Definition 2. A belief distribution τ is **public** if $t_i^1 = t_j^1$ and $\mathrm{marg}_{T_{-i}}\beta_i^*(\cdot|t_i) = \delta_{t_{-i}}$ (where δ is the Dirac measure) for all $t \in \operatorname{supp}\tau$ and $|\operatorname{supp}\tau| \geq 2$. A belief distribution τ is **private** if it is not public.

The first part says that players share the same first-order beliefs and this is commonly believed among them. This is the natural translation in terms of beliefs of the standard notion of public information. Notice also that we categorize the degenerate case $|\operatorname{supp}\tau| = 1$ as private. When the support is a singleton this distinction is indeed mostly a matter of semantics; yet the current choice makes our characterization later on more transparent.

3.2. Manipulation

Consider an information design problem with $\theta \in \{0, 1\}$,

$$\begin{aligned} u_i(a_i, \theta) &= -(a_i - \theta)^2 \quad i = 1, 2 \\ v &= u_1 - u_2, \end{aligned}$$

[6] To see why, note that $\Pr(s_1, s_1) = \frac{3}{4}$, $\Pr(s_2, s_2) = \frac{1}{4}$, and a player i receiving message s_ℓ has beliefs $(2\ell - 1)/3$ that $\theta = 1$ and is certain that j also received s_ℓ.

where $a_i \in \{0,1\}$ for $i = 1,2$ and each player cares only about matching the state with his action. The designer wants to favor the first player while harming the second. She could obtain her maximal payoff of 1, if she could somehow reveal the state perfectly to player 1, while persuading player 2 that the opposite state has realized. If this were possible, player 1 would be certain that the state is θ, player 2 would be certain that the state is $1-\theta$, and this disagreement would be commonly known. Since Aumann (1976), we have known that Bayesian agents cannot agree to disagree if they have a common prior. Say that $p \in \Delta(\Theta \times T)$ is a **common prior** if

$$p(\theta, t) = \beta_i^*(\theta, t_{-i}|t_i)p(t_i) \tag{4}$$

for all θ, t and i. That is, all players i obtain their belief map β_i^* by Bayesian updating of the same distribution p. Denote by Δ^f the probability measures with finite support. Define

$$\mathcal{C} := \Big\{\tau \in \Delta^f T : \exists \text{a common prior}\, p \text{ s.t. } \tau = \text{marg}_T p\Big\} \tag{5}$$

to be the space of **consistent** (belief-hierarchy) distributions. In a consistent distribution, all players' beliefs arise from a common prior that draws every t with the same probability as τ, i.e., $\tau = \text{marg}_T p$. Let p_τ be the unique distribution p in (5) (uniqueness follows from Mertens and Zamir (1985, Proposition 4.5)).

A distribution $\tau \in \Delta^f T$ is **Bayes plausible** if the expected first-order belief of at least one player equals the prior:

$$\sum_{t_i} \text{marg}_\Theta \beta_i^*(\cdot|t_i)\tau_i(t_i) = \mu_0 \quad \text{for some } i.$$

Proposition 1. *There exists (S,π) that induces $\tau \in \Delta^f T$, if and only if, τ is consistent and Bayes plausible.*

This characterization, which builds upon Mertens and Zamir (1985), disciplines the designer's freedom in shaping players' beliefs, but only to the extent that they are consistent and Bayes plausible. In the one-agent case, information disclosure is equivalent to choosing a Bayes plausible distribution over first-order beliefs. In games, it is equivalent to choosing a Bayes plausible and *consistent* distribution over hierarchies of beliefs. Importantly, it does not matter which player i satisfies Bayes plausibility, because by consistency, if it is true for one player, then it will hold for all.

Returning to the simple example above, if the designer wants to obtain a payoff of 1 with certainty, she must give full information to player 1, for otherwise she will not get an expected payoff of 0 from him. At the same time, she must fool player 2 all the time. Thus, to reach the upper bound of 1, it would have to be that $\beta_1^*(\theta = 1, t_2|t_1) = 1$ and $\beta_2^*(\theta = 0, t_1|t_2) = 1$ for some (t_1, t_2), which violates (4). So, no information structure can deliver an expected payoff of 1. While the designer wants the agents to "maximally disagree", she is constrained by Proposition 1, as their beliefs must satisfy consistency. From here, it is a small step to conclude that it is optimal for 1 to know the value of the state and for 2 to know as little as possible. In the language of information disclosure, it is optimal to give full information to 1 and no information to 2.

Consistency combines individual beliefs into a feasible whole, as formalized by (4). This condition, imposed in addition to Bayes plausibility, is the main difference and technical challenge compared to the single-agent case. One approach to implementing it is to design the individual distributions of players' beliefs and then couple them in a consistent way (see Ely (2017) for a related procedure). A different approach, formulated in the next proposition, interprets the designer's information design problem as a choice of a distribution over her own beliefs about the state, and then views players' first-order belief distributions as a special garbling of the designer's information.

Proposition 2. *If τ is consistent and Bayes plausible, then there exists $\nu : supp\,\tau \to \Delta\Theta$ such that:*

$$\sum_t \tau(t)\nu(t) = \mu_0 \tag{6}$$

and

$$\sum_{t_{-i}} \tau(t_{-i}|t_i)\nu(t_i, t_{-i}) = marg_\Theta \beta_i^*(\cdot|t_i) \quad \forall i, t_i \tag{7}$$

Conversely, if $\xi \in \Delta^f(\Delta\Theta)^n$ and $\nu : supp\,\xi \to \Delta\Theta$ satisfy

$$\sum_{\boldsymbol{\mu}=(\mu_i)_i} \xi(\boldsymbol{\mu})\nu(\boldsymbol{\mu}) = \mu_0, \tag{8}$$

and

$$\sum_{\boldsymbol{\mu_{-i}}=(\mu_j)_{j\neq i}} \xi(\boldsymbol{\mu_{-i}}|\mu_i)\nu(\boldsymbol{\mu_{-i}}, \mu_i) = \mu_i \quad \forall i, \mu_i, \tag{9}$$

then there exist a consistent and Bayes plausible τ such that $supp\,\tau_i \cong supp\,\xi_i$ (i.e., there is a bijection $\phi_i : supp\,\xi_i \to supp\,\tau_i$) and $\mu_i = marg_\Theta \beta_i^*(\cdot|\phi_i(\mu_i))$ for all i, and $\tau(t) = \xi(\phi^{-1}(t))$ for all t where $\phi = (\phi_i)$.

The designer's beliefs about the state are obtained by conditioning on the entire hierarchy profile $t = (t_1, \ldots, t_n)$, whereas each player i only conditions upon t_i. In the first part of the proposition, $\nu(t)$ represents the designer's beliefs about the state given t, and τ can be interpreted as the distribution of the designer's beliefs. If τ is consistent and Bayes plausible, then on average the designer's beliefs must be equal to the prior, which is equation (6). Additionally, each player's first-order belief, $\text{marg}_\Theta \beta_i^*(\cdot|t_i)$, *must* be derived from the designer's beliefs through (7). The lhs of condition (7) describes a mean-preserving transformation of the designer's beliefs. Thus, every player's first-order beliefs in expectation will also equal the prior, due to (6). In fact, (7) describes every first-order belief of every player as an average of the designer's beliefs, hence each player's distribution of first-order beliefs is a mean-preserving contraction (mpc) of the designer's distribution. Put differently, the designer's belief distribution is a mean-preserving spread of players' first-order belief distributions. This means that the designer is better informed about the state than each player.[7] While this can be achieved in many ways, since there are many possible mpc's, only the mpc made of the conditional distributions derived from τ is compatible with consistency.

The second part of the proposition provides a method for constructing *all* consistent and Bayes plausible belief-hierarchy distributions, subject to hierarchies being identified with their first-order beliefs.[8] Let ξ be a distribution over players' first-order beliefs $\boldsymbol{\mu} = (\mu_i)$, and let $\nu(\boldsymbol{\mu}) \in \Delta\Theta$ be the designer's belief about the state given $\boldsymbol{\mu}$. Indirectly through $\boldsymbol{\mu}$, ξ can therefore be thought of as a distribution over the designer's beliefs ν, required to average μ_0 by (8). Observe that ξ and ν fully determine a prior $p(\theta, \boldsymbol{\mu}) = \xi(\boldsymbol{\mu})\nu(\theta|\boldsymbol{\mu})$ for all $(\theta, \boldsymbol{\mu})$. Given p, every player i formulates beliefs about the other players' first-order beliefs conditional on his own, $p(\boldsymbol{\mu}_{-i}|\mu_i)$. Players' entire hierarchies are built inductively in this way. In that construction, players' updated beliefs coincide with $\boldsymbol{\mu}$ when (9) holds (i.e., $p(\theta|\mu_i) = \mu_i$), and τ is the belief hierarchy distribution induced by p.

3.3. Outcomes from Belief Distributions

To complete the formulation of information design in the space of beliefs, the equivalence between information structures and belief distributions should be more than epistemic, it should be about outcomes. Given a consistent distribution τ, denote a solution concept in the space of beliefs as $\Sigma^{\mathrm{B}}(\tau) \subseteq \{\sigma = (\sigma_i)|\sigma_i : \operatorname{supp}\tau_i \to \Delta A_i \text{ for all } i\}$. This set describes players' behavior in the Bayesian game $\langle G, p_\tau\rangle$,

[7] Mean-preserving spreads (contractions) are closely connected to the usual garbling definition of Blackwell informativeness (Theorem 12.2.2. in Blackwell and Girshick (1954)).

[8] In this version of the proposition, no two hierarchies have the same first-order beliefs. For some applications, we may want to induce hierarchies with the same first-order beliefs but different higher-order beliefs. This can be done provided ν is measurable with respect to the players' hierarchies.

where p_τ is the unique common prior p such that $\text{marg}_T p = \tau$. The outcomes that ensue are

$$O_{\Sigma^{\text{B}}}(\tau) := \{\gamma \in \Delta(A \times \Theta) : \text{there exists } \sigma \in \Sigma(\tau) \text{ such that} \\ \gamma(a,\theta) = \textstyle\sum_t \sigma(a|t) p_\tau(t,\theta) \text{ for all } (a,\theta)\}.$$

It is well-known that a given solution concept may not yield the same set of outcomes in the space of information as in the space of beliefs (e.g., Ely and Peski (2006) and Liu (2009)). That is, the fact that (S,π) induces τ does *not* guarantee that $O_\Sigma(S,\pi) = O_{\Sigma^{\text{B}}}(\tau)$ when $\Sigma = \Sigma^{\text{B}}$. For example, when $\Sigma = \Sigma^{\text{B}} = \text{BNE}$, it can happen that $O_\Sigma(S,\pi) \supsetneq O_{\Sigma^{\text{B}}}(\tau)$, as there may be multiple message profiles s inducing the same hierarchy profile t, which creates opportunities for correlation of players' behavior in (S,π) that are not possible in τ.

Since Σ^{B} is relevant only insofar as it captures the outcomes from Σ, let Σ^{B} be such that

$$O_{\Sigma^{\text{B}}}(\tau) = \bigcup_{(S,\pi) \text{ induces } \tau} O_\Sigma(S,\pi) \tag{10}$$

for all consistent τ. For various solution concepts Σ, it is clear from the literature which Σ^{B} satisfies (10). For example, for $\Sigma = \text{BNE}$, Σ^{B} is Liu (2015)'s (belief-preserving) correlated equilibrium. Alternatively, for the solution concept of interim correlated rationalizability (Dekel, Fudenberg, and Morris (2007)), $\Sigma = \Sigma^{\text{B}} = \text{ICR}$.

4. Representation of Optimal Solutions

In this section, we prove that optimal solutions to information design problems in games can be seen as a combination of special distributions. As a consequence, all optimal solutions can be decomposed into optimal purely private and optimal public components, where the latter come from concavification.

