THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

On information design in games

Citation for published version:
Mathevet, L, Perego, J & Taneva, I 2020, 'On information design in games', *Journal of Political Economy*, vol. 128, no. 4, pp. 1370-1404. https://doi.org/10.1086/705332

Digital Object Identifier (DOI):
10.1086/705332

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Journal of Political Economy

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 24. Feb. 2026

On Information Design in Games

Laurent Mathevet

New York University

Jacopo Perego

Columbia University

Ina Taneva

University of Edinburgh

Information provision in games influences behavior by affecting agents' beliefs about the state as well as their higher-order beliefs. We first characterize the extent to which a designer can manipulate agents' beliefs by disclosing information. We then describe the structure of optimal belief distributions, including a concave-envelope representation that subsumes the single-agent result of Kamenica and Gentzkow. This result holds under various solution concepts and outcome selection rules. Finally, we use our approach to compute an optimal information structure in an investment game under adversarial equilibrium selection.

I. 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,

We are grateful to the editor, Emir Kamenica, and to three anonymous referees for their comments and suggestions, which significantly improved the paper. Special thanks are due to Anne-Katrin Roesler for her insightful comments as a discussant at the 2016 Cowles

Electronically published March 3, 2020
[*Journal of Political Economy*, 2020, vol. 128, no. 4]
© 2020 by The University of Chicago. All rights reserved. 0022-3808/2020/12804-00XX$10.00

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

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 or not to invest 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 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, for example, the designer is constrained to distributions over beliefs about the state that on average equal the prior, a condition known as *Bayes plausibility* (Kamenica and Gentzkow 2011). In the multiagent case, an additional requirement emerges: agents' beliefs should be consistent with one another. We further establish an equivalence between (Bayes-plausible and consistent) distributions of agents' beliefs and distributions of the designer's beliefs, which is particularly useful in applications.

Second, we represent the designer's problem in a way that exploits the structure of consistent belief distributions. We show that every consistent

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, the Paris School of Economics, Stony Brook University, the University of Cambridge, the University of Edinburgh, the University of Texas at Austin, the SIRE (Scottish Institute for Research in Economics) BIC (Behaviors, Incentives and Contracts) 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 the SAET (Society for the Advancement of Economic Theory) Conference 2017. Taneva acknowledges financial support from ESRC (Economic and Social Research Council) grant ES/N00776X/1.

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

belief distribution can be represented as a (convex) combination of more "basic" elements, themselves belief distributions. Therefore, the problem of choosing an optimal information structure is equivalent to choosing an optimal combination of such basic elements, subject to Bayes plausibility. From this follows a two-step approach to the information-design problem: the first step optimizes among the basic elements only, and the second step optimally combines the elements selected in the first step, subject to Bayes plausibility. The latter step corresponds to the concavification of a value function derived in the first step. Therefore, this representation subsumes the single-agent result of Kamenica and Gentzkow (2011). This two-step decomposition can be interpreted as optimizing over private information in the first step and then adding an optimal public signal in the second step.

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 have only bounded depths of reasoning, that they can deviate in coalitions, or that they can communicate. A designer may also want to hedge against the possibility that, when there are multiple outcomes (consistent with the solution concept), agents might coordinate on the outcome most unfavorable to her. This can be achieved by choosing a robust information structure, which maximizes the designer's payoff under an adversarial selection rule. 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 may want to 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. We use our approach to compute the optimal information structure that takes on a simple form: every agent receives either a private message that makes investing uniquely optimal because of a combination of first- and higher-order beliefs or a public message that makes it common knowledge that the state is low, and 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 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.

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

The single-agent problem has been a rich subject of study since the influential work of Kamenica and Gentzkow (2011). 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.[1] By contrast, the theory of information design in games is not as well understood. Bergemann and Morris (2016) and Taneva (2019) 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 multiagent analog to Kamenica and Gentzkow (2011)'s single-agent formulation. Earlier works have studied the effects of public and private information on equilibrium behavior, efficiency, and welfare (e.g., Vives 1988; Morris and Shin 2002; Angeletos and Pavan 2007). More recent papers study the optimal design of information in voting games (Alonso and Câmara 2016; Chan et al. 2019), dynamic bank runs (Ely 2017), stress testing (Inostroza and Pavan 2017), auctions (Bergemann, Brooks, and Morris 2017), or contests (Zhang and Zhou 2016) or focus on public information in games (Laclau and Renou 2016).

II. The Information-Design Problem

A set $N = \{1, \ldots, n\}$ of agents interact in an uncertain environment. The variable $\theta \in \Theta$ describes the uncertain state of the world, where the set Θ is finite. Every agent $i \in N$ 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 agents know only 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 agents about the payoff-relevant state θ. This is modeled by an *information structure* (S, π), where $S_i \subseteq \mathcal{S}$ is the finite set of messages that agent i can receive, $S = \prod_i S_i$ is the set of message profiles, and $\pi : \Theta \to \Delta(S)$ is the information

[1] Other early contributions to single-agent information design include models by Brocas and Carrillo (2007) and Rayo and Segal (2010).

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

map.[2] In any state θ, the message profile $s = (s_i)$ is drawn according to $\pi(s|\theta)$, and agent 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 test. 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 knows only its distribution μ_0.[3] 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* $\Sigma(\mathcal{G}) \subseteq \{\sigma = (\sigma_i) \mid \sigma_i : S_i \to \Delta A_i \text{ for all } i\}$. 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 nonempty and compact valued. Given that G is finite, this holds when Σ is a 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 that associates to any $D \subseteq \Delta(\Theta \times A)$ an element $g(D) \in D$. 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

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

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

[2] We restrict our attention to finite message spaces, because doing so guarantees existence of a BNE for all information structures. Any infinite set $\mathcal{S}$ would serve the purpose.

[3] 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; Mathevet, Pearce, and Stacchetti 2018).

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

III. Information Design as Belief Manipulation

We reformulate information design into a belief-manipulation problem, analogously to Kamenica and Gentzkow's (2011) 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 agents' beliefs are, in addition, consistent with each other.

A. *Belief Distributions*

A *belief hierarchy* t_i for agent i is an infinite sequence $(t_i^1, t_i^2, \ldots)$ whose components are coherent beliefs of all orders:[4] $t_i^1 \in \Delta\Theta$ 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. Even if the belief hierarchies are coherent, they may assign positive probability to other agents' 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.[5] This map describes i's beliefs about (θ, t_{-i}), given t_i, and shows that there are sets of coherent belief hierarchies for all agents that put only positive probabilities on each other, making coherency common knowledge. Let $T := \prod_i T_i$ be the space of hierarchy profiles with common knowledge of coherency.

Given μ_0 and an information structure (S, π), and upon receiving a message s_i, agent i formulates beliefs $\mu_i(s_i) \in \Delta(\Theta \times S_{-i})$ about the state and other agents' messages in $S_{-i} := \prod_{j \neq i} S_j$ by using Bayes's rule. 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.[6] 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 belief hierarchy $h_i(s_i) \in T_i$ for agent i, and every message profile s induces a profile of belief hierarchies $h(s) := (h_i(s_i))_{i \in N}$.

[4] 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 $X_{k-1} = \text{supp}\, t_i^{k-1}$.

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

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

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

TABLE 1
A (Public) INFORMATION STRUCTURE

	s_1'	s_2''
$\pi(\cdot\|0)$:		
s_1'	1	0
s_2''	0	0
$\pi(\cdot\|1)$:		
s_1'	1/2	0
s_2''	0	1/2

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

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

for all t.

For example, the information structure in table 1 induces $\tau = (3/4)t_{1/3} + (1/4)t_1$ when $\mu_0 := \mu_0(\theta = 1) = 1/2$, where t_μ is the hierarchy profile in which $\mu := \mu(\theta = 1)$ is commonly believed.[7]

We categorize belief distributions as public or 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 $\text{marg}_{T_{-i}}\beta_i^*(\cdot|t_i) = \delta_{t_{-i}}$ (where δ is the Dirac measure) for all $t \in \text{supp}\,\tau$ and $|\text{supp}\,\tau| \geq 2$. A belief distribution τ is *private* if it is not public.

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

B. *Manipulation*

Consider an environment with two agents and two equally likely states $\theta \in \{0, 1\}$. For each agent $i = 1, 2$, consider two belief hierarchies, t_i and t_i', such that

$$\beta_i^*(t_j|t_i) = \beta_i^*(t_j'|t_i') = 1 \quad \forall\ i, j \neq i, \tag{4}$$

[7] To see why, note that $\Pr(s_1', s_2') = 3/4$, $\Pr(s_1'', s_2'') = 1/4$, and an agent i receiving message s_i' has belief 1/3 that $\theta = 1$ and is certain that j also received s_j', while s_i'' has belief 1 that $\theta = 1$ and is certain that j also received s_j''.

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

$\beta_1^*(\theta = 1|t_1) = \beta_1^*(\theta = 1|t_2') = 0.8$, and $\beta_1^*(\theta = 1|t_1') = \beta_1^*(\theta = 1|t_2) = 0.2$. In words, at (t_1, t_2) and (t_1', t_2'), the agents believe that $\theta = 1$ with different probabilities, 0.8 or 0.2, and this disagreement is commonly known. Can the belief-hierarchy distribution $\tau(t) = \tau(t') = 1/2$ be induced by some information structure? More generally, can the designer ever get agents to agree to disagree? 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

$$p(\theta, t) = \beta_i^*(\theta, t_{-i}|t_i)p(t_i) \tag{5}$$

for all θ, t, and i. That is, all agents 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

$$\mathcal{C} := \left\{\tau \in \Delta^f T : \exists \text{ a common prior } p \text{ such that } \tau = \text{marg}_T p\right\} \tag{6}$$

to be the space of *consistent* (belief-hierarchy) distributions. In a consistent distribution, all agents' beliefs arise from a common prior that draws every t with the same probability as τ, that is, $\tau = \text{marg}_T p$. Let p_τ be the unique distribution p in (6) (uniqueness follows from proposition 4.5 of Mertens and Zamir 1985).

Note that consistency does not require that $\text{marg}_\Theta p = \mu_0$, which is a conceptually different point. A distribution $\tau \in \Delta^f T$ is *Bayes plausible for agent* i if

$$\sum_{t_i} \text{marg}_\Theta \beta_i^*(\cdot|t_i)\tau_i(t_i) = \mu_0,$$

that is, if agent i's expected first-order belief equals the state distribution.

Proposition 1. There exists an information structure that induces $\tau \in \Delta^f T$ if and only if τ is consistent and Bayes plausible for some agent i.

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

Note, however, that merely ensuring that Bayes plausibility holds for all agents does not guarantee consistency. In the simple example above, τ is Bayes plausible for both agents and yet fails consistency, because $\beta_1^*(\theta = 1, t_2'|t_1') \cdot (1/2) = 0.2 \cdot (1/2) \neq \beta_2^*(\theta = 1, t_1'|t_2') \cdot (1/2) = 0.8 \cdot (1/2)$ violates equation (5).

From an operational viewpoint, there are two distinct ways of designing a consistent belief-hierarchy distribution. The first one is to design

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

the distributions of agents' beliefs individually and then couple them so as to ensure consistency (see Ely 2017 for a related procedure). A different approach, formulated in the next proposition, is to design the distribution of designer's beliefs and then to derive from it the resulting consistent distribution of agents' beliefs.

PROPOSITION 2. $\tau \in \Delta^{f}(T)$ is consistent and Bayes plausible if and only if there exists $\nu : \text{supp}\,\tau \to \Delta\Theta$ such that, for all t, θ, and i,

$$\sum_{t} \tau(t)\nu(\theta|t) = \mu_0(\theta), \tag{7}$$

$$\nu(\theta|t_i, t_{-i}) = \beta_i^*(\theta|t_i, t_{-i}) := \frac{\beta_i^*(\theta, t_{-i}|t_i)}{\beta_i^*(t_{-i}|t_i)}, \tag{8}$$

$$\tau(t_{-i}|t_i) = \beta_i^*(t_{-i}|t_i). \tag{9}$$

Whereas each agent i observes only t_i, the designer observes the entire hierarchy profile $t = (t_1, ..., t_n)$ and hence is better informed than every single agent. In the above, $\nu(t)$ is interpreted as the designer's beliefs about the state, obtained by conditioning on t. The result demonstrates that any consistent and Bayes-plausible belief-hierarchy distribution corresponds to a specific distribution of designer's beliefs, and vice versa—a relationship governed by three conditions. The first one states that the average of the designer's beliefs is equal to the distribution of states. This is a form of Bayes plausibility at the designer's level. The second requires that the designer's beliefs about the state be the same as what any agent would believe if he also knew the information of all the other agents. The last condition requires that i's conditional beliefs about the other agents' hierarchies can be derived from τ by conditioning on i's belief hierarchy. This result suggests a different way of approaching the design of (consistent and Bayes-plausible) belief-hierarchy distributions: it is equivalent to designing the distribution of the designer's beliefs subject to Bayes plausibility in equation (7) and deriving the agents' hierarchies via equations (8) and (9). We demonstrate how this can be a useful way of implementing consistency in section V.

C. *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. For any consistent distribution τ, let a solution concept be a collection $\Sigma^{B}(\tau)$ of sets $\Lambda \subseteq \{\sigma : \text{supp}\tau \to \Delta A\}$.[8] This describes agents' behavior in the

[8] We thank Daniel Clark for pointing out an error; this led to a substantial improvement of this section.

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

Bayesian game $\langle G, p_\tau \rangle$, where p_τ is the unique common prior p such that $\text{marg}_T p = \tau$. The different Λs capture possibly correlated behaviors, enabled by the addition of different (belief-preserving) correlating devices to τ (see Liu 2009). Given any $\Lambda \in \Sigma^B(\tau)$, the resulting outcomes are

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

from which it follows that the designer's ex ante expected payoff from a consistent distribution τ is

$$w(\tau) := \sup_{\Lambda \in \Sigma^B(\tau)} \sum_{\theta, a} g(O_\Lambda(\tau))(a, \theta) v(a, \theta). \tag{10}$$

Note that Σ^B is useful only insofar as it captures the appropriate outcomes from solution concept Σ. Hence, we assume that $\Lambda \in \Sigma^B(\tau)$ if and only if Λ is such that $O_\Lambda(\tau) = O_\Sigma(S, \pi)$ for some (S, π) inducing τ. In general, it is well known that a given solution concept may not yield the same set of outcomes when applied to an information structure as when applied to the corresponding belief distribution (e.g., Ely and Peski 2006; Liu 2009). Indeed, even if (S, π) induces τ, there may be multiple message profiles s inducing the same hierarchy profile t, thus creating opportunities for redundant correlations of agents' behavior that are not possible in τ. In some cases, the literature makes clear which Σ^B should be chosen, given Σ. For example, if Σ = BNE, then said correlations can be recovered (without affecting the beliefs in τ) by taking Σ^B to be Liu's (2015) belief-preserving correlated equilibrium. Alternatively, if Σ is interim correlated rationalizability (ICR; Dekel, Fudenberg, and Morris 2007), then Σ^B should also be ICR.

IV. 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.

A. *Assumptions*

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

Linear Selection. For all $D', D'' \subseteq \Delta(\Theta \times A)$ and $0 \leq \alpha \leq 1$, $g(\alpha D' + (1 - \alpha)D'') = \alpha g(D') + (1 - \alpha) g(D'')$.

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

INVARIANT SOLUTION. Fix $\tau, \tau', \tau'' \in \mathcal{C}$ such that $\operatorname{supp}\tau = \operatorname{supp}\tau' \cup \operatorname{supp}\tau''$ and $\operatorname{supp}\tau' \cap \operatorname{supp}\tau'' = \varnothing$. $\Lambda \in \Sigma^{B}(\tau)$ if and only if there exist $\Lambda' \in \Sigma^{B}(\tau')$ and $\Lambda'' \in \Sigma^{B}(\tau'')$ such that $\Lambda = \{\sigma : \operatorname{supp}\tau \to \Delta A : \sigma|_{\operatorname{supp}\tau'} \in \Lambda' \text{ and } \sigma|_{\operatorname{supp}\tau''} \in \Lambda''\}$.

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's (2015) correlated equilibrium satisfies invariance. And when $\Sigma = \text{ICR}$ (Dekel, Fudenberg, and Morris 2007), $\Sigma^{B} = \Sigma$, which also satisfies invariance. Invariance is important because it allows us to recover the outcomes from any consistent distribution through appropriate randomizations over distributions with smaller supports and, thus, ensures linearity of the ex-ante payoff (proposition 4 and lemma 3 in sec. A3).

B. *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.[9] In the tradition of extremal representation theorems,[10] 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 sec. A3). 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 $\mathcal{C}^{M}$ denote the set of all minimal distributions,[11] 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,

[9] 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 $\tau'' = \tau$.