4.1. Assumptions

Our approach can handle various selection rules and solution concepts, provided the following assumptions hold:

(Linear Selection). g is linear.[9]

(Invariant Solution). For all consistent τ and τ', if $\sigma \in \Sigma^{\text{B}}(\tau)$, then there exists $\sigma' \in \Sigma^{\text{B}}(\tau')$ such that $\sigma(t) = \sigma'(t)$ for all $t \in \text{supp}\,\tau \cap \text{supp}\,\tau'$.

[9] For all D', D'' and $0 \le \alpha \le 1$, $g(\alpha D' + (1-\alpha)D'') = \alpha g(D') + (1-\alpha) g(D'')$.

Linearity of g is a natural assumption that requires the selection criterion to be independent of the subsets of outcomes to which it is applied. The best and the worst outcomes, defined in (1), are linear selection criteria. However, selecting the best outcome within one subset and the worst in another breaks linearity, unless the outcome is always unique.

Invariance says that play at a profile of belief hierarchies t under Σ^{B} is independent of the ambient distribution from which t is drawn. For instance, Liu (2015)'s correlated equilibrium satisfies invariance. And when $\Sigma = \mathrm{ICR}$ (that is, Interim Correlated Rationalizability by Dekel, Fudenberg, and Morris (2007)), it is well-known that $\Sigma^{\mathrm{B}} = \Sigma$, which also satisfies invariance. Invariance is important because:

Proposition 3. *If Σ^B is invariant, then O_{Σ^B} is linear.*

4.2. Representations

Information design exhibits a convex structure when seen as belief manipulation. From any consistent τ' and τ'', the designer can build a third distribution, $\tau = \alpha\tau' + (1-\alpha)\tau''$, which can be induced by an information structure provided that it is Bayes plausible. In particular, this is true even if τ' and τ'' are themselves *not* Bayes plausible. In technical terms, $\mathcal{C}$ is convex and, moreover, admits extreme points.[10] In the tradition of extremal representation theorems,[11] the designer generates a consistent and Bayes plausible distribution by randomizing over extreme points, and *any* consistent and Bayes plausible distribution can be generated in this way. By Proposition 1, any information structure can thus be interpreted as a convex combination of extreme points. Importantly, these extreme points have a useful characterization: they are the minimal consistent distributions (see Lemma 2 in Appendix A.4). A consistent distribution $\tau \in \mathcal{C}$ is **minimal** if there is no $\tau' \in \mathcal{C}$ such that $\operatorname{supp}\tau' \subsetneq \operatorname{supp}\tau$. Let $\mathcal{C}^{\mathrm{M}}$ denote the set of all minimal distributions,[12] which is nonempty by basic inclusion arguments. From this definition follows a nice interpretation in terms of information. The minimal distributions correspond to *purely private* information. By definition, any non-minimal distribution τ contains two consistent components with support $\operatorname{supp}\tau'$ and $\operatorname{supp}\tau \backslash \operatorname{supp}\tau'$. A public signal makes it common knowledge among the players which of these two components their beliefs are in. Since a minimal belief distribution has only one component, it contains

[10] An extreme point of $\mathcal{C}$ is an element $\tau \in \mathcal{C}$ with the property that if $\tau = \alpha\tau' + (1-\alpha)\tau''$, given $\tau', \tau'' \in \mathcal{C}$ and $\alpha \in [0,1]$, then $\tau' = \tau$ or $\tau'' = \tau$.

[11] E.g., Minkowski–Caratheodory theorem, Krein-Milman theorem, and Choquet's integral representation theorem.

[12] Minimal belief subspaces appeared in contexts other than information design in Heifetz and Neeman (2006), Barelli (2009), and Yildiz (2015).

no such public signal. As such, it is purely private information (possibly degenerate).

Owing to their mathematical status as extreme points, the minimal consistent distributions correspond to the basic communication schemes at a given distribution of states, from which *all* others can be constructed. In the single-agent case, the minimal distributions are the agent's first-order beliefs. The results below formalize their special role in information design.[13]

Given any consistent distribution τ and the selected outcome $g^\tau := g(O_{\Sigma^{\mathrm{B}}}(\tau))$, the designer's ex ante expected payoff is given by

$$w(\tau) := \sum_{\theta,a} g^\tau(a,\theta)\, v(a,\theta). \tag{11}$$

Theorem 1 (Representation Theorem). *The designer's maximization problem can be represented as*

$$\begin{aligned} \sup_{(S,\pi)} V(S,\pi) = \quad & \sup_{\lambda\in\Delta^f(\mathcal{C}^M)} \quad \sum_{e\in\mathcal{C}^M} w(e)\lambda(e) \\ & \text{subject to} \quad \sum_{e\in\mathcal{C}^M} marg_\Theta p_e\, \lambda(e) = \mu_0. \end{aligned} \tag{12}$$

Corollary 1 (Within–Between Maximizations). *For any $\mu\in\Delta\Theta$, let*

$$w^*(\mu) := \sup_{e\in\mathcal{C}^M : marg_\Theta p_e = \mu} w(e). \tag{13}$$

Then, the designer's maximization problem can be represented as

$$\begin{aligned} \sup_{(S,\pi)} V(S,\pi) = \quad & \sup_{\lambda\in\Delta^f\Delta\Theta} \quad \sum_{\operatorname{supp}\lambda} w^*(\mu)\lambda(\mu) \\ & \text{subject to} \quad \sum_{\operatorname{supp}\lambda} \mu\lambda(\mu) = \mu_0. \end{aligned} \tag{14}$$

From the representation theorem, the designer maximizes her expected payoff *as if* she were optimally randomizing over minimal consistent distributions, subject to posterior beliefs averaging to μ_0 across those distributions. Every minimal distribution e induces a Bayesian game and leads to an outcome for which the designer receives expected payoff $w(e)$. Every minimal distribution has a distribution over states, $\mathrm{marg}_\Theta p_e = \mu$, and the "further" that is from μ_0 , the "costlier" it is for the

[13]We further illustrate the notion of minimal distribution in Appendix B by characterizing minimal distributions for public and conditionally independent information.

designer to use it. In this sense, the constraint in (12) can be seen as a form of budget constraint.

The corollary decomposes the representation theorem into two steps. First, there is a **maximization within**—given by (13)—that takes place among all the minimal distributions with $\text{marg}_\Theta p_e = \mu$ and for all μ. All minimal distributions with the same μ contribute equally toward the Bayes plausibility constraint; hence, the designer should choose the best one among them, i.e., the one that gives the highest value of $w(e)$. Interestingly, maximization within delivers the optimal value of private information, which takes the form of a value function $\mu \mapsto w^*(\mu)$. The possibility to identify the optimal value of private information comes from the fact that all minimal distributions represent purely private information.

Second, there is a **maximization between** that concavifies the value function, thereby optimally randomizing between the minimal distributions that maximize within. This step is akin to a public signal λ that "sends" all players to different minimal distributions e, thus making e common knowledge. From standard arguments (Rockafellar (1970, p.36)), the rhs of (14) is a characterization of the **concave envelope** of w^*, defined as $(\text{cav}\, w^*)(\mu) = \inf\{g(\mu) : \ g \text{ concave and } g \geq w^*\}$. Hence, the corollary delivers a concave-envelope characterization of optimal design. In the one-agent case, $\{e \in \mathcal{C}^{\text{M}} \,\text{s.t.}\, \text{marg}_\Theta p_e = \mu\} = \{\mu\}$, hence $w^* = w$ in (13) and the theorem reduces to maximization between.

While any belief distribution, and hence any information structure, can be decomposed into private and public information, this might potentially be done in different ways. However, the decomposition from the corollary, which applies to all information structures, has a special feature: its private components cannot be decomposed further without violating consistency. That is to say, such decomposition is maximal, as it filters out all public information from the private components and puts it into the public component. Therefore, the minimal distributions do not only correspond to private information, but to *purely* private information. This gives the public component its largest possible weight in the decomposition. It also implies that concavification takes on the largest possible role in the optimization.

5. Application: Adversarial Selection

Monetary incentives have been used to stimulate investment and technology adoption (e.g., tax incentives by governments), to stabilize banks and currencies (e.g., financial interventions by central banks), and to increase efforts in organizations (e.g., compensation schemes by companies). In such situations, often characterized by coordination motives, informational incentives can also be used to leverage the underlying complementarities .

We consider the problem of fostering investment in an interaction where two players are choosing whether or not to invest, $\{I, N\}$, given an uncertain state $\theta \in \{-1, 2\}$. The payoffs of the interaction are summarized in Table 2. Let $\mu_0 := \text{Prob}(\theta = 2) > 0$ denote the probability of the high state.

(u_1, u_2)	I	N
I	θ, θ	$\theta - 1, 0$
N	$0, \theta - 1$	$0, 0$

Table 2: Investment Game

Under complete information, each player has a dominant strategy to invest when $\theta = 2$ and not to invest when $\theta = -1$. Under incomplete information, however, a coordination problem arises. A player with a "moderate" belief that $\theta = 2$ will invest if and only if he believes that the other player is likely enough to invest as well. This gives rise to multiple equilibria.

Note that this game differs in an important way from Carlsson and van Damme (1993) and Morris and Shin (2003), because the coordination problem arises only under incomplete information. Relatedly, it also differs from Rubinstein (1989), Kajii and Morris (1997), and Hoshino (2018), because no equilibrium of the complete information game is robust to the introduction of incomplete information. For example, (N, N) is a dominant equilibrium in the low state $\theta = -1$, which makes (I, I) very fragile to the introduction of incomplete information.

Consider an information design problem with the following features:

(a) The designer wants to stimulate investment, which is modeled by a symmetric and monotone payoff function: $v(I, I) > v(I, N) = v(N, I) > v(N, N)$.

(b) The solution concept is Bayes Nash Equilibrium (BNE).

(c) The *min* selection rule, defined in (1), chooses the worst BNE outcome.

The adversarial equilibrium selection, defined by the *min* in (c), corresponds to a form of robust information design: the designer aims to maximize her minimum equilibrium payoff. As it is difficult to know which equilibrium the agents will coordinate on, it is important that the optimal information structure fares well in all equilibrium situations. This matters in various settings, and especially in public policy applications, where erroneous coordination can be detrimental to social welfare.

The current method available for Bayes-Nash information design, based on BCE (see Bergemann and Morris (2016) and Taneva (2014)), outputs a *sub*optimal information structure in this problem. If we apply the BCE program

$$\begin{aligned} &\max_{\pi} \sum_{\theta,a} v(a,\theta)\pi(a|\theta)\mu_0(\theta) \\ &\text{s.t. } \pi \text{ is a BCE} \end{aligned} \tag{15}$$

to any designer's payoff satisfying (a) and for any $\mu_0 \in [\frac{1}{2}, \frac{2}{3})$, it outputs 'no information' as the solution (i.e., $\pi^*(I, I|\theta) = 1$ for all θ). The reason for this is that the *max* selection rule is implicitly built into (15) (see Bergemann and Morris (2018) for further discussion). Since (I, I) and (N, N) are both BNE in the absence of any further information than the prior μ_0, the max rule selects (I, I). Instead, under the min rule, (N, N) will be selected, which results in the smallest possible payoff for the designer. We use our approach to compute an optimal information structure, and show that the designer can achieve a higher expected payoff by revealing some information for this range of priors.

In design environments where equilibrium multiplicity is not simply resolved in favor of the designer, it is well-known that relevant mechanisms can have infinite message spaces.[14] In this paper, however, we have restricted attention to *finite* message spaces, as this guarantees existence of BNE for all belief-hierarchy distributions. Given our objective to develop a general theory, it is important that it applies to a large class of games under the standard solution concept. Finite messages are also the topic of a literature in mechanism design and implementation theory. Hurwicz (1986) and Hurwicz and Marschak (2003) discuss the physical impossibility of transmitting infinite messages, hence the need of 'rounding off' messages and finiteness. In our investment game, we restrict attention to at most $m = 4$ messages in the maximization within. This will be without loss for some priors and constraining for others, which is true for any finite m. Our approach can be applied analogously to find the optimal information structure for any m. Having methods for computing optimal finite information structures is useful, because it allows us to evaluate their performance, switch from m to $m + 1$ messages only when necessary, and avoid unbounded induction.