[10] For example, the Minkowski-Caratheodory theorem, the Krein-Milman theorem, and Choquet's integral representation theorem.

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

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

any nonminimal distribution τ contains two consistent components with support supp τ' and supp$\tau\backslash$suppτ'. A public signal makes it common knowledge among the agents which of these two components their beliefs are in. Since a minimal belief distribution has only one component, it contains 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 elements from which all other consistent distributions 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.[12]

THEOREM 1 (Representation theorem). The designer's maximization problem can be represented as

$$\sup_{(S,\pi)} V(S,\pi) = \sup_{\lambda\in\Delta^{f}(\mathcal{C}^{M})} \sum_{e\in\text{supp}\lambda} w(e)\lambda(e) \quad \text{subject to} \sum_{e\in\text{supp}\lambda} \text{marg}_{\Theta} p_e \lambda(e) = \mu_0. \tag{11}$$

COROLLARY 1 (Within-between maximizations). For any $\mu \in \Delta\Theta$, let

$$w^{*}(\mu) := \sup_{e\in\mathcal{C}^{M}:\,\text{marg}_{\Theta} p_e=\mu} w(e). \tag{12}$$

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

$$\sup_{(S,\pi)} V(S,\pi) = \sup_{\lambda\in\Delta^{f}\Delta\Theta} \sum_{\text{supp}\lambda} w^{*}(\mu)\lambda(\mu), \quad \text{subject to} \sum_{\text{supp}\lambda} \mu\lambda(\mu) = \mu_0. \tag{13}$$

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 (marginal) distribution on states, $\text{marg}_{\Theta} p_e = \mu$, and the "farther" that is from μ_0, the "costlier" it is for the designer to use it. In this sense, the constraint in equation (11) can be seen as a form of budget constraint.

[12] We further illustrate the notion of minimal distribution in the supplementary appendix (available online), by characterizing minimal distributions for public and conditionally independent information.

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

The corollary decomposes the representation theorem into two steps. First, there is a *maximization within*—given by equation (12)—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, that is, 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 agents to different minimal distributions e, thus making e common knowledge. From standard arguments (Rockafellar 1970, 36), the right-hand side of equation (13) 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, $\{e \in \mathcal{C}^M \text{ such that } \text{marg}_\Theta p_e = \mu\} = \{\mu\}$; hence, $w^* = w$ in equation (12) and the theorem reduces to maximization between.

The above decomposition is most useful when maximization within is performed over a restricted set of minimal belief distributions. Such restrictions are made in the context of constrained information design. For example, in some applications, it may be appropriate to limit the maximization within to conditionally independent private information. Alternatively, there are environments in which imposing restrictions can be done without loss of generality. This is shown in the application of the next section, where consistency, Bayes plausibility, and the payoff externalities impose enough structure that the unconstrained optimal solution can be computed by restricting maximization within to a small subset of minimal belief distributions. Whether these restrictions constrain the optimal solution or not, the restricted set of minimal belief distributions forms the smallest class of distributions from which that optimal solution can be derived by means of concavification.

V. Application: Fostering Investment under 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,

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

TABLE 2
INVESTMENT GAME

(u_1, u_2)	Invest (I)	Do Not Invest (N)
Invest (I)	θ, θ	$\theta - 1, 0$
Do not invest (N)	$0, \theta - 1$	$0, 0$

the strategic provision of information by a third party constitutes a different way of leveraging the underlying complementarities.

We consider the problem of fostering investment in an interaction where two agents 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.

Under complete information, each agent has a dominant strategy to invest when $\theta = 2$ and not to invest when $\theta = -1$. Under incomplete information, however, a coordination problem arises. An agent with a "moderate" belief that $\theta = 2$ will invest if and only if he believes that the other agent is likely enough to invest as well. This gives rise to multiple equilibria.[13]

Consider an information-design problem with the following features.

1. The designer wants to stimulate investment and values the complementarities between agents. This is modeled by a symmetric and monotone payoff function: $v(\text{I}, \text{I}) > v(\text{I}, \text{N}) = v(\text{N}, \text{I}) > v(\text{N}, \text{N}) = 0$, such that $v(\text{I}, \text{I}) \geq (3/2)v(\text{I}, \text{N})$. The latter condition includes all supermodular designers.[14]
2. The solution concept is BNE.
3. The "min" selection rule, defined in (1), chooses the worst BNE outcome.

Adversarial equilibrium selection, defined by the min in feature 3, corresponds to a form of robust information design: the designer aims to maximize the payoff she would obtain if the worst equilibrium for her were played. When there are multiple equilibria, it is difficult to predict which one the agents will coordinate on. Therefore, in environments

[13] 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. 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) sensitive to the introduction of incomplete information.

[14] Supermodular (submodular) designers are those for which $v(\text{I}, \text{I}) + v(\text{N}, \text{N}) \geq (\leq)\ v(\text{N}, \text{I}) + v(\text{I}, \text{N})$.

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

where erroneous coordination can be particularly detrimental, it is especially important that information be designed to secure the highest possible payoff.

The current method available for Bayes-Nash information design, based on BCE (see Bergemann and Morris 2016 and Taneva 2019), cannot be used to solve this problem, because it does not apply to adversarial selection. To see why, consider any $\mu_0 \in [1/2, 2/3)$. The solution to the BCE program prescribes to provide no information to the agents, namely, $\pi^*(\mathrm{I}, \mathrm{I}|\theta) = 1$ for all θ. This is because the BCE method implicitly invokes the max selection rule (see Bergemann and Morris 2019 for further discussion). Since (I, I) and (N, N) are both BNE in the absence of any additional information, 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, under adversarial selection, the designer can achieve a higher expected payoff for $\mu_0 \in [1/2, 2/3)$ by revealing some information privately to one of the agents while leaving the other agent uninformed.

A. *Worst-Equilibrium Characterization*

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

Using the concepts from section III, 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 agent 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 agent i assigns high enough probability either to $\theta = 2$ or to agent j playing I (because $t_j \in \rho_j^1$), 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 an agent with belief in ρ_i is I. This implies that, in all BNEs, agent i's equilibrium strategy must choose I with certainty when his hierarchy is in ρ_i.

[15] Formally, uniquely rationalizable.

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

Given a belief distribution $\tau \in \mathcal{C}$, the worst equilibrium for our designer is such that all agents play I only when their beliefs belong to ρ_i (that is, only when I is uniquely rationalizable) and play N otherwise: for all i and t_i,

$$\sigma_i^{\text{MIN}}(\text{I}|t_i) = \begin{cases} 1 & \text{if } t_i \in \rho_i, \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

B. Solution

Following theorem 1 and corollary 1, we solve the information-design problem in two steps: maximization within and maximization between.

1. Maximization Within

We first characterize the state distributions for which the maximization within induces joint investment (i.e., investment by both agents) with certainty in the worst equilibrium. For $\mu > 2/3$, each agent invests regardless of the other's decision. Therefore, it is optimal to provide no information. For $\mu < 2/3$, instead, information can be used, at least in some cases, to induce joint investment with certainty by exploiting the equilibrium structure. In these cases, the designer induces some agent (say agent 1) to invest at a type t_1 only on the basis of his first-order belief being greater than $2/3$, that is, $t_1 \in \rho_1^1$. Leveraging the presence of t_1, the designer can induce a type t_2 of agent 2 to invest, with a first-order belief lower than $2/3$. This type finds investing uniquely optimal because of the sufficiently high probability it assigns to t_1, that is, $t_2 \in \rho_2^2$. This allows for a type t_1' of agent 1 to invest, with even lower first-order beliefs, on the basis of the probability it assigns to t_2, that is, $t_1' \in \rho_1^3$. This chain of reasoning continues for as many messages as the designer sends, each message corresponding to a different hierarchy.

We combine this insight with the conditions of proposition 2 to obtain a system of inequalities and one equation ([B1]–[B4] in sec. B2), which characterizes the minimal distributions that induce joint investment with certainty at μ. Simple addition of the inequalities shows that μ must be greater than $1/2$, as stated in the following result.

Claim 1. Joint investment with certainty can be achieved if and only if $\mu > 1/2$.

In the maximization within, revealing no information for $\mu \in (1/2, 2/3)$ yields no investment in the worst equilibrium. Therefore, the designer has to reveal some (private) information to induce joint investment with certainty for this range of μ. In fact, for all $\mu \in (1/2, 2/3)$ this can be achieved with the simple minimal distributions presented in table 3, where ε is a

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

TABLE 3
OPTIMAL MINIMAL DISTRIBUTIONS FOR $\mu \in (1/2, 2/3)$

$\hat{\sigma}^*_\mu$	t_2: $\mu_2 = \mu$
t_1: $\mu_1 = 2/3 + \varepsilon$	$3(\mu - \varepsilon) - 1$
t_1' : $\mu_1' = 1/3 + \varepsilon$	$2 - 3(\mu - \varepsilon)$

small positive number. From claim 1 it follows that the designer can no longer generate joint investment with certainty for $\mu \leq 1/2$. But can she still generate enough investment that her value from the maximization within is relevant for the concavification stage? As we show in the next section, for many monotone designers the answer is no; hence, maximization within for $\mu \leq 1/2$ can be dispensed with.

2. Maximization Between

Assuming that the designer does not maximize within for $\mu < 1/2$, we obtain a value function $\tilde{w}^*(\mu)$ equal to $v(\mathrm{I}, \mathrm{I})$ if $\mu > 1/2$ and zero otherwise. This function is plotted (dashed lines) in figure 1 for a designer with $v(\mathrm{I}, \mathrm{I}) = 2$ and $v(\mathrm{N}, \mathrm{I}) = v(\mathrm{I}, \mathrm{N}) = 1$. The concave envelope of $\tilde{w}^*$,

$$(\mathrm{cav}\tilde{w}^*)(\mu) = \begin{cases} v(\mathrm{I}, \mathrm{I}) & \text{if } \mu > 1/2, \\ 2\mu v(\mathrm{I}, \mathrm{I}) & \text{if } \mu \leq 1/2, \end{cases}$$

is also depicted (solid line) in figure 1. Our next result states that this concave envelope represents the optimal solution. We prove that the value of maximization within for $\mu \leq 1/2$ cannot be relevant for the concavification,

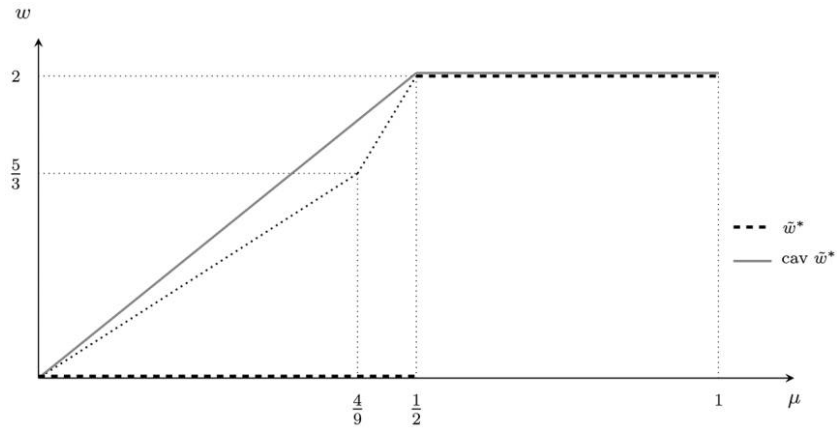

FIG. 1.—Optimal value. A color version of this figure is available online.

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

and therefore its computation within this range can be omitted without loss of generality.

CLAIM 2. $(\text{cav}\tilde{w}^*)(\mu_0)$ is the designer's optimal expected value at μ_0.

The optimal belief-hierarchy distribution is as follows: (i) for $\mu_0 > 2/3$, reveal no information; (ii) for $\mu_0 \in (1/2, 2/3]$, use e^*_μ with probability 1; and (iii) for $\mu_0 \leq 1/2$, use a public signal that either draws $e^*_{1/2+\varepsilon}$ with probability $2\mu_0/(1 + 2\varepsilon)$ or makes it common knowledge that $\theta = -1$ with the remaining probability.[16] The corresponding optimal information structure for $\mu_0 \leq 1/2$ is presented in table 4.

C. Extension

When the designer values unilateral investment enough that claim 2 no longer holds, maximization within may become relevant for $\mu \leq 1/2$, because its value may dominate $(\text{cav}\tilde{w}^*)(\mu)$. For such designers and state distributions, the optimal minimal distribution will necessarily put mass on at least one hierarchy at which N is also rationalizable and will thus be played in the worst equilibrium. For these hierarchies, lemma 5 in section B1 establishes that it is optimal to set the first-order beliefs to zero, because, for a fixed μ, this allows the designer to put more weight on other hierarchies that do invest. Therefore, an optimal minimal distribution can be found by solving our system (eqq. [B1]–[B4]), with as few first-order beliefs equal to zero as possible to find a solution at each μ.[17]

D. Discussion

Private information plays a central role in the optimal design of robust incentives under adversarial selection. In our application, private information fosters robust investment by making agents uncertain about each other's beliefs about the state while, at the same time, making them certain about each other's behavior. By giving some information to one of the agents while leaving the other one uninformed, the designer makes the latter uncertain about the beliefs of the former in a way that induces joint investment with certainty. It is sufficient for the uninformed agent to believe with high enough probability that investing is dominant for

[16] Under adversarial selection, an optimal solution may not exist for all μ_0 (for instance, at $\mu_0 = 1/2$ in our example). However, the supremum of theorem 1 can always be approached with arbitrary precision for all μ_0. In our example, $(\text{cav}\tilde{w}^*)(\mu_0)$ gives the exact value of the supremum for all μ_0. It is obtained by concavifying the function $\tilde{w}^*_k$, equal to 0 for all $\mu < (1/2) + \varepsilon$ and equal to 2 otherwise, and taking the pointwise limit as $\varepsilon \downarrow 0$. The function $\tilde{w}^*_k$ corresponds to using e^*_μ from table 3 (which depends on ε) for $\mu \geq (1/2) + \varepsilon$ and giving no information for lower μ.

[17] We do not need to compute the maximization within in our example for $\mu < 1/2$, because of claim 2. However, we used our system to compute its value for $R = L = 2$ (i.e., two messages per player) and have plotted the result (the dotted line) in fig. 1 for completeness.

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

TABLE 4
OPTIMAL INFORMATION STRUCTURE FOR $\mu_0 \leq 1/2$

	s	q
$\pi^*(\cdot \mid \theta = -1)$:		
s, s_1'	$\frac{\mu_0}{1-\mu_0}\frac{1/3-\varepsilon}{1+2\varepsilon}$	0
s, s_1''	$\frac{\mu_0}{1-\mu_0}\frac{2/3-\varepsilon}{1+2\varepsilon}$	0
q	0	$\frac{1-2\mu_0+2\varepsilon}{(1-\mu_0)(1+2\varepsilon)}$
$\pi^*(\cdot \mid \theta = 2)$:		
s, s_1'	$\frac{2/3+\varepsilon}{1+2\varepsilon}$	0
s, s_1''	$\frac{1/3+\varepsilon}{1+2\varepsilon}$	0
q	0	0

the other agent, in order to make investment uniquely optimal for him as well. In contrast, it is public information that plays the central role under favorable equilibrium selection. Note that public information can also provide robust incentives to invest under adversarial selection, but only if it makes both agents extremely optimistic (with first-order beliefs greater than 2/3). This use of information, however, is suboptimal, because joint investment can be achieved through private information, even when agents are not as optimistic. How this is possible is explained next.

Optimal information disclosure under adversarial selection brings out two important principles underpinning the role of private information: the bandwagon effect and extreme pessimism. The bandwagon effect is the force that provides robust incentives to invest without relying on agents' (mutual) extreme optimism about the state. There need be only one type of one agent (agent 1 of type t_1) with first-order beliefs about the state high enough to make investment strictly dominant at that type. Then, investing becomes uniquely optimal for the other agent (agent 2 of type t_2) under a more pessimistic belief about the state, if he puts sufficiently high probability on agent 1 being extremely optimistic (i.e., on agent 1 being of type t_1). In turn, investing becomes uniquely optimal for the second type of agent 1 (type t_1'), with an even lower first-order belief, because he is certain that player 2 will always invest. This contagion process trades off pessimism about the state for optimism about the other agent's investment by taking advantage of the two kinds of payoff complementarities. This trade-off can be exploited only by informative private signals.