[14]See Maskin (1999)'s mechanism and the literature on implementation theory. See Hoshino (2018) and the survey by Bergemann and Morris (2018), in particular the application of Kajii and Morris (1997) to information design problems.

5.1. Worst-Equilibrium Characterization

When a player believes that $\theta = 2$ with probability larger than $2/3$, investing is uniquely optimal[15] for him, irrespective of his belief about the other player's action. Investing can also be uniquely optimal even when a player's belief that $\theta = 2$ is less than $2/3$, if that player believes the other player will invest with large enough probability.

Using the concepts from Section 3, let ρ_i^k be the set of hierarchies defined inductively as follows:

$$\rho_i^1 = \left\{ t_i : \beta_i^*(\{\theta = 2\} \times T_j | t_i) > \frac{2}{3} \right\}$$

$$\rho_i^k = \left\{ t_i : \beta_i^*(\{\theta = 2\} \times T_j | t_i) + \frac{1}{3}\beta_i^*(\Theta \times \rho_j^{k-1} | t_i) > \frac{2}{3} \right\}.$$

If player i's hierarchy t_i is in ρ_i^1, then he believes with probability greater than 2/3 that $\theta = 2$, and his unique optimal response is to play I. If $t_i \in \rho_i^2$, then player i assigns high enough probability either to $\theta = 2$ or to player j playing I (due to $t_j \in \rho_j^1$), so that I is again uniquely optimal. By induction, the same conclusion holds for hierarchies in ρ_i^k for any $k \geq 1$. Letting $\rho_i := \cup_{k \geq 1} \rho_i^k$, the unique optimal action for a player with belief in ρ_i is I. This implies that, in *all* BNEs, player i's equilibrium strategy must choose I with certainty when his hierarchy is in ρ_i.

Given a belief distribution $\tau \in \Delta(T_1 \times T_2)$, the worst equilibrium for our designer is such that all players play I *only* when their beliefs belong to ρ_i and play N otherwise: for all i and t_i,

$$\sigma_i^{\text{MIN}}(I|t_i) = \begin{cases} 1 & \text{if } t_i \in \rho_i \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

The strategy profile $(\sigma_i^{\text{MIN}})_i$ is the worst equilibrium, because it minimizes the probability that either player plays I, since I is played only when it is uniquely rationalizable and N is played otherwise — both when N is uniquely rationalizable and when both actions are rationalizable.

Given the structure of the worst equilibrium, we show that it is never optimal to induce a belief hierarchy at which a player has multiple rationalizable actions. This implies that equilibrium multiplicity is never beneficial to our designer under adversarial selection. This conclusion is not immediate, as it could well be that the worst equilibrium (out of many) under some belief distribution might dominate the unique equilibrium under another distribution for the same prior.

[15] Formally, uniquely rationalizable.

Proposition 4. *If τ^* is an optimal belief-hierarchy distribution and player i plays N with positive probability at $t_i \in supp\, \tau_i^*$ in any BNE, then $\beta_i^*(\theta = 2|t_i) = 0$.*

An immediate corollary is that any optimal belief distribution will generate a unique BNE under the min selection criterion, since all players have a unique rationalizable action at each hierarchy.

Corollary 2. *All optimal belief-hierarchy distributions generate a unique BNE.*

Equilibrium multiplicity would imply that there exist hierarchies for which both I and N are rationalizable. These hierarchies would have moderate beliefs: they are neither optimistic enough about the state to play I even when the other player does not invest, nor are they pessimistic enough to play N given that the other player invests. Instead of inducing moderate beliefs that do not contribute to investment under adversarial selection, the designer should instead create very pessimistic hierarchies for which N is uniquely rationalizable, allowing her to put more weight on optimistic hierarchies, for which I is uniquely rationalizable. That is, the designer optimally fosters investment in the worst equilibrium by redistributing the mass from hierarchies with multiple rationalizable actions to hierarchies with a uniquely rationalizable action while preserving Bayes plausibility.

The next proposition says that if the designer considers sending at most m private messages in total, then she need only design hierarchies of order m or lower in order to induce investment.

Proposition 5. *Suppose τ^* is an optimal belief-hierarchy distribution. Let $m = \sum_{i=1,2} |supp\, \tau_i^*|$. Then, for all i and $t_i \in \tau_i^*$, $\sigma_i^{MIN}(I|t_i) = 1$ if and only if $t_i \in \cup_{k=1}^m \rho_i^k$.*

The equilibrium characterization in (16) and the above results simplify the maximization within by pointing to a tractable set of belief hierarchies: namely the investing hierarchies, which must be in $\cup_{k=1}^m \rho_i^k$, and the non-investing hierarchies, at which a player is certain that $\theta = -1$.

5.2. Solution

Following Theorem 1 and Corollary 1 we solve the information design problem in two steps: maximization within and maximization between.

5.2.1. Maximization Within

Suppose the designer can or wants to send at most two messages per player, that is, $m = 4$. Then, the largest minimal distributions to consider have at most two different hierarchies per player, denoted t_i' and t_i'' for $i = 1, 2$. By Propositions 4 and 5, it is without loss to look only at hierarchies in the sets ρ_i^k with $k = 1, \dots, 4$ for all i. Given a commonly known $\mu := \text{Prob}(\theta = 2)$, denote an optimal minimal distribution e_μ^* by

e_μ^*	t_2'	t_2''
t_1'	A	B
t_1''	C	D

where all entries are positive and add up to 1. Let $\mu_i' := \beta_i^*(\theta = 2|t_i')$ and $\mu_i'' := \beta_i^*(\theta = 2|t_i'')$ denote i's first-order beliefs at the respective hierarchies. For $\mu > \frac{2}{3}$, the optimal minimal distribution is given by

e_μ^*	t_2'
t_1'	1

where $\mu_i' = \mu$ for $i = 1, 2$. The designer simply lets players act under common knowledge of μ, as investment is uniquely optimal in this range. For $\mu \leq \frac{2}{3}$, the following system of equations yields the minimal distributions e_μ^* such that all players invest at all t_i:

$$
\begin{cases}
\frac{A}{A+B}\mu_A + \frac{B}{A+B}\mu_B > \frac{2}{3} & (17) \\
\left(\frac{A}{A+C}\mu_A + \frac{C}{A+C}\mu_C\right) + \frac{1}{3}\frac{A}{A+C} > \frac{2}{3} & (18) \\
\left(\frac{C}{C+D}\mu_C + \frac{D}{C+D}\mu_D\right) + \frac{1}{3}\frac{C}{C+D} > \frac{2}{3} & (19) \\
\left(\frac{B}{B+D}\mu_B + \frac{D}{B+D}\mu_D\right) + \frac{1}{3} > \frac{2}{3} & (20) \\
A\mu_A + B\mu_B + C\mu_C + D\mu_D = \mu & (21)
\end{cases}
$$

The variables $\mu_A, \dots, \mu_D \in [0, 1]$ are the designer's beliefs that $\theta = 2$ given the corresponding pair of players' hierarchies.Inequalities (17) to (20) incorporate the (sufficient) conditions for consistency from Proposition 2: all players' first-order beliefs should be a weighted average of the designer's beliefs. For example, in (17),

$$\mu_1' := \frac{A}{A+B}\mu_A + \frac{B}{A+B}\mu_B$$

represents player 1's (first-order) belief that $\theta = 2$ at t_1'. Likewise, from (18) to (20), the expressions in parentheses represent first-order beliefs μ_2', μ_1'' and μ_2'', respectively. Condition (21) requires the average of the designer's beliefs to be equal to μ, which guarantees Bayes plausibility, also by Proposition 2.

Inequalities (17) to (20) are the investment constraints inherited from the equilibrium characterization in (16) and Proposition 5. For any t_i to be in $\cup_{k=1}^4 \rho_i^k$ for any $i = 1, 2$, some player must have a hierarchy at which he invests irrespective of the other player's action. That is, some player must have a first-order belief which, by itself, is enough to incentivize investment. Without loss, choose t_1' to play that role and require $\mu_1' > \frac{2}{3}$ (this is (17)). Given that player 1 invests based on first-order beliefs alone at t_1', player 2 invests based on second-order beliefs if $t_2' \in \rho_2^2$. This is equivalent to $\mu_2' + \frac{1}{3}\beta_2^*(t_1'|t_2') > \frac{2}{3}$, where

$$\begin{aligned} \mu_2' &:= \frac{A}{A+C}\mu_A + \frac{C}{A+C}\mu_C \\ \beta_2^*(t_1'|t_2') &:= \frac{A}{A+C}. \end{aligned} \tag{22}$$

This is precisely (18), where (22) is 2's second-order belief that $t_1 \in \rho_1^1$. Given that player 2 invests based on second-order beliefs at t_2', player 1 invests based on third-order beliefs if $t_1'' \in \rho_1^3$. This is equivalent to $\mu_1'' + \frac{1}{3}\beta_1^*(t_2'|t_2'') > \frac{2}{3}$ (condition (19)). Finally, player 2 invests at t_2'' based on fourth-order beliefs if $t_2'' \in \rho_2^4$. Since 1 invests no matter his hierarchy, $t_2'' \in \rho_2^4$ is equivalent to $\mu_2'' > \frac{1}{3}$ (condition (20)). In conclusion, the system of inequalities (17)-(21) describes the most efficient way of inducing investment with certainty. If the system has a solution, it must be e_μ^*.

The system has a solution for all $\mu > \frac{1}{2}$. At $\mu = \frac{1}{2}$, a solution exists if inequalities (17)-(20) are weak:

$e_{1/2}^*$	$t_2' : \mu_2' = 3/5$	$t_2'' : \mu_2'' = 1/3$
$t_1' : \mu_1' = 2/3$	1/8	1/8
$t_1'' : \mu_1'' = 4/9$	1/2	1/4

Since a small perturbation of $e_{1/2}^*$ solves the actual system for $\mu \downarrow \frac{1}{2}$, we use it as a reference.[16]

For $\mu \leq 1/2$, the designer can no longer ensure I will be played at all hierarchies. By Proposition 4, the hierarchy at which N is played must have a first-order belief of zero. Without loss, choose t_2'' to be that hieararchy: set $\mu_2'' = 0$ and replace (20) with

$$\frac{B}{B+D}\mu_B + \frac{D}{B+D}\mu_D = 0$$

in the above system. In this new system, player 1 plays I at both t_1' and t_1'', while player 2 plays I at t_2' and N at t_2''. The system has a solution for all $\mu > \frac{4}{9}$. For

[16] Under adversarial selection, there are priors at which an optimal solution does not exist. However, we can get arbitrarily close to the supremum value in Theorem 1 (for example, by using a perturbation of $e_{1/2}^*$ in our application).

example, at $\mu = \frac{9}{20}$, the solution is

$e^*_{9/20}$	$t'_2 : \mu'_2 = 9/14$	$t''_2 : \mu''_2 = 0$
$t'_1 : \mu'_1 = 2/3$	$1/20$	0
$t''_1 : \mu''_1 = 25/57$	$13/20$	$6/20$

For $\mu < 4/9$, the designer can only ensure I will be played at two hierarchies. In the original system, set both $\mu''_2 = 0$ and $\mu''_1 = 0$ instead of (20) and (19), respectively. Now, each player i plays I at t'_i and N at t''_i. This new system has a solution for any $\mu < \frac{4}{9}$. For example, at $\mu = \frac{1}{3}$, the solution is

$e^*_{1/3}$	$t'_2 : \mu'_2 = 4/9$	$t''_2 : \mu''_2 = 0$
$t'_1 : \mu'_1 = 2/3$	$1/2$	0
$t''_1 : \mu''_1 = 0$	$1/4$	$1/4$

Notice that the conditions from all of the above systems necessarily imply minimality of e^*_μ. Indeed, non-minimality would require either $A = D = 0$ or $B = C = 0$, neither of which can hold jointly with any of the systems for $\mu \leq \frac{2}{3}$.