Just as investment should be induced efficiently, so should be noninvestment. In our application, the designer causes an extreme form of pessimism, according to which an agent is certain that the state is bad whenever he chooses not to invest. If, instead, the designer were to induce a moderate belief about the state, at which both actions could be optimal, depending on coordination, then not investing would be played in the worst equilibrium. Therefore, the designer should make him as

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

pessimistic as possible, since this does not change his behavior in the worst equilibrium while allowing the designer to increase the probability of robust investment. This step also takes advantage of the complementarities: extremely pessimistic beliefs incentivize noninvestment because the complementarities between an agent's action and the state overpower the strategic complementarities. This allows the designer to put more probability on higher beliefs at which the bandwagon effect induces robust investment. Extreme pessimism also emerges in the solution to the single-agent problem of Kamenica and Gentzkow (2011), in which the agent is certain about the state when taking the undesirable (to the designer) action. As in our application, this in turn allows the designer to induce the desirable action with maximal probability, subject to Bayes plausibility.

VI. 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 maximizations 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, and 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

A1. Proof of Proposition 1

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

$$\tau(t) = \sum_{\theta} \pi(\{s : h(s) = t\}|\theta)\mu_0(\theta) \tag{A1}$$

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

$$p(\theta, t) = \pi(\{s : h(s) = t\}|\theta)\mu_0(\theta) \tag{A2}$$

for all θ and $t \in \operatorname{supp} \tau$. It is immediate from equations (A1) and (A2) that $\operatorname{marg}_T p = \tau$ and so $\operatorname{marg}_{T_i} p = \tau_i$ for all i. Further, when any agent i forms his

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

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,

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

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

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

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 (S, π). Define information structure $(\text{supp}\,\tau, \pi_\tau)$, where

$$\pi_\tau(t|\cdot) : \theta \mapsto \frac{1}{\mu_0(\theta)} \beta_i^*(\theta, t_{-i}|t_i)\tau_i(t_i) \tag{A3}$$

for all $t \in \text{supp}\,\tau$, 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

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

which guarantees that

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

(i.e., $\pi(\cdot|\theta)$ is a probability distribution for every θ). By construction, this information structure is such that, when any agent 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

$$\tau(t) = \sum_\theta \pi_\tau(t|\theta)\mu_0(\theta) \tag{A4}$$

for all $t \in \text{supp}\,\tau$. By (A3), the right-hand side of equation (A4) 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 $\text{marg}_\Theta p = \tau$). QED

A2. Proof of Proposition 2

Suppose τ is consistent and Bayes plausible. By definition, there exists a common prior $p \in \Delta(\Theta \times T)$ such that, for all t, θ, and i,

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

and $\text{marg}_T p = \tau$. Moreover, by Bayes plausibility, $\text{marg}_\Theta p = \mu_0$. Define $\nu(\theta|t) \coloneqq p(\theta|t)$ for all t and θ. We show that conditions (7)–(9) are satisfied. First,

$$\begin{aligned} \textstyle\sum_t \tau(t)\nu(\theta|t) &= \textstyle\sum_t p(t)p(\theta|t) \\ &= \textstyle\sum_t p(t)\dfrac{p(\theta, t)}{p(t)} \\ &= \mu_0(\theta). \end{aligned}$$

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

Moreover, since p is a common prior, $p(\theta, t_{-i}|t_i) = \beta_i^*(\theta, t_{-i}|t_i)$ and $p(t_{-i}|t_i) = \beta_i^*(t_{-i}|t_i)$ for all i, t, and θ. Thus, for all i, t, and θ,

$$\nu(\theta|t_i, t_{-i}) \coloneqq p(\theta|t_i, t_{-i}) = \frac{p(\theta, t_{-i}|t_i)}{p(t_{-i}|t_i)} = \beta_i^*(\theta|t_i, t_{-i}).$$

Finally, since $\text{marg}_T p = \tau$ and p is a common prior,

$$\tau(t_{-i}|t_i) = p(t_{-i}|t_i) = \beta_i^*(t_{-i}|t_i)$$

for all i and t.

Now, consider $\tau \in \Delta^f(T)$ and $\nu : \text{supp}\,\tau \to \Delta\Theta$ such that conditions (7)–(9) hold. Define $p \in \Delta(\Theta \times T)$ as $p(\theta, t) \coloneqq \tau(t)\nu(\theta|t)$ for all θ and t. We want to show that p is a common prior for τ. First, note that, by definition of p, $p(t) = \Sigma_\theta p(\theta, t) = \tau(t)\Sigma_\theta \nu(\theta|t) = \tau(t)$, for all t. Therefore, $\text{marg}_T p = \tau$. Second, for all i, t, and θ,

$$\begin{aligned} p(\theta, t_{-i}|t_i) &= \frac{p(\theta, t)}{p(t_i)} \\ &= \frac{\tau(t)\nu(\theta|t)}{\tau(t_i)} \\ &= \tau(t_{-i}|t_i)\nu(\theta|t) \\ &= \beta_i^*(t_{-i}|t_i)\beta_i^*(\theta|t) = \beta_i^*(\theta, t_{-i}|t_i). \end{aligned}$$

The first equality comes from Bayes's rule and the second equality comes from the definition of p. In the fourth equality, we used conditions (8) and (9). Finally, in the last equality, we used the definition of $\beta_i^*(\theta|t_i, t_{-i})$. This implies that $p(\theta, t) = \beta_i^*(\theta, t_{-i}|t_i)p(t_i)$ for all θ, t, and i. Therefore, p is a common prior for τ or, equivalently, $\tau \in \mathcal{C}$. Finally, τ is Bayes plausible because, given condition (7),

$$p(\theta) = \sum_t p(\theta, t) = \sum_t \tau(t)\nu(\theta|t) = \mu_0(\theta),$$

for all θ. QED

A3. 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 $\text{marg}_T p_{\tau'} = \tau'$ and $\text{marg}_T p_{\tau''} = \tau''$ and

$$\begin{aligned} p_{\tau'}(\theta, t) &= \beta_i^*(\theta, t_{-i}|t_i)\tau_i'(t_i), \\ p_{\tau''}(\theta, t) &= \beta_i^*(\theta, t_{-i}|t_i)\tau_i''(t_i), \end{aligned} \tag{A5}$$

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

$$p_\tau(\theta, t) \coloneqq \beta_i^*(\theta, t_{-i}|t_i)\tau_i(t_i)$$

for all i, θ, and $t \in \text{supp}\,\tau$. Note that p_τ is well defined, because of equation (A5). Thus,

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

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

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

Although $\mathcal{C}$ is convex, it is not closed because we can build sequences in $\mathcal{C}$ with growing supports, converging to a belief-hierarchy distribution only 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 $\mathcal{C}^M \subseteq \mathcal{E}$. Take $\tau \in \mathcal{C}^M$. If $\tau \notin \mathcal{E}$, then there exist $\tau', \tau'' \in \mathcal{C}$ with $\tau' \neq \tau''$ and $\alpha \in (0,1)$ such that $\tau := \alpha\tau' + (1-\alpha)\tau''$. Since $\tau \in \mathcal{C}^M$, $\tau \in \mathcal{C}$. Moreover, $\text{supp}\,\tau' \cup \text{supp}\,\tau'' \subseteq \text{supp}\,\tau$. 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, $\Sigma_{t\in\text{supp}\tau}\tau_\lambda(t) = 1$ for all $\lambda \in \mathbb{R}_+$. Define $\Lambda := \{\lambda \geq 0 : \forall\, t \in \text{supp}\,\tau, 0 \leq \tau_\lambda(t) \leq 1\}$ 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 nonempty, since both $\lambda = 0$ and $\lambda = (1-\alpha)/\alpha$ 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 that $\lambda \in \Lambda$. For all $t \in \text{supp}\tau$, by definition of τ, the sequence $\tau_{\lambda_m}(t)$ is monotone (either nondecreasing or nonincreasing), 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 $\tilde{\lambda} := \max \Lambda \geq (1-\alpha)/\alpha > 0$.

We want to show that $\text{supp}\,\tau_{\tilde{\lambda}} \subsetneq \text{supp}\,\tau$. To see this, let $\tilde{t} \in \text{supp}\,\tau_{\tilde{\lambda}}$, and suppose that $\tilde{t} \notin \text{supp}\,\tau$. Then, by definition, $\tau_{\tilde{\lambda}}(\tilde{t}) = -\tilde{\lambda}\tau''(\tilde{t}) \leq 0$, which is impossible. Moreover, there must exist $t \in \text{supp}\,\tau$ such that $\tau_{\tilde{\lambda}}(t) = 0$. To see this, suppose not; that is, suppose that $\text{supp}\tau = \text{supp}\tau_{\tilde{\lambda}}$. Then, for all $t \in \text{supp}\,\tau$, we would have that $\tau_{\tilde{\lambda}}(t) > 0$. Since $\tau_{\tilde{\lambda}} \in \mathcal{C}$ by construction, and since $|\text{supp}\,\tau| > 1$ (otherwise $\tau' = \tau''$), we also have that $\tau_{\tilde{\lambda}}(t) < 1$ for all $t \in \text{supp}\,\tau$. Let $T^- := \{t \in \text{supp}\,\tau : \tau(t) - \tau''(t) < 0\}$ and $T^+ := \{t \in \text{supp}\,\tau : \tau(t) - \tau''(t) > 0\}$. These sets are nonempty by assumption ($\tau' \neq \tau''$). For $t \in T^-$, let $\lambda(t) := -\tau(t)/(\tau(t) - \tau''(t))$, and note that $0 = \tau(t) + \lambda(t)(\tau(t) - \tau''(t)) < \tau_{\tilde{\lambda}}(t)$, implying that $\lambda(t) > \tilde{\lambda}$. Similarly, for $t \in T^+$, let $\lambda(t) := (1 - \tau(t))/(\tau(t) - \tau''(t))$, and note that $1 = \tau(t) + \lambda(t)(\tau(t) - \tau''(t)) > \tau_{\tilde{\lambda}}(t)$, implying that $\lambda(t) > \tilde{\lambda}$. 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 maximum. Therefore, we conclude that $\text{supp}\,\tau_{\tilde{\lambda}} \subsetneq \text{supp}\,\tau$ and thus that $\tau \notin \mathcal{C}^M$.

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

$$\tau = \alpha\tau' + (1-\alpha)\tau'', \tag{A6}$$

where $\alpha = \tau(\text{supp}\tilde{\tau}) \in (0,1)$. Since $\text{supp}\tilde{\tau}$ is belief closed, so is $\text{supp}\,\tau \setminus \text{supp}\tilde{\tau}$. To see why, note that for any $t \in \text{supp}\,\tau \setminus \text{supp}\tilde{\tau}$, if there were $i, \tilde{t} \in \text{supp}\tilde{\tau}$, and $\theta \in \Theta$ such that $p_\tau(\theta, \tilde{t}_{-i}|t_i) > 0$, then this would imply $p_\tau(\theta, t_i, \tilde{t}_{-i}) > 0$ and, thus, $p_\tau(\theta, t_i, \tilde{t}_{-(ij)}|\tilde{t}_j) > 0$ (where $\tilde{t}_{-(ij)}$ are the hierarchy profiles for agents other than i

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

and j). This implies that at $\tilde{t}_j$—a hierarchy that agent j can have in $\tilde{\tau}$—agent j assigns strictly positive probability to a hierarchy of agent 1 that is not in $\text{supp}\tilde{\tau}$. This contradicts the fact that $\text{supp}\tilde{\tau}$ is belief closed. 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 $\tau'' \neq \tau'$, equation (A6) implies that τ is not an extreme point. QED

PROPOSITION 3. For any $\tau \in \mathcal{C}$, there exist unique $\{e_\ell\}_{\ell=1}^{L} \subseteq \mathcal{C}^{\mathrm{M}}$ and weakly positive numbers $\{\alpha_\ell\}_{\ell=1}^{L}$ such that $\Sigma_{\ell=1}^{L}\alpha_\ell = 1$ and $\tau = \Sigma_{\ell=1}^{L}\alpha_\ell e_\ell$.

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 $\text{supp}\tau' \subsetneq \text{supp}\,\tau$. Similarly, either τ' is minimal, in which case we conclude that there exists a minimal $e^1 := \tau'$ such that $\text{supp}e^1 \subsetneq \text{supp}\,\tau$, or there is $\tau'' \in \mathcal{C}$ such that $\text{supp}\tau'' \subsetneq \text{supp}\,\tau'$. Given that τ has finite support, finitely many steps of this procedure deliver a minimal consistent belief-hierarchy distribution e^1, $\text{supp}e^1 \subsetneq \text{supp}\,\tau$. Since τ and e^1 are both consistent, and hence their supports are belief closed, $\text{supp}\tau \setminus \text{supp}e^1$ must be belief closed. Given that $\text{supp}\,\tau \setminus \text{supp}e^1$ is a belief-closed subset of $\text{supp}\,\tau$ and τ is consistent, define a new distribution τ as

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

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

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

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

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

for all $\theta \in \Theta$ and $t \in \text{supp}\tau^2$. Hence, $\tau^2 \in \mathcal{C}$. Therefore, we can repeat the procedure for distribution $\tau^2 \in \mathcal{C}$. Since τ has finite support, there exists $L \in N$ such that, after L steps of this procedure, we obtain a consistent 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 $t \in \text{supp}\,\tau$ there exists a unique $\ell \in \{1, \ldots, L\}$ such that $\tau = e^\ell\tau(\text{supp}e^\ell)$. Therefore, $\tau = \Sigma_{\ell=1}^{L}\alpha_\ell e^\ell$, where $\alpha_\ell := \tau(\text{supp}e^\ell) > 0$ and $\Sigma_{\ell=1}^{L}\alpha_\ell = \Sigma_{\ell=1}^{L}\tau(\text{supp}e^\ell) = 1$.

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

i. $\text{supp}\,e^\ell \neq \text{supp}\,\hat{e}^k$ Since $e^\ell, \hat{e}^k \in \mathcal{C}$, $\text{supp}e^\ell$ and $\text{supp}\,\hat{e}^k$ are belief closed, which in turn implies that $T^{\ell,k} := \text{supp}e^\ell \cap \text{supp}\hat{e}^k$ (nonempty by assumption)

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

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 $\text{supp}e^{\ell,k} \subsetneq \text{supp}e^{\ell}$. This contradicts the minimality of e^{ℓ}.

ii. $\text{supp}\, e^{\ell} = \text{supp}\, \hat{e}^{k}$ Since $e^{\ell}, \hat{e}^{k} \in \mathcal{C}$, there exist common priors p and $\hat{p}$ in $\Delta(\Theta \times T)$ such that $\text{marg}_T p = e_\ell$ and $\text{marg}_T \hat{p} = \hat{e}^{k}$. Thus, $\text{suppmarg}_T p = \text{suppmarg}_T \hat{p}$. This implies that $\text{supp}p = \text{supp}\hat{p}$ (if not, without loss there would exist some i, $\tilde{t} \in \text{suppmarg}_T p$, and $\tilde{\theta} \in \Theta$, such that $\beta_i(\tilde{\theta}, \tilde{t}_{-i}|\tilde{t}_i) > 0$, while for all i and $t \in \text{suppmarg}_T \hat{p}$, $\beta_i(\tilde{\theta}, t_{-i}|t_i) = 0$. This would contradict $\text{suppmarg}_T p = \text{suppmarg}_T \hat{p}$). 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 $e^{\ell} \neq \hat{e}^{k}$. QED

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.

For any $\tau', \tau'' \in \mathcal{C}$, let

$$\alpha O_{\Lambda'}(\tau') + (1-\alpha) O_{\Lambda''}(\tau'') := \{\alpha\gamma' + (1-\alpha)\gamma'' : \gamma' \in O_{\Lambda'}(\tau') \text{ and } \gamma'' \in O_{\Lambda''}(\tau'')\}. \tag{A7}$$

PROPOSITION 4. If Σ^{B} is invariant, then for all $\tau', \tau'' \in \mathcal{C}$ and $\tau = \alpha\tau' + (1-\alpha)\tau''$ with $\alpha \in [0, 1]$ and for all $\Lambda \in \Sigma^{B}(\tau)$, there are $\Lambda' \in \Sigma^{B}(\tau')$ and $\Lambda'' \in \Sigma^{B}(\tau'')$ such that

$$O_\Lambda(\tau) = \alpha O_{\Lambda'}(\tau') + (1-\alpha) O_{\Lambda''}(\tau''). \tag{A8}$$

Proof. Take any $\tau', \tau'' \in \mathcal{C}$ and $\alpha \in [0, 1]$, and let $\tau = \alpha\tau' + (1-\alpha)\tau''$. Take any $\Lambda \in \Sigma^{B}(\tau)$ and $\sigma \in \Lambda$. Define

$$\gamma_\sigma(a, \theta) := \sum_{t \in \text{supp}\tau} \sigma(a|t) p_\tau(t, \theta) \quad \forall\ (a, \theta), \tag{A9}$$

so that $O_\Lambda(\tau) = \{\gamma_\sigma : \sigma \in \Lambda\}$. It follows from invariance that $\Lambda' := \{\sigma|_{\text{supp}\tau'} : \sigma \in \Lambda\}$ is in $\Sigma^{B}(\tau')$ and that $\Lambda'' := \{\sigma|_{\text{supp}\,\tau''} : \sigma \in \Lambda\}$ is in $\Sigma^{B}(\tau'')$. Moreover, note that

$$\begin{aligned}\left\{\sigma|_{\text{supp}\,\tau' \cap \text{supp}\tau''} : \sigma \in \Lambda\right\} &= \left\{\sigma'|_{\text{supp}\tau' \cap \text{supp}\tau''} : \sigma' \in \Lambda'\right\} \\ &= \left\{\sigma''|_{\text{supp}\tau' \cap \text{supp}\tau''} : \sigma'' \in \Lambda''\right\}.\end{aligned}$$