Figure 1 plots (solid line) the designer's maximum-within value for $v(I, I) = 2$, $v(N, I) = v(I, N) = 1$, and $v(N, N) = 0$. Even without the maximization-between, private information plays an important role, for otherwise the designer would obtain a payoff of zero for all $\mu < \frac{2}{3}$ if the players acted under their prior information alone.

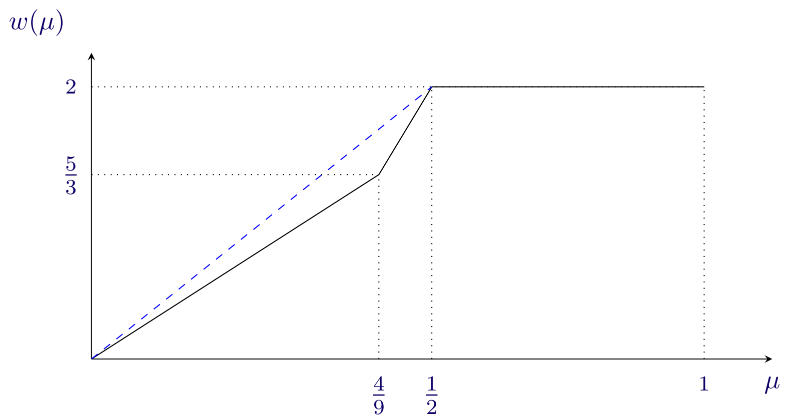

Figure 1: Value of maximization within (solid) and between (dashed)

Notice that under a direct approach based on a revelation principle-style argument, every player is given an action recommendation in place of the message inducing that action. For example, based on e^*_μ with $\mu > \frac{1}{2}$, the players should always be told to invest, which is completely uninformative. When players are completely uninformed for priors $\mu \in (\frac{1}{2}, \frac{2}{3})$, both (I, I) and (N, N) are BNE of the game. Since (N, N) would be picked by the adversarial selection rule, this shows that a direct approach fails to solve this type of information design problems.

5.2.2. Maximization Between

Concavification completes the solution (dashed line in Figure 1). For all $\mu \le \frac{1}{2}$, the designer uses a public signal and sends both players to $e^*_{1/2}$ with probability 2μ, where they both always play I, and to e^*_0 with probability $1 - 2\mu$, where they both know $\theta = -1$ and hence play N. The designer's value is given by $w^*(\mu) = 4\mu$ for $\mu \le \frac{1}{2}$ and $w^*(\mu) = 2$ for $\mu > \frac{1}{2}$. For $\mu \le \frac{1}{2}$, the optimal information structure that corresponds to the maximizations within and between is in Table 3.

$\pi^*(\cdot\,\vert\theta = -1)$	s'_2	s''_2	s
s'_1	$\frac{\mu}{30(1-\mu)}$	$\frac{4\mu}{30(1-\mu)}$	0
s''_1	$\frac{14\mu}{30(1-\mu)}$	$\frac{11\mu}{30(1-\mu)}$	0
s	0	0	$\frac{1-2\mu}{1-\mu}$

$\pi^*(\cdot\,\vert\theta = 2)$	s'_2	s''_2	s
s'_1	$\frac{13}{60}$	$\frac{7}{60}$	0
s''_1	$\frac{32}{60}$	$\frac{8}{60}$	0
s	0	0	0

Table 3: Optimal information structure

In the optimal information structure, each player i receives one of two private messages, s'_i or s''_i, or the public message s.[17] When players receive the public message s, it is common knowledge that $\theta = -1$, thus no one invests. Upon receiving private message s'_1, player 1 updates his first-order belief that $\theta = 2$ to $2/3$ and invests just based on this optimistic forecast. Player 2, upon observing s'_2, has a first-order belief of $4/9$, which is not optimistic enough to make him invest based on it alone. However, he puts sufficiently high probability on player 1 having received s'_1 that his higher-order belief that player 1 will invest incentivizes him to invest. The same logic is replicated for s''_1 and finally for s''_2. We refer to this contagion process à la Rubinstein (1989) as the *bandwagon effect*: starting from a "seed belief" that induces independent investment, each additional message is sent with the *optimal* probability in order to leverage beliefs of increasingly higher order about the other player's decision.

[17] Technically speaking, there are two public messages: one makes it common knowledge that $\theta = -1$ (message s), and one (call it $\tilde{s}$) makes it common knowledge that messages are in $\{s'_1, s''_1\} \times \{s'_2, s''_2\}$.

6. Conclusion

This paper contributes to the foundations of information design with multiple interacting agents. Our representation results formulate the belief-based approach to the problem, and decompose it into maximization within and between, where the latter is concavification. This approach accommodates various equilibrium selection rules and solution concepts, which can be used to analyze diverse topics, such as robustness, bounded rationality, collusion, or communication. We provide an economic application based on a two-agent investment game, and apply our approach to solve the information design problem under adversarial equilibrium selection. An obvious avenue for future research is to generalize this robust information design to a class of games with strategic complementarities, for which our results from the investment game provide the fundamental logic. Examining the implications of heterogenous prior distributions among the agents is another interesting extension of the current framework.

Appendix

A. Proofs of Main Results

A.1. Proof of Proposition 1

Let τ be induced by some (S, π), so that

$$\tau(t) = \sum_{\theta} \pi\big(\{s : h(s) = t\}|\theta\big)\mu_0(\theta) \tag{23}$$

for all $t \in \operatorname{supp} \tau$. Define $p \in \Delta(\Theta \times T)$ as

$$p(\theta, t) = \pi\big(\{s : h(s) = t\}|\theta\big)\mu_0(\theta) \tag{24}$$

for all θ and $t \in \operatorname{supp} \tau$. It is immediate from (23) and (24) that $\operatorname{marg}_T p = \tau$ and so $\operatorname{marg}_{T_i} p = \tau_i$ for all i. Further, when any player i forms his beliefs under (S, π), he computes $\mu_i : S_i \to \Delta(\Theta \times S_{-i})$ by conditioning $\pi(s|\theta)\mu_0(\theta)$ on s_i, so that $\beta_i^* : T_i \to \Delta(\Theta \times T_{-i})$ is given by the conditional of p given t_i. That is,

$$p(\theta, t) = \beta_i^*(\theta, t_{-i}|t_i)\operatorname{marg}_{T_i} p(t_i)$$

for all i, θ, and $t \in \operatorname{supp} \tau$. This shows $\tau \in \mathcal{C}$. Finally,

$$\sum_{t_i \in \operatorname{supp} \tau_i} \beta_i^*(\theta|t_i)\tau_i(t_i) := \sum_{t \in \operatorname{supp} \tau} p(\theta, t) = \sum_{t} \pi\big(\{s : h(s) = t\}|\theta\big)\mu_0(\theta) = \mu_0(\theta)$$

for all θ, which proves Bayes plausibility.

Suppose now that $\tau \in \mathcal{C}$ and satisfies Bayes plausibility. Let us show that these conditions are sufficient for τ to be induced by some some (S, π). Define information structure $(\operatorname{supp} \tau, \pi_\tau)$ where

$$\pi_\tau(t|\cdot) : \theta \mapsto \frac{1}{\mu_0(\theta)} \beta_i^*(\theta, t_{-i}|t_i)\tau_i(t_i) \tag{25}$$

for all $t \in \operatorname{supp} \tau$, which is defined independently of the choice of i because $\tau \in \mathcal{C}$. First, let us verify that π_τ is a valid information structure. Bayes plausibility says

$$\sum_{t_i \in \operatorname{supp} \tau_i} \beta_i^*(\theta|t_i)\tau_i(t_i) = \mu_0(\theta),$$

which guarantees that

$$\sum_{t \in \operatorname{supp} \tau} \pi_\tau(t|\theta) = \frac{1}{\mu_0(\theta)} \sum_{t \in \in \operatorname{supp} \tau} \beta_i^*(\theta, t_{-i}|t_i)\tau_i(t_i) = 1,$$

(i.e., $\pi(\cdot|\theta)$ is a probability distribution for every θ). By construction, this information structure is such that, when any player j receives t_j, his beliefs are $\mu_j(\cdot|t_j) = \beta_j^*(\cdot|t_j)$, also because $\tau \in \mathcal{C}$. To prove that π_τ generates τ, we also need to check that

$$\tau(t) = \sum_\theta \pi_\tau(t|\theta)\mu_0(\theta) \tag{26}$$

for all $t \in \operatorname{supp}\tau$. By (25), the rhs of (26) is equal to $\beta_i^*(t_{-i}|t_i)\tau_i(t_i)$, which equals $\tau(t)$ because $\tau \in \mathcal{C}$ (in particular, because $\operatorname{marg}_\Theta p = \tau$). □

A.2. Proof of Proposition 2

Suppose τ is consistent and Bayes plausible. Then, there exists $p \in \Delta(\Theta \times T)$ such that

$$p(\theta, t) = \beta_i^*(\theta, t_{-i}|t_i)p(t_i) \quad \forall \theta, t$$

and $\operatorname{marg}_T p = \tau$ (hence $p(t_i) = \tau_i(t_i)$ for all t_i). Define $\nu(t) := p(\cdot|t)$ for all t, which is the conditional distribution of θ given $t = (t_i)_i$ under p. Then,

$$\begin{aligned} \textstyle\sum_t \tau(t)\nu(t) &= \textstyle\sum_t \operatorname{marg}_T p(t) p(\cdot|t) \\ &= \textstyle\sum_{t_i} \operatorname{marg}_\Theta \beta_i^*(\cdot|t_i)p(t_i) \\ &= \mu_0. \end{aligned}$$

Moreover, for all t_i,

$$\begin{aligned} \textstyle\sum_{t_{-i}} \tau(t_{-i}|t_i)\nu(t_i, t_{-i}) &= \sum_{t_{-i}} \frac{p(\cdot|t)\tau(t)}{\tau_i(t_i)} \\ &= p(\cdot|t_i) \\ &= \operatorname{marg}_\Theta \beta_i^*(\cdot|t_i). \end{aligned}$$

Now, consider $\xi \in \Delta^f(\Delta\Theta)^n$ and $\nu : \operatorname{supp}\xi \to \Delta\Theta$ such that (8) and (9) hold. Denote $\nu(\boldsymbol{\mu}) = (\nu(\theta|\boldsymbol{\mu}))_\theta$ and define $g(\theta, \boldsymbol{\mu}) := \xi(\boldsymbol{\mu})\nu(\theta|\boldsymbol{\mu})$ for all θ and $\boldsymbol{\mu}$. We can think of

$$g(\boldsymbol{\mu}|\theta) = \frac{\nu(\theta|\boldsymbol{\mu})\xi(\boldsymbol{\mu})}{\mu_0(\theta)} \tag{27}$$

as the information map of information structure $(\operatorname{supp}\xi, g)$, because, by (8), we have $\sum_{\boldsymbol{\mu}} \nu(\theta|\boldsymbol{\mu})\xi(\boldsymbol{\mu}) = \mu_0(\theta)$, so that (27) is less than 1 for all θ and $\boldsymbol{\mu}$ and $\sum_{\boldsymbol{\mu}} g(\boldsymbol{\mu}|\theta) = 1$ for all θ. Denote by τ the belief hierarchy distribution induced by $(\operatorname{supp}\xi, g)$ (thus, $\tau(t) = \xi(\phi^{-1}(t))$ for all t). For any $t_i \in \operatorname{supp}\tau_i$ and $\mu_i = \phi^{-1}(t_i)$,

$$\beta_i^*(\theta|t_i) = \frac{\sum_{\boldsymbol{\mu}_{-i}} \nu(\theta|\boldsymbol{\mu})\xi(\boldsymbol{\mu})}{\sum_\theta \sum_{\boldsymbol{\mu}_{-i}} \nu(\theta|\boldsymbol{\mu})\xi(\boldsymbol{\mu})} = \frac{\mu_i(\theta)\xi(\mu_i)}{\xi(\mu_i)} = \mu_i(\theta)$$

for all θ. □

A.3. Proof of Proposition 3

We want to show that for all $\tau', \tau'' \in \mathcal{C}$ and $\alpha \in [0,1]$,

$$\alpha O_{\Sigma^{\mathrm{B}}}(\tau') + (1-\alpha) O_{\Sigma^{\mathrm{B}}}(\tau'') = O_{\Sigma^{\mathrm{B}}}(\alpha\tau' + (1-\alpha)\tau''),$$

where $\alpha O_{\Sigma^{\mathrm{B}}}(\tau) = \{\alpha\gamma : \gamma \in O_{\Sigma^{\mathrm{B}}}(\tau)\}$.

Take any $\tau', \tau'' \in \mathcal{C}^{\mathrm{M}}$ and $\alpha \in [0,1]$, let $\tau = \alpha\tau' + (1-\alpha)\tau''$ and take an arbitrary $\gamma \in O_{\Sigma^{\mathrm{B}}}(\tau)$. By definition, there exists $\sigma \in \Sigma^{\mathrm{B}}(\tau)$ such that

$$\gamma(a,\theta) = \sum_t \sigma(a|t) p_\tau(t,\theta) \tag{28}$$

for all (a,θ). Since Σ^{B} is invariant, there is $\sigma' \in \Sigma^{\mathrm{B}}(\tau')$ and $\sigma'' \in \Sigma^{\mathrm{B}}(\tau'')$ such that $\sigma(t) = \sigma'(t)$ for all $t \in \operatorname{supp}\tau'$, and $\sigma(t) = \sigma''(t)$ for all $t \in \operatorname{supp}\tau''$. Since $p_\tau = \alpha p_{\tau'} + (1-\alpha) p_{\tau''}$, letting

$$\gamma'(a,\theta) := \sum_{t \in \operatorname{supp}\tau'} \sigma'(a|t) p_{\tau'}(\theta,t) \tag{29}$$

$$\gamma''(a,\theta) := \sum_{t \in \operatorname{supp}\tau''} \sigma''(a|t) p_{\tau''}(\theta,t) \tag{30}$$

for all (a,θ), it follows from (28) that

$$\begin{aligned}
\gamma(a,\theta) &= \sum_t \sigma(a|t)(\alpha p_{\tau'}(\theta,t) + (1-\alpha) p_{\tau''}(\theta,t)) \\
&= \alpha \sum_{t \in \operatorname{supp}\tau'} \sigma'(a|t) p_{\tau'}(\theta,t) + (1-\alpha) \sum_{t \in \operatorname{supp}\tau''} \sigma''(a|t) p_{\tau''}(\theta,t) \\
&= \alpha\gamma'(a,\theta) + (1-\alpha)\gamma''(a,\theta).
\end{aligned}$$

Clearly, $\gamma' \in O_{\Sigma^{\mathrm{B}}}(\tau')$ and $\gamma'' \in O_{\Sigma^{\mathrm{B}}}(\tau'')$, so this completes the $\subseteq$ direction. For the $\supseteq$ direction, take $\gamma' \in O_{\Sigma^{\mathrm{B}}}(\tau')$ and $\gamma'' \in O_{\Sigma^{\mathrm{B}}}(\tau'')$. By definition, these distributions can be written as (29) and (30). From here, the argument is similar to the above, leading to $\alpha\gamma' + (1-\alpha)\gamma'' \in O_{\Sigma^{\mathrm{B}}}(\alpha\tau' + (1-\alpha)\tau'')$. □

A.4. Proof of Theorem 1

Lemma 1. *$\mathcal{C}$ is convex.*

Proof. Take $\alpha \in [0,1]$ and $\tau', \tau'' \in \mathcal{C}$. By definition of $\mathcal{C}$, there are $p_{\tau'}$ and $p_{\tau''}$ such that $\operatorname{marg}_T p_{\tau'} = \tau'$ and $\operatorname{marg}_T p_{\tau'} = \tau'$ and

$$\begin{aligned}
p_{\tau'}(\theta,t) &= \beta_i^*(\theta, t_{-i}|t_i)\tau_i'(t_i), \\
p_{\tau''}(\theta,t) &= \beta_i^*(\theta, t_{-i}|t_i)\tau_i''(t_i),
\end{aligned} \tag{31}$$

for all θ, i and t. Define $\tau := \alpha\tau' + (1-\alpha)\tau''$ and note that $\tau_i = \alpha\tau_i' + (1-\alpha)\tau_i''$, by the linearity of Lebesgue integral. Define

$$p_\tau(\theta, t) := \beta_i^*(\theta, t_{-i}|t_i)\tau_{\alpha,i}(t_i)$$

for all i, θ, and $t \in \operatorname{supp}\tau$. Notice that p_τ is well-defined, because of (31). Thus,

$$\operatorname{marg}_T p_\tau = \alpha \operatorname{marg}_T p_{\tau'} + (1-\alpha)\operatorname{marg}_T p_{\tau''} = \alpha\tau' + (1-\alpha)\tau'' = \tau$$

and we conclude that $\tau \in \mathcal{C}$. □

Although $\mathcal{C}$ is convex, it is not closed because we can build sequences in $\mathcal{C}$ with growing supports, only converging to a belief-hierarchy distribution with an infinite support. Still, the next lemma proves that minimal (consistent) distributions are the extreme points of the set of consistent distributions.

Lemma 2. *$\mathcal{E} = \mathcal{C}^M$.*

Proof. First, we show that $\mathcal{C}^M \subseteq \mathcal{E}$. If $\tau \notin \mathcal{E}$, then there exist $\tau', \tau'' \in \mathcal{C}$ with $\tau' \neq \tau''$ and $\alpha \in (0,1)$ such that $\tau := \alpha\tau' + (1-\alpha)\tau''$. By convexity of $\mathcal{C}$, $\tau \in \mathcal{C}$. Moreover, $\operatorname{supp}\tau' \cup \operatorname{supp}\tau'' \subseteq \operatorname{supp}\tau$. For $\lambda \in \mathbb{R}_+$, let $\tau_\lambda := \tau + \lambda(\tau - \tau'')$, which is a linear combination between τ' and τ''. Indeed, by construction $\tau - \tau'' = \alpha(\tau' - \tau'')$ and, therefore, we can rewrite $\tau_\lambda = \alpha(1+\lambda)\tau' + (1-\alpha(1+\lambda))\tau''$. Clearly, $\sum_{t \in \operatorname{supp}\tau} \tau_\lambda(t) = 1$ for all $\lambda \in \mathbb{R}_+$. Define $\Lambda := \{\lambda \geq 0 : \forall t \in \operatorname{supp}\tau, 0 \leq \tau_\lambda(t) \leq 1\}$ so that, by construction, $\tau_\lambda \in \mathcal{C}$ for all $\lambda \in \Lambda$. Next, we establish a number of simple properties of Λ. The set Λ is non-empty since both $\lambda = 0$ and $\lambda = \frac{1-\alpha}{\alpha}$ belong to it, which can be verified by substitution. Moreover, it is easy to check that Λ is convex. The set Λ is closed. To see this, it is enough to consider any increasing sequence $(\lambda_m) \subset \Lambda^\infty$ such that $\lambda_m \nearrow \lambda$. We want to show $\lambda \in \Lambda$. For all $t \in \operatorname{supp}\tau$, by definition of τ, the sequence $\tau_{\lambda_m}(t)$ is monotone (either non-decreasing or non-increasing) and, by definition of Λ, it lives in the compact interval $[0,1]$. Therefore, it converges in $\lim_m \tau_{\lambda_m}(t) \in [0,1]$. Hence, $\lambda \in \Lambda$. Therefore, we can write $\Lambda = [0, \tilde{\lambda}]$ where $\tilde{\lambda} := \max\Lambda \geq \frac{1-\alpha}{\alpha} > 0$.

We want to show that $\operatorname{supp}\tilde{\tau} \subsetneq \operatorname{supp}\tau$. To see this, let $\tilde{t} \in \operatorname{supp}\tilde{\tau}$ and suppose $\tilde{t} \notin \operatorname{supp}\tau$. Then, by definition, $\tilde{\tau}(\tilde{t}) = -\tilde{\lambda}\tau''(\tilde{t}) \leq 0$, which is impossible. Moreover, there must exist $t \in \operatorname{supp}\tau$ such that $\tau_{\tilde{\lambda}}(t) = 0$. To see this, suppose not, i.e. suppose $\operatorname{supp}\tau = \operatorname{supp}\tau_{\tilde{\lambda}}$. Then for all $t \in \operatorname{supp}\tau$ we would have that $\tau_{\tilde{\lambda}}(t) > 0$. Since $\tau_{\tilde{\lambda}} \in \mathcal{C}$ by construction and since $|\operatorname{supp}\tau| > 1$ (otherwise $\tau' = \tau''$), we also have that $\tau_{\tilde{\lambda}}(t) < 1$ for all $t \in \operatorname{supp}\tau$. Let $T^- := \{t \in \operatorname{supp}\tau : \tau(t) - \tau''(t) < 0\}$ and $T^+ := \{t \in \operatorname{supp}\tau : \tau(t) - \tau''(t) > 0\}$. These sets are non-empty by assumption ($\tau' \neq \tau''$). For $t \in T^-$ let $\lambda(t) := \frac{-\tau(t)}{\tau(t) - \tau''(t)}$ and notice that $0 = \tau(t) + \lambda(t)(\tau(t) -$

$\tau''(t)) < \tau_{\tilde{\lambda}}(t)$, implying $\lambda(t) > \tilde{\lambda}$. Similarly, for $t \in T^+$ let $\lambda(t) := \frac{1-\tau(t)}{\tau(t)-\tau''(t)}$ and notice that $1 = \tau(t) + \lambda(t)(\tau(t) - \tau''(t)) > \tau_{\tilde{\lambda}}(t)$, implying $\lambda(t) > \tilde{\lambda}$. Now define $\lambda' := \min\{\lambda(t) : t \in T^+ \cup T^-\}$, which is well-defined since $T^+ \cup T^-$ is finite. By construction, $\tau_{\lambda'} \in \mathcal{C}$ and $\lambda' > \tilde{\lambda}$, a contradiction to the fact that $\tilde{\lambda}$ is the max. Therefore, we conclude that $\operatorname{supp} \tilde{\tau} \subsetneq \operatorname{supp} \tau$ and thus $\tau \notin \mathcal{C}^{\mathrm{M}}$.