Now define

$$\gamma'_\sigma(a, \theta) := \sum_{t \in \text{supp}\,\tau'} \sigma|_{\text{supp}\tau'}(a|t) p_{\tau'}(\theta, t) \quad \forall\ (a, \theta),$$

$$\gamma''_\sigma(a, \theta) := \sum_{t \in \text{supp}\,\tau''} \sigma|_{\text{supp}\tau''}(a|t) p_{\tau''}(\theta, t) \quad \forall\ (a, \theta),$$

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

so that $O_{\Lambda'}(\tau') = \{\gamma'_\sigma : \sigma \in \Lambda\}$ and $O_{\Lambda''}(\tau'') = \{\gamma''_\sigma : \sigma \in \Lambda\}$. Since $p_\tau = \alpha p_{\tau'} + (1 - \alpha)p_{\tau''}$, (A9) implies that for all $\sigma \in \Lambda$, a, and θ,

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

Hence, $O_\Lambda(\tau) = \alpha O_{\Lambda'}(\tau') + (1 - \alpha)O_{\Lambda''}(\tau'')$. QED

LEMMA 3. The function w is linear over $\mathcal{C}$.

Proof. Take any $\tau', \tau'' \in \mathcal{C}$ and $\alpha \in [0, 1]$, and let $\tau = \alpha\tau' + (1 - \alpha)\tau''$. By proposition 4, we know that for all sequences (Λ_n) in $\Sigma^B(\tau)$, there exist sequences (Λ'_n) in $\Sigma^B(\tau')$ and (Λ''_n) in $\Sigma^B(\tau'')$ such that

$$O_{\Lambda_n}(\tau) = \alpha O_{\Lambda'_n}(\tau') + (1 - \alpha)O_{\Lambda''_n}(\tau'') \quad \forall \ n.$$

Since g is linear,

$$\begin{aligned}\sum_{\theta,a} g(O_{\Lambda_n}(\tau))(a, \theta)v(a, \theta) &= \sum_{\theta,a} g(\alpha O_{\Lambda'_n}(\tau') + (1 - \alpha)O_{\Lambda''_n}(\tau''))(a, \theta)v(a, \theta) \\ &= \alpha \sum_{\theta,a} g(O_{\Lambda'_n}(\tau'))v(a, \theta) + (1 - \alpha) \sum_{\theta,a} g(O_{\Lambda''_n}(\tau''))(\theta, a)v(a, \theta).\end{aligned} \tag{A10}$$

Choose $(\bar{\Lambda}_n)$ in $\Sigma^B(\tau)$ such that $w(\tau) = \lim_{x \to \infty} \Sigma_{\theta,a} g(O_{\bar{\Lambda}_n}(\tau))(a, \theta)v(a, \theta)$, $(\bar{\Lambda}'_n)$ in $\Sigma^B(\tau')$ such that $w(\tau') = \lim_{x \to \infty} \Sigma_{\theta,a} g(O_{\bar{\Lambda}'_n}(\tau'))(a, \theta)v(a, \theta)$, and $(\bar{\Lambda}''_n)$ in $\Sigma^B(\tau'')$ such that $w(\tau'') = \lim_{x \to \infty} \Sigma_{\theta,a} g(O_{\bar{\Lambda}''_n}(\tau''))(a, \theta)v(a, \theta)$. By equation (A10), it must be that

$$\begin{aligned}w(\tau) &= \lim_{n \to \infty} \sum_{\theta,a} g(O_{\bar{\Lambda}_n}(\tau))(a, \theta)v(a, \theta) \\ &\leq \alpha \lim_{n \to \infty} \sum_{\theta,a} g(O_{\bar{\Lambda}'_n}(\tau'))v(a, \theta) + (1 - \alpha) \lim_{n \to \infty} \sum_{\theta,a} g(O_{\bar{\Lambda}''_n}(\tau''))(a, \theta)v(a, \theta) \\ &= \alpha w(\tau') + (1 - \alpha)w(\tau'').\end{aligned}$$

Next, choose (Λ'_n) in $\Sigma^B(\tau')$ and (Λ''_n) in $\Sigma^B(\tau'')$ such that

$$\begin{aligned}\alpha w(\tau') + (1 - \alpha)w(\tau'') = \alpha &\lim_{n \to \infty} \sum_{\theta,a} g(O_{\Lambda'_n}(\tau'))v(a, \theta) \\ &+ (1 - \alpha) \lim_{n \to \infty} \sum_{\theta,a} g(O_{\Lambda''_n}(\tau''))(\theta, a)v(a, \theta)\end{aligned} \tag{A11}$$

and such that

$$\left\{\sigma'|_{\text{supp}\tau' \cap \text{supp}\tau''} : \sigma' \in \Lambda'_n\right\} = \left\{\sigma''|_{\text{supp}\tau' \cap \text{supp}\tau''} : \sigma'' \in \Lambda''_n\right\} =: \Lambda'''_n.$$

Note that the above restriction to sequences describing the same behavior over $\text{supp}\tau' \cap \text{supp}\tau''$ is without loss, because a maximizing designer must be

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

indifferent between their selected outcomes. Therefore, the right-hand side of equation (A11) can be written as

$$\begin{aligned}\alpha \lim_{n\to\infty} \sum_{\theta,a} g(O_{\tilde{\Lambda}'_n}(\tilde{\tau}'))v(a,\theta) + (1-\alpha)\lim_{n\to\infty} \sum_{\theta,a} g(O_{\tilde{\Lambda}''_n}(\tilde{\tau}''))(\theta,a)v(a,\theta) \\ + \lim_{n\to\infty} \sum_{\theta,a} g(O_{\Lambda'''_n}(\tau'''))(\theta,a)v(a,\theta)\end{aligned} \tag{A12}$$

for some sequences $(\tilde{\Lambda}'_n)$ in $\Sigma^{\mathrm{B}}(\tilde{\tau}')$, $(\tilde{\Lambda}''_n)$ in $\Sigma^{\mathrm{B}}(\tilde{\tau}'')$, and (Λ'''_n) in $\Sigma^{\mathrm{B}}(\tau''')$, where $\tilde{\tau}'$ is the consistent distribution with support $\mathrm{supp}\tau' \setminus \mathrm{supp}\tau''$, $\tilde{\tau}''$ is the consistent distribution with support $\mathrm{supp}\tau'' \setminus \mathrm{supp}\tau'$, and τ''' is the consistent distribution with support $\mathrm{supp}\tau'' \cap \mathrm{supp}\tau'$. By applying invariance (twice),

$$\tilde{\Lambda}_n := \left\{\sigma : \mathrm{supp}\,\tau \to \Delta A | \sigma|_{\mathrm{supp}\tilde{\tau}'} \in \tilde{\Lambda}'_n, \sigma|_{\mathrm{supp}\tilde{\tau}''} \in \tilde{\Lambda}''_n \text{ and } \sigma|_{\mathrm{supp}\tau'''} \in \Lambda'''_n\right\}$$

must be in $\Sigma^{\mathrm{B}}(\tau)$, which, together with linearity of g, ensures that equation (A12) is equal to

$$\lim_{n\to\infty} \sum_{\theta,a} g(O_{\tilde{\Lambda}_n}(\tau))v(a,\theta) = w(\tau).$$

Hence, we obtain $w(\tau) = \alpha w(\tau') + (1-\alpha)w(\tau'')$. QED

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 $V(S,\pi) \le w(\tau)$ and, thus, $\sup_{(S,\pi)} V(S,\pi) \le \sup\{w(\tau) | \tau \in \mathcal{C} \text{ and } \mathrm{marg}_\Theta p_\tau = \mu_0\}$. 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, $\sup_{(S,\pi)} V(S,\pi) \ge \sup\{w(\tau) | \tau \in \mathcal{C} \text{ and } \mathrm{marg}_\Theta p_\tau = \mu_0\}$. We conclude that

$$\sup_{(S,\pi)} V(S,\pi) = \sup_{\substack{\tau \in \mathcal{C} \\ \mathrm{marg}_\Theta p_\tau = \mu_0}} w(\tau). \tag{A13}$$