We now show the converse, $\mathcal{C}^{\mathrm{M}} \supseteq \mathcal{E}$. Suppose τ is not minimal, i.e., there is a $\tilde{\tau} \in \mathcal{C}$ such that $\operatorname{supp} \tilde{\tau} \subsetneq \operatorname{supp} \tau$. Define $\tau', \tau'' \in \Delta T$ as $\tau'(\cdot) := \tau(\,\cdot \mid \operatorname{supp} \tilde{\tau})$ and $\tau''(\cdot) := \tau(\,\cdot \mid \operatorname{supp} \tau \setminus \operatorname{supp} \tilde{\tau})$, the conditional distributions of τ on subsets $\operatorname{supp} \tilde{\tau}$ and $\operatorname{supp} \tau \setminus \operatorname{supp} \tilde{\tau}$. Clearly,

$$\tau = \alpha \tau' + (1-\alpha)\tau'' \tag{32}$$

where $\alpha = \tau(\operatorname{supp} \tilde{\tau}) \in (0,1)$. Since $\operatorname{supp} \tilde{\tau}$ is belief-closed, so is $\operatorname{supp} \tau \setminus \operatorname{supp} \tilde{\tau}$. Since τ' and τ'' are derived from a consistent τ and are supported on a belief-closed subspace, τ' and τ'' are consistent. Given that $\tau'' \neq \tau'$, (32) implies that τ is not an extreme point. $\square$

Proposition 6. *For any $\tau \in \mathcal{C}$, there exist unique $\{e_\ell\}_{\ell=1}^{L} \subseteq \mathcal{C}^{M}$ and weakly positive numbers $\{\alpha_\ell\}_{\ell=1}^{L}$ such that $\sum_{\ell=1}^{L} \alpha_\ell = 1$ and $\tau = \sum_{\ell=1}^{L} \alpha_\ell e_\ell$.*

Proof. Take any $\tau \in \mathcal{C}$. Either τ is minimal, in which case we are done, or it is not, in which case there is $\tau' \in \mathcal{C}$ such that $\operatorname{supp} \tau' \subsetneq \operatorname{supp} \tau$. Similarly, either τ' is minimal, in which case we conclude that there exists a minimal $e^1 := \tau'$ such that $\operatorname{supp} e^1 \subsetneq \operatorname{supp} \tau$, or there is $\tau'' \in \mathcal{C}$ such that $\operatorname{supp} \tau'' \subsetneq \operatorname{supp} \tau'$. Given that τ has finite support, finitely many steps of this procedure deliver a minimal consistent belief-hierarchy distribution e^1, $\operatorname{supp} e^1 \subsetneq \operatorname{supp} \tau$. Since τ and e^1 are both consistent and hence, their supports are belief-closed, $\operatorname{supp} \tau \setminus \operatorname{supp} e^1$ must be belief-closed. To see why, note that for any $t \in \operatorname{supp} \tau \setminus \operatorname{supp} e^1$, if there were $i, \hat{t} \in \operatorname{supp} e^1$ and $\theta \in \Theta$ such that $p_\tau(\theta, \hat{t}_{-i} | t_i) > 0$, then this would imply $p_\tau(\theta, t_i, \hat{t}_{-i}) > 0$ and, thus, $p_\tau(\theta, t_i, \hat{t}_{-(ij)} | \hat{t}_j) > 0$ (where $\hat{t}_{-(ij)}$ is the belief hierarchies of players other than i and j). This implies that at $\hat{t}_j$ — a hierarchy that player j can have in e^1 — player j assigns strictly positive probability to a type of player 1 that is not in the $\operatorname{supp} e^1$. This contradicts the fact that $\operatorname{supp} e^1$ is belief-closed. Given that $\operatorname{supp} \tau \setminus \operatorname{supp} e^1$ is a belief-closed subset of $\operatorname{supp} \tau$ and τ is consistent, define a new distribution τ^2 as

$$p_{\tau^2}(\theta, t) := \frac{p_\tau(\theta, t)}{\tau(\operatorname{supp} \tau \setminus \operatorname{supp} e^1)}$$

for all $\theta \in \Theta$ and $t \in \operatorname{supp} \tau \setminus \operatorname{supp} e^1$. By construction, $\operatorname{supp} \tau^2 = \operatorname{supp} \tau \setminus \operatorname{supp} e^1$. Moreover, since $\tau \in \mathcal{C}$, $p_\tau(\theta, t) = \beta_i^*(\theta, t_{-i} | t_i)\tau(t_i)$ for all $\theta \in \Theta$, $t \in \operatorname{supp} \tau$, and i. Hence,

$$p_{\tau^2}(\theta, t) = \frac{p_\tau(\theta, t)}{\tau(\operatorname{supp} \tau^2)} = \frac{\beta_i^*(\theta, t_{-i} | t_i)\tau(t_i)}{\tau(\operatorname{supp} \tau^2)} = \beta_i^*(\theta, t_{-i} | t_i)\tau^2(t_i)$$

for all $\theta \in \Theta$, $t \in \operatorname{supp} \tau^2$, and i. In addition,

$$\operatorname{marg}_T p_{\tau^2}(\theta, t) = \frac{\operatorname{marg}_T p_\tau(\theta, t)}{\tau(\operatorname{supp} \tau^2)} = \frac{\tau(t)}{\tau(\operatorname{supp} \tau^2)} = \tau^2(t)$$

for all $\theta \in \Theta$ and $t \in \operatorname{supp} \tau^2$. Hence, $\tau^2 \in \mathcal{C}$. Therefore, we can repeat the procedure again for distribution $\tau^2 \in \mathcal{C}$. Since τ has finite support, there exists $L \in \mathbb{N}$ such that, after L steps of this procedure, we obtain consistent a belief-hierarchy distribution τ^L that is also minimal. We denote $e^L := \tau^L$ and our procedure terminates. By construction, we have that for each $t \in \operatorname{supp} \tau$, there exists a unique $\ell \in \{1, \ldots, L\}$ such that $\tau = e^\ell \tau(\operatorname{supp} e^\ell)$. Therefore, $\tau = \sum_{\ell=1}^L \alpha_\ell e^\ell$ where $\alpha_\ell := \tau(\operatorname{supp} e^\ell) > 0$ and $\sum_{\ell=1}^L \alpha_\ell = \sum_{\ell=1}^L \tau(\operatorname{supp} e^\ell) = 1$.

Finally, we prove uniqueness. By way of contradiction, suppose that τ admits two minimal representations, that is,

$$\tau = \sum_\ell \alpha_\ell e^\ell = \sum_k \xi_k \hat{e}^k$$

such that $e^\ell \neq \hat{e}^k$ for some ℓ, k. This implies that for some $t \in \operatorname{supp} \tau$ and some ℓ, k, it holds that $t \in \operatorname{supp} e^\ell \cap \operatorname{supp} \hat{e}^k$ with $e^\ell \neq \hat{e}^k$. Two cases are possible:

(i) $\operatorname{supp} e^\ell \neq \operatorname{supp} \hat{e}^k$

Since $e^\ell, \hat{e}^k \in \mathcal{C}$, $\operatorname{supp} e^\ell$ and $\operatorname{supp} \hat{e}^k$ are belief-closed, which in turn implies $T^{\ell,k} := \operatorname{supp} e^\ell \cap \operatorname{supp} \hat{e}^k$ (nonempty by assumption) is also belief-closed. Therefore, there exists a distribution $e^{\ell,k}$ supported on $T^{\ell,k}$ and described by

$$p_{e^{\ell,k}}(\theta, t) := \frac{p_\tau(\theta, t)}{\tau(T^{\ell,k})}$$

for all $\theta \in \Theta$ and $t \in T^{\ell,k}$, which is consistent and $\operatorname{supp} e^{\ell,k} \subsetneq \operatorname{supp} e^\ell$. This contradicts the minimality of e^ℓ.

(ii) $\operatorname{supp} e^\ell = \operatorname{supp} \hat{e}^k$

Since $e^\ell, \hat{e}^k \in \mathcal{C}$, there exist common priors p and $\hat{p}$ in $\Delta(\Theta \times T)$ such that $\operatorname{marg}_T p = e_\ell$ and $\operatorname{marg}_T \hat{p} = \hat{e}^k$. Thus, $\operatorname{supp} \operatorname{marg}_T p = \operatorname{supp} \operatorname{marg}_T \hat{p}$. This implies $\operatorname{supp} p = \operatorname{supp} \hat{p}$ (If not, without loss there would exist some i, $\tilde{t} \in \operatorname{supp} \operatorname{marg}_T p$ and $\tilde{\theta} \in \Theta$, such that $\beta_i(\tilde{\theta}, \tilde{t}_{-i}|\tilde{t}_i) > 0$, while for all i and $t \in \operatorname{supp} \operatorname{marg}_T \hat{p}$, $\beta_i(\tilde{\theta}, t_{-i}|t_i) = 0$. This would contradict $\operatorname{supp} \operatorname{marg}_T p = \operatorname{supp} \operatorname{marg}_T \hat{p}$). By Propositions 4.4 and 4.5 in Mertens and Zamir (1985), there can be only one common prior with a given finite support in $\Delta(\Theta \times T)$, hence $p = \hat{p}$. In turn, $e^\ell = \hat{e}^k$, which contradicts that $e^\ell \neq \hat{e}^k$. $\square$

Now, we prove linearity of w. The point is to show that the set of outcomes of a mixture of subspaces of the universal type space can be written as a similar mixture of the sets of outcomes of these respective subspaces.

Lemma 3. *The function w is linear over $\mathcal{C}^M$.*

Proof. Let $\tau', \tau'' \in \mathcal{C}^M$ and $\alpha \in [0,1]$. Define $\tau = \alpha\tau' + (1-\alpha)\tau''$. Proposition 3 shows linearity of $O_{\Sigma^{\mathrm{B}}}$, so we have

$$\begin{aligned} w(\tau) &= \sum_{\theta,a} g(O_{\Sigma^{\mathrm{B}}}(\tau))[\theta,a]v(a,\theta) \\ &= \sum_{\theta,a} g\Big(\alpha O_{\Sigma^{\mathrm{B}}}(\tau') + (1-\alpha)O_{\Sigma^{\mathrm{B}}}(\tau'')\Big)[\theta,a]v(a,\theta) \end{aligned}$$

Since g is linear, this becomes

$$\begin{aligned} &\alpha\sum_{\theta,a} g\big(O_{\Sigma^{\mathrm{B}}}(\tau')\big)[\theta,a]v(a,\theta) + (1-\alpha)\sum_{\theta,a} g\big(O_{\Sigma^{\mathrm{B}}}(\tau'')\big)[\theta,a]v(a,\theta) \\ = \;& \alpha w(\tau') + (1-\alpha)w(\tau''), \end{aligned}$$

which completes the proof. ☐

Proof of Theorem 1. Fix a prior $\mu_0 \in \Delta(\Theta)$ and take any information structure (S,π). From Proposition 1, it follows that (S,π) induces a consistent belief-hierarchy distribution $\tau \in \mathcal{C}$ such that $\mathrm{marg}_\Theta\, p_\tau = \mu_0$. By definition of Σ^{B} and w, we have $V(S,\pi) \leq w(\tau)$ and, thus, $\sup_{(S,\pi)} V(S,\pi) \leq \sup\{w(\tau)|\tau \in \mathcal{C} \text{ and } \mathrm{marg}_\Theta p_\tau = \mu_0\}$. Moreover, Proposition 1 also implies that, for $\tau \in \mathcal{C}$ such that $\mathrm{marg}_\Theta\, p_\tau = \mu_0$, there exists an information structure (S,π) that induces τ and such that $V(S,\pi) = w(\tau)$. Therefore, $\sup_{(S,\pi)} V(S,\pi) \geq \sup\{w(\tau)|\tau \in \mathcal{C} \text{ and } \mathrm{marg}_\Theta\, p_\tau = \mu_0\}$. We conclude that

$$\sup_{(S,\pi)} V(S,\pi) = \sup_{\substack{\tau\in\mathcal{C} \\ \mathrm{marg}_\Theta\, p_\tau = \mu_0}} w(\tau). \tag{33}$$

By Proposition 6, there exists a unique $\lambda \in \Delta^f(\mathcal{C}^{\mathrm{M}})$ such that $\tau = \sum_{e\in\mathrm{supp}\,\lambda} \lambda(e)e$. Since p and marg are linear,

$$\mathrm{marg}_\Theta p_\tau = \mathrm{marg}_\Theta p_{\sum_e \lambda(e)e} = \sum_{e\in\mathrm{supp}\,\lambda} \lambda(e)\mathrm{marg}_\Theta p_e.$$

Then, by Lemma 3 and (33), we have

$$\begin{aligned} \sup_{(S,\pi)} V(S,\pi) = \sup_{\lambda\in\Delta^f(\mathcal{C}^{\mathrm{M}})} &\sum_e w(e)\lambda(e) \\ \text{subject to} &\sum_e \mathrm{marg}_\Theta p_e\, \lambda(e) = \mu_0, \end{aligned} \tag{34}$$

which concludes the proof. ☐

A.5. Proof of Proposition 4

Consider a consistent and Bayes plausible optimal minimal distribution τ centered at $\mu > 0$. First, we argue that, since τ is optimal, it must be that for both i there exists $t_i \in \operatorname{supp} \tau_i$ such that $t_i \in \rho_i$. If not, τ would be strictly dominated by $\hat{\tau} = x \cdot t_{\frac{2}{3}+\varepsilon} + (1-x) \cdot t_0$ with $x = \min\left\{\frac{\mu}{\frac{2}{3}+\varepsilon}, 1\right\}$, where $t_{\bar{\mu}}$ is a hierarchy profile at which $\beta_i^*(\{\theta = 2\}|t_i) = \bar{\mu}$ for all i is common knowledge. Indeed, the designer's expected payoff under $\hat{\tau}$ in the worst equilibrium is $x \cdot v(I, I) + (1-x) \cdot v(N, N)$. However, if $\rho_i = \emptyset$ for some i under τ, the designer's expected payoff is bounded above by $x \cdot v(I, N) + (1-x) \cdot v(N, N)$, which is strictly smaller than the payoff under $\hat{\tau}$.

By way of contradiction, suppose that for some player i we have $\tilde{t}_i \in \operatorname{supp} \tau_i$ such that $\tilde{t}_i \in \neg\rho_i$ and $\beta_i^*(\theta = 2|\tilde{t}_i) > 0$. Therefore, in the worst equilibrium, player i will play N at $\tilde{t}_i$, i.e., $\sigma_i^{\mathrm{MIN}}(I|\tilde{t}_i) = 0$. Denote

$$\underline{t}_i \in \operatorname*{argmin}_{t_i \in \rho_i \cap \operatorname{supp} \tau_i} \beta_i^*(\theta = 2|t_i)$$

for all i. Consider the distribution p' obtained from p_τ by transferring all the probability mass $\sum_{t_j} p_\tau(\tilde{t}_i, t_j, \theta = 2)$ to $p_\tau(\underline{t}_i, \underline{t}_j, \theta = 2)$. This operation changes the hierarchies in $\operatorname{supp} p_\tau$ according to the maps $f_i : \operatorname{supp} \operatorname{marg}_{T_i} p_\tau \to \operatorname{supp} \operatorname{marg}_{T_i} p'$ for all i. Formally,

(A) $p'(f_i(\underline{t}_i), f_j(\underline{t}_j), \theta = 2) = p_\tau(\underline{t}_i, \underline{t}_j, \theta = 2) + \sum_{t_j} p_\tau(\tilde{t}_i, t_j, \theta = 2)$,

(B) $p'(f_i(\tilde{t}_i), f_j(t_j), \theta = 2) = 0$ for all t_j, and

(C) $p'(f_i(t_i), f_j(t_j), \theta) = p_\tau(t_i, t_j, \theta)$ for all $t_i \neq \underline{t}_i, \tilde{t}_i$, t_j, and θ.

Now, consider the belief-hierarchy distribution τ' induced by p'. Conditions (A)-(B) imply:

(a) $\beta_j^*(\theta = 2|f_j(t_j)) \geq \beta_j^*(\theta = 2|t_j)$ for all $t_j \in \operatorname{supp} \tau_j$ with strict inequality for at least $\underline{t}_j$,

(b) $\beta_i^*(\theta = 2|f_i(t_i)) = \beta_i^*(\theta = 2|t_i)$ for all $t_i \in \operatorname{supp} \tau_i \setminus \{\tilde{t}_i, \underline{t}_i\}$,

(c) $\beta_i^*(\theta = 2|f_i(\tilde{t}_i)) = 0$, and

(d) $\beta_i^*(\theta = 2|f_i(\underline{t}_i)) > \beta_i^*(\theta = 2|\underline{t}_i)$ for all i.

By construction of p', we have

$$\sum_{T_i \times T_j} p'(f_i(t_i), f_j(t_j), \theta = 2) = \sum_{T_i \times T_j} p_\tau(t_i, t_j, \theta = 2) = \mu, \tag{35}$$

and hence, τ' is also centered at μ.

By (a)-(b), the first-order beliefs at all hierarchies except for $\tilde{t}_i$ have weakly increased (strictly increased for $\underline{t}_i$ and $\underline{t}_j$). (A)-(C) ensure that the higher-order beliefs of all t_i and i have increased, except for $\tilde{t}_i$. Therefore, if player i with hierarchy $t_i \in T_i \setminus \{\tilde{t}_i\}$ has uniquely rationalizable action I, and hence plays I in any equilibrium under τ, this also holds for player i with hierarchy $f_i(t_i)$ under τ'. On the other hand, player i with hierarchy $\tilde{t}_i$ played N in the worst equilibrium under τ, and this continues to hold for player i with hierarchy $f_i(\tilde{t}_i)$ under τ'.

Finally, note that by (A)-(C) we have

$$\tau'(f_i(\underline{t}_i), f_j(t_j)) \geq \tau(\underline{t}_i, t_j)$$

and

$$\tau'(f_i(\tilde{t}_i), f_j(t_j)) \leq \tau(\tilde{t}_i, t_j)$$

for all $t_j \in \operatorname{supp} \tau_j$ with strict inequality for some t_j. Hence, a designer with a symmetric and monotone payoffs would have a strictly higher expected payoff under τ' than under τ, a contradiction. $\square$

A.6. Proof of Proposition 5

The "if" part follows from (16) and $\cup_{k=1}^{m} \rho_i^k \subseteq \cup_{k\geq 1} \rho_i^k$. The "only if" part we prove by contradiction. Suppose $\hat{t}_i \in \tau_i^*$ and $\sigma_i^{\text{MIN}}(I|\hat{t}_i) = 1$, but $\hat{t}_i \notin \cup_{k=1}^{m} \rho_i^k$. Then, by (16) it must be that $\hat{t}_i \in \cup_{k\geq m+1} \rho_i^k$. Notice that $\rho_i^k \subseteq \rho_i^{k+1}$ for all $k \geq 1$. Thus, for any $t_i \in \rho_i$, there exists a number $k^*(t_i) \in \mathbb{N}^+$ such that $t_i \in \rho_i^{k^*(t_i)}$ for all $k^* \geq 1$ and $t_i \notin \cup_{k^*(t_i)-1\geq k\geq 1} \rho_i^k$ for all $k^* \geq 2$. That is, $k^*(t_i)$ is the smallest k such that $t_i \in \rho_i^k$. Let $n \geq 1$ be such that $k^*(\hat{t}_i) = m + n$. That is, $\hat{t}_i \in \rho_i^{m+n}$ and by definition

$$\beta_i^*(\{\theta = 2\} \times T_j|\hat{t}_i) + \frac{1}{3}\beta_i^*(\Theta \times \rho_j^{m+n-1}|\hat{t}_i) > \frac{2}{3},$$

while $\hat{t}_i \notin \rho_i^{m+n-1}$ and thus

$$\beta_i^*(\{\theta = 2\} \times T_j|\hat{t}_i) + \frac{1}{3}\beta_i^*(\Theta \times \rho_j^{m+n-2}|\hat{t}_i) \leq \frac{2}{3}.$$

This implies $\beta_i^*(\Theta \times \rho_j^{m+n-1}|\hat{t}_i) > \beta_i^*(\Theta \times \rho_j^{m+n-2}|\hat{t}_i)$, hence $\rho_j^{m+n-2} \subsetneq \rho_j^{m+n-1}$. Therefore, there exists $\tilde{t}_j \in \rho_j^{m+n-1}$ such that $\tilde{t}_j \notin \rho_j^{m+n-2}$, hence $k^*(\tilde{t}_j) = m+n-1$. By the same argument, there exists $\tilde{t}_i$ such that $k^*(\tilde{t}_i) = m+n-2$, and so on. This process continues for $m + n - 1$ steps in total, i.e. there needs to be $\bar{t}_j$ such that $k^*(\bar{t}_j) = 1$ if $m + n$ is even or $\bar{t}_i$ such that $k^*(\bar{t}_i) = 1$ if $m + n$ is odd. Hence, there need to be at least $m+n$ different hierarchies, which contradicts $m = \sum_{i=1,2} |\operatorname{supp} \tau_i^*|$. $\square$

B. ONLINE APPENDIX

Proposition 7.

(i) *Suppose that $\tau \in \mathcal{C}$ is conditionally independent. If $\mu = marg_{\Theta} p_\tau$ is not degenerate, then τ is minimal iff it is not perfectly informative. If μ is degenerate, then τ is minimal.*

(ii) *A public $\tau \in \mathcal{C}$ is minimal iff $\operatorname{supp} \tau$ is a singleton.*

Proof. Part (i). Suppose that τ is conditionally independent and μ is non-degenerate. First, if τ is perfectly informative, then it can be written

$$\tau = \sum_\theta \mu(\theta)\tau_\theta,$$

where τ_θ is a distribution that gives probability 1 to belief hierarchies representing common knowledge that θ has realized. Given $\mu(\theta) \in (0,1)$, τ is therefore a convex combination of belief-hierarchy distributions, hence it is not minimal.

Second, we show that if τ is non-minimal, then it must be perfectly informative. Let τ be non-minimal. By Proposition 6, there exist $\alpha \in (0,1)$ and $\tau' \neq \tau''$ such that $\operatorname{supp} \tau' \cap \operatorname{supp} \tau'' = \emptyset$ and $\tau = \alpha\tau' + (1-\alpha)\tau''$.

Now, take $t' \in \operatorname{supp} \tau'$, $t'' \in \operatorname{supp} \tau''$ and note

$$p_\tau(t'_i, t''_{-i}|\theta) = \alpha p_{\tau'}(t'_i, t''_{-i}|\theta) + (1-\alpha)p_{\tau''}(t'_i, t''_{-i}|\theta) = 0 \tag{36}$$

for all θ and i. If τ were conditionally independent,

$$\begin{aligned} & p_\tau(t'_i, t''_{-i}|\theta) \\ = \; & p_\tau(t'_i|\theta)\textstyle\prod_{j\neq i} p_\tau(t''_j|\theta) \\ = \; & \left(\alpha p_{\tau'}(t'_i|\theta) + (1-\alpha)p_{\tau''}(t'_i|\theta)\right)\textstyle\prod_{j\neq i}\left(\alpha p_{\tau'}(t''_j|\theta) + (1-\alpha)p_{\tau''}(t''_j|\theta)\right), \end{aligned}$$

which is strictly positive for some θ when τ is not perfectly informative, and thus contradicts (36). This implies that a non-minimal conditionally independent τ must be perfectly informative.

Part (ii). If τ is public, then every $\{t\}$ such that $t \in \operatorname{supp} \tau$ is a consistent distribution. Therefore, if $\operatorname{supp} \tau$ is a singleton, then it is clearly minimal. But if $\operatorname{supp} \tau$ is not a singleton, then τ is a convex combination of multiple consistent distributions, in which case τ is not minimal. □

Next we show that $\mathcal{C}^{\mathrm{M}}$ is small in a measure-theoretic sense relative to $\mathcal{C}$. Since there is no analog of the Lebesgue measure in infinite dimensional spaces, we use the notion of finite shyness proposed by Anderson and Zame (2001), which captures the idea of Lebesgue measure 0.

Definition 3. A measurable subset A of a convex subset C of a vector space $\mathcal{S}$ is finitely shy if there exists a finite dimensional vector space $V \subseteq \mathcal{S}$ for which $\lambda_V(C+s) > 0$ for some $s \in \mathcal{S}$, and $\lambda_V(A+s) = 0$ for all $s \in \mathcal{S}$, where λ_V is the Lebesgue measure defined on V.

Proposition 8. *$\mathcal{C}^M$ is finitely shy in $\mathcal{C}$.*

Proof. By Lemma 1, $\mathcal{C}$ is a convex subset of the vector space $\mathcal{S}$ of all signed measures on T. Choose any distinct $e, e' \in \mathcal{C}^M$ and let $V = \{\alpha(e-e') : \alpha \in \mathbb{R}\} \subseteq \mathcal{S}$. By construction, V is a one-dimensional subspace of $\mathcal{S}$. Let $\lambda_V \in \Delta V$ represent the Lebesgue measure on V. Notice that $\alpha(e-e') = \alpha e + (1-\alpha)e' - s$ for $s := e'$ and that $(\mathcal{C}-s) \cap V = \{\alpha(e-e') : \alpha \in [0,1]\}$ by convexity of $\mathcal{C}$. Hence, $\lambda_V(\mathcal{C}-s) > 0$. However, since $\mathcal{C}^M$ is the set of extreme points of $\mathcal{C}$, for every $s \in \mathcal{S}$, $(\mathcal{C}^M - s) \cap V$ contains at most two points. This gives $\lambda_V(\mathcal{C}^M - s) = 0$, since points have Lebesgue measure zero in V. $\square$

Proposition 9. *Let Σ be BNE and suppose that the selection criterion is max. For any minimal belief-hierarchy distribution $e \in \mathcal{C}^M$, there exist v and G for which $\lambda^* = \delta_e$ is the essentially unique optimal solution.*[18]

Proof. Fix $e \in \mathcal{C}^M$, $\varepsilon > 0$ and let $\mu_0 = p_e$. Denote by $G_\varepsilon(e) = (N, \{A_i, u_i\})$ the (base) game defined in Chen et al. (2010)'s Lemma 1 where $A_i \supseteq \operatorname{supp} e_i$. In this game, player i's actions include belief hierarchies from e_i. The (base) game $G_\varepsilon(e)$ is so conceived that, in the Bayesian game $(G_\varepsilon(e), e)$, every player has a strict incentive to truthfully report his true belief hierarchy, but for any τ that is suitably distant from e, some i in the Bayesian game $(G_\varepsilon(e), \tau)$ has a strict incentive *not* to report any $t_i \in \operatorname{supp} e_i$. For any i and $t_i, t_i' \in T_i$, let

$$d_i(t_i, t_i') := \sup_{k \geq 1} d^k(t_i, t_i')$$

where d^k is the standard metric (over k-order beliefs) that metrizes the topology of weak convergence. Let R_i and R be the ICR actions and profiles. Lemma 1 and Proposition 3 from Chen et al. (2010) imply that for every i and $t_i \in \operatorname{supp} e_i$,

$$t_i \in \operatorname*{argmax}_{a_i \in A_i} \sum_{\theta} \sum_{t_{-i} \in \operatorname{supp} e_{-i}} u_i(a_i, t_{-i}, \theta) \beta_i^*(\theta, t_{-i} | t_i)$$

[18]The proof establishes the stronger claim with $\Sigma :=$ ICR by using a result from Chen et al. (2010). A minimal distribution e is the essentially unique optimal solution if for all $\varepsilon > 0$, there is a game G_ε such that all $\tau \in \mathcal{C}$ with $d(e, \tau) > \varepsilon$ are strictly suboptimal (the metric is defined in (37)). By choosing a constant game or a constant designer's utility, it is easy to make all minimal distributions optimal, since the designer is indifferent among them. Uniqueness makes the result much stronger.

and for every $\tau \in \mathcal{C}$ such that

$$d(e,\tau) := \max_i d_i^H(\operatorname{supp} e_i, \operatorname{supp} \tau_i) \geq \varepsilon, \tag{37}$$

where d_i^H is the standard Hausdorff metric, there exist i and $t_i' \in \operatorname{supp} \tau_i$ such that $\operatorname{supp} e_i \cap R_i(t_i') = \emptyset$. To see why, note that since e is minimal, there can be no sequence $(\tau_n) \subseteq \mathcal{C}$ such that $d(e,\tau_n) \geq \varepsilon$ for all n, while

$$\max_i \max_{t_i \in \operatorname{supp} \tau_{n,i}} \min_{t_i' \in \operatorname{supp} e_i} d_i(t_i, t_i') \to 0. \tag{38}$$

That is, for all τ such that $d(e,\tau) \geq \varepsilon$, there is $\delta > 0$ such that

$$\max_i \max_{t_i \in \operatorname{supp} \tau_{n,i}} \min_{t_i' \in \operatorname{supp} e_i} d_i(t_i, t_i') \geq \delta. \tag{39}$$

Put differently, there exist i and $t_i' \in \operatorname{supp} \tau_i$ such that $d_i(t_i, t_i') \geq \delta > 0$ for all $t_i \in \operatorname{supp} e_i$. From (the proof of) Proposition 3 in Chen et al. (2010), we conclude that $\operatorname{supp} e_i \cap R^k(t_i') = \emptyset$ for some k. Given that $R_i(t_i') = \cap_{k=1}^{\infty} R_i^k(t_i')$, we have $\operatorname{supp} e_i \cap R_i(t_i') = \emptyset$. Now, define the designer's utility as $v(a) := \mathbb{1}(a \in \operatorname{supp} e)$ for all $a \in A$. Then, the designer's expected payoff is

$$w(\tau) = \begin{cases} 1 & \text{if } \tau = e, \\ x & \text{if } d(e,\tau) < \varepsilon, \\ y & \text{if } d(e,\tau) \geq \varepsilon \end{cases}$$

where $x \leq 1$ and $y < 1$. When $d(e,\tau) < \varepsilon$, it is not excluded that $x = 1$, because all of $\operatorname{supp} \tau$, by virtue of being close to some hierarchy in $\operatorname{supp} e$, might report only in $\operatorname{supp} e$. However, whenever $d(e,\tau) \geq \varepsilon$, there is a hierarchy profile t occurring with positive probability that reports outside $\operatorname{supp} e$. Thus, the designer maximizes her expected payoff by setting $\lambda^* = \delta_e$, which is Bayes plausible since $\mu_0 = p_e$. □

References

Alonso, Ricardo and Odilon Câmara. 2016. "Persuading Voters." *American Economic Review* 106 (11):3590–3605.

Anderson, Robert M. and William Zame. 2001. "Genericity with Infinitely Many Parameters." *The B.E. Journal of Theoretical Economics* 1 (1):1–64.

Angeletos, George-Marios and Alessandro Pavan. 2007. "Efficient Use of Information and Social Value of Information." *Econometrica* 75 (4):1103–1142.

Aumann, Robert J. 1976. "Agreeing to Disagree." *The Annals of Statistics* 4 (6):1236–1239.

Aumann, Robert J. and Michael B. Maschler. 1995. *Repeated Games with Incomplete Information*. Cambridge: MA: MIT Press.

Barelli, Paulo. 2009. "On the genericity of full surplus extraction in mechanism design." *Journal of Economic Theory* 144 (3):1320–1332.

Benoît, Jean-Pierre and Juan Dubra. 2011. "Apparent overconfidence." *Econometrica* 79 (5):1591–1625.

Bergemann, Dirk, Benjamin Brooks, and Stephen Morris. 2017. "First-Price Auctions with General Information Structures: Implications for Bidding and Revenue." *Econometrica* 85 (1):107–143.

Bergemann, Dirk and Stephen Morris. 2016. "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games." *Theoretical Economics* 11:487–522.

———. 2018. "Information Design: A Unified Perspective." *Journal of Economic Literature* .

Best, James and Daniel Quigley. 2018. "Persuasion for the Long-Run." Oxford University WP.

Blackwell, David and M. A. Girshick. 1954. *Theory of Games And Statistical Decisions*. New York: John Wiley and Sons.

Brandenburger, Adam and Eddie Dekel. 1993. "Hierarchies of Beliefs and Common Knowledge." *Journal of Economic Theory* 59 (1):189–198.

Brocas, Isabelle and Juan. D Carrillo. 2007. "Influence through ignorance." *The RAND Journal of Economics* 38 (4):931–947.

Carlsson, H. and E. van Damme. 1993. "Global games and equilibrium selection." *Econometrica* 61 (5):989–1018.

Chan, Jimmy, Seher Gupta, Fei Li, and Yun Wang. 2016. "Pivotal Persuasion." WP.

Chen, Yi-Chun, Alfredo Di Tillio, Eduardo Faingold, and Siyang Xiong. 2010. "Uniform topologies on types." *Theoretical Economics* 5 (3).

Dekel, Edie, Drew Fudenberg, and Stephen Morris. 2007. "Interim Correlated Rationalizability." *Theoretical Economics* 2:15–40.

Ely, Jeffrey. 2017. "Beeps." *American Economic Review* 107 (1):31–53.

Ely, Jeffrey, Alexander Frankel, and Emir Kamenica. 2015. "Suspense and Surprise." *Journal of Political Economy* 123 (1):215–260.

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

Heifetz, Aviad and Zvika Neeman. 2006. "On the Generic (Im)Possibility of Full Surplus Extraction in Mechanism Design." *Econometrica* 74 (1):213–233.

Hoshino, Tetsuya. 2018. "Using Strategic Uncertainty to Persuade Multiple Agents." Pennsylvania State University, WP.

Hurwicz, Leonid. 1986. "On Informational Decentralization and Efficiency in Resource Allocation Mechanisms." In *Studies in Mathematical Economics*, edited by S. Reiter. Washington, DC: Mathematical Association of America.

Hurwicz, Leonid and Thomas Marschak. 2003. "Comparing finite mechanisms." *Economic Theory* 21 (4):783–841.

Inostroza, Nicolas and Alessandro Pavan. 2017. "Persuasion in Global Games with Application to Stress Testing." WP, Northwestern University.

Kajii, Atsushi and Stephen Morris. 1997. "The Robustness of Equilibria to Incomplete Information." *Econometrica* 65 (6):1283–1309.

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

Kolotilin, Anton, Ming Li, Tymofiy Mylovanov, and Andriy Zapechelnyuk. 2017 (forthcoming). "Persuasion of a Privately Informed Receiver." *Econometrica* 85 (6):1949–1964.

Laclau, Marie and Ludovic Renou. 2016. "Public Persuasion." WP.

Lipnowski, Elliot and Laurent Mathevet. 2018. "Disclosure to a Psychological audience." *AEJ: Microeconomics.*

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

———. 2015. “Correlation and common priors in games with incomplete information.” *Journal of Economic Theory* 157 (C):49–75.

Maskin, Eric. 1999. “Nash Equilibrium and Welfare Optimality.” *Review of Economic Studies* 66 (1):23–38.

Mathevet, Laurent, David Pearce, and Ennio Stacchetti. 2018. “Reputation and Information Design.” WP, New York University.

Mertens, Jean-Francois and Shmuel Zamir. 1985. “Formulation of Bayesian analysis for games with incomplete information.” *International Journal of Game Theory* 14 (1):1–29.

Morris, Stephen and Hyun Song Shin. 2002. “Social Value of Public Information.” *American Economic Review* 92 (5):1521–1534.

———. 2003. “Global games: theory and applications.” In *Advances in Economics and Econometrics, the Eighth World Congress*. Cambridge University Press.

Rayo, Luis and Ilya Segal. 2010. “Optimal Information Disclosure.” *Journal of Political Economy* 118 (5):949 – 987.

Rockafellar, R. Tyrrell. 1970. *Convex Analysis*. Princeton, New Jersey: Princeton University Press, 2nd ed.

Rubinstein, Ariel. 1989. “The Electronic Mail Game: Strategic Behavior under “Almost Common Knowledge”.” *American Economic Review* 79 (3):385–91.

Taneva, Ina. 2014. “Information Design.” Economics Department, Edinburgh University.

Vives, Xavier. 1988. “Aggregation of Information in Large Cournot Markets.” *Econometrica* 56 (4):851–876.

Yildiz, Muhamet. 2015. “Invariance to representation of information.” *Games and Economic Behavior* 94 (C):142–156.

Zhang, Jun and Junjie Zhou. 2016. “Information Disclosure in Contests: A Bayesian Persuasion Approach.” *Economic Journal* 126:2197–2217.