By proposition 3, there exists a unique $\lambda \in \Delta^{\mathrm{f}}(\mathcal{C}^{\mathrm{M}})$ such that $\tau = \Sigma_{e\in\mathrm{supp}\lambda}\lambda(e)e$. 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 equation (A13), we have

$$\begin{aligned}\sup_{(S,\pi)} V(S,\pi) = \sup_{\lambda\in\Delta^{\mathrm{f}}(\mathcal{C}^{\mathrm{M}})} \sum_e w(e)\lambda(e) \\ \text{subject to} \sum_e \mathrm{marg}_\Theta p_e\, \lambda(e) = \mu_0,\end{aligned} \tag{A14}$$

which concludes the proof. QED

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

Appendix B

Appendix of the Application

B1. Bandwagon Effect and Extreme Pessimism

We first show that, given the structure of the worst equilibrium, in any optimal distribution with m hierarchies in total, an agent invests at hierarchy t_i in the worst equilibrium if and only if t_i rationalizes I uniquely on the basis of beliefs of order m or lower.

Lemma 4 (Bandwagon effect). Suppose that τ^* is an optimal belief-hierarchy distribution. Let $m = \Sigma_{i=1,2}|\text{supp}\tau_i^*|$. Then, for all i and $t_i \in \text{supp}\tau_i^*$, $\sigma_i^{\text{MIN}}(\text{I}|t_i) = 1$ if and only if $t_i \in \cup_{k=1}^m \rho_i^k$.

Proof. The "if" part follows from equation (14) 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 that $\hat{t}_i \in \text{supp}\tau_i^*$ and $\sigma_i^{\text{MIN}}(\text{I}|\hat{t}_i) = 1$, but $\hat{t}_i \notin \cup_{k=1}^m \rho_i^k$. Then, by equation (14) it must be that $\hat{t}_i \in \cup_{k\geq m+1}\rho_i^k$. Note 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 $k^*(t_i) \in N^+$ such that $t_i \in \rho_i^{k^*(t_i)}$ and $t_i \notin \rho_i^{k^*(t_i)-1}$, where $\rho_i^0 = \varnothing$. 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 that $\beta_i^*(\Theta \times \rho_j^{m+n-1}|\hat{t}_i) > \beta_i^*(\Theta \times \rho_j^{m+n-2}|\hat{t}_i)$, 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; that is, there must 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 must be at least $m + n$ different hierarchies, which contradicts $m = \Sigma_{i=1,2}|\text{supp}\tau_i^*|$. QED

An implication of lemma 4 is that there is an optimal τ^* such that for every $k = 2, \ldots, m$ it holds that either $\text{supp}\tau_i^* \cap \rho_i^k = \varnothing$ for $i = 1, 2$ or $\text{supp}\tau_i^* \cap \rho_i^k = t_i^k$, $\text{supp}\tau_i^* \cap \rho_i^{k-1} = \varnothing$, and $\text{supp}\tau_j^* \cap \rho_j^{k-1} = t_j^{k-1}$ for $j \neq i$, $i = 1, 2$. Next, we show that it is never optimal to induce a belief hierarchy at which an agent has multiple rationalizable actions. In particular, the hierarchies that rationalize action N are optimally set to have first-order beliefs of zero, that is, extreme pessimism.

Lemma 5 (Extreme pessimism). If τ^* is an optimal belief-hierarchy distribution and if there is a BNE σ such that $\sigma_i(N|t_i) > 0$ for some i and $t_i \in \text{supp}\tau_i^*$, then $\beta_i^*(\theta = 2|t_i) = 0$.

Proof. 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 $t_i \in \text{supp}\tau_i$ such that $t_i \in \rho_i$. If not, τ would be strictly dominated by $\hat{\tau} = x \cdot t_{(2/3)+\varepsilon} + (1 - x) \cdot t_0$ with $x = \min\{\mu/[(2/3) + \varepsilon], 1\}$, 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

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

$x \cdot v(\mathrm{I}, \mathrm{I}) + (1 - x) \cdot v(\mathrm{N}, \mathrm{N})$. Instead, if for some i, $t_i \notin \rho_i$ for all $t_i \in \mathrm{supp}\tau_i$, the designer's expected payoff is bounded above by $x \cdot v(\mathrm{I}, \mathrm{N}) + (1 - x) \cdot v(\mathrm{N}, \mathrm{N})$, which is strictly smaller than the payoff under $\hat{\tau}$.

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

By lemma 4, we know that any investing t_i is in $\cup_{k=1}^{m} \rho_i^k$ and there is a unique $t_i^k \in (\rho_1^k \cup \rho_2^k) \cap \mathrm{supp}\tau_i$. Consider the distribution p' obtained from p_τ by transferring all the probability mass $\Sigma_{t_j} p_\tau(\tilde{t}_i, t_j, \theta = 2)$, as described below. This operation changes the hierarchies in $\mathrm{supp}\, p_\tau$ according to $f_i : \mathrm{suppmarg}_{T_i} p_\tau \to \mathrm{suppmarg}_{T_i} p'$ for all i.

A.

i. If $\mathrm{supp}\tau_i \cap \rho_i^1 \neq \varnothing$, then $p'(f_i(t_i^1), f_j(t_j), \theta = 2) = p_\tau(t_i^1, t_j, \theta = 2) + p_\tau(\tilde{t}_i, t_j, \theta = 2)$ for all t_j.

ii. If $\mathrm{supp}\tau_i \cap \rho_i^1 = \varnothing$ and m is even, then $p'(f_i(t_i^{2k}), f_j(t_j^{2k-1}), \theta = 2) = p_\tau(t_i^{2k}, t_j^{2k-1}, \theta = 2) + p_\tau(\tilde{t}_i, t_j^{2k-1}, \theta = 2)$ for all $k = 1, \ldots, m/2$.

iii. If $\mathrm{supp}\tau_i \cap \rho_i^1 = \varnothing$ and m is odd, then $p'(f_i(t_i^{2k}), f_j(t_j^{2k-1}), \theta = 2) = p_\tau(t_i^{2k}, t_j^{2k-1}, \theta = 2) + p_\tau(\tilde{t}_i, t_j^{2k-1}, \theta = 2)$ for all $k = 1, \ldots, (m + 1)/2$, where t_i^{m+1} is a new hierarchy with $p'(t_i^{m+1}, f_j(t_j), \theta = -1) = 0$ for all t_j.

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

C. $p'(f_i(t_i), f_j(t_j), \theta) = p_\tau(t_i, t_j, \theta)$ otherwise.

Now, consider the belief-hierarchy distribution τ' induced by p'. By construction of p', conditions A–C ensure that the beliefs of all (relevant) orders have weakly increased for both agents in all hierarchies, except for $\tilde{t}_i$. Therefore, if agent 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 agent i with hierarchy $f_i(t_i)$ under τ'. On the other hand, agent i with hierarchy $\tilde{t}_i$ played N in the worst equilibrium under τ, and this continues to hold for agent i with hierarchy $f_i(\tilde{t}_i)$ under τ'. Moreover, τ' necessarily Bayes plausible (since it is induced by p'). Finally, again by construction of p', probability mass has been transferred only from noninvesting to investing types; hence, a designer with symmetric and monotone payoffs must have a strictly higher expected payoff under τ' than under τ, a contradiction. QED

B2. Characterization of the Optimal Minimal Distributions

The above lemmas imply that optimal design takes a special form in this example. First, if it is necessary to induce N, then by lemma 5 the designer should do it by making the agent extremely pessimistic about the state. In this way, for a given μ, the designer can put higher probability on hierarchies that induce I. Second, lemma 4 implies that optimal design incentivizes joint investment through contagion, which we call the "bandwagon effect." To uniquely rationalize joint investment, on the basis of agents' beliefs of bounded order, some hierarchy of an agent must initiate investment by having first-order beliefs greater than 2/3. For any given μ, it is never optimal to put mass on more than one such "first-order investor" hierarchy, because we can induce the other agent to invest at lower (than 2/3) first-order beliefs, if he puts a sufficiently high probability on that first-order

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

investor. Likewise, it is optimal to have only one such "second-order investor," as we may induce the agent with the first-order investor hierarchy to now invest at a different hierarchy with even lower first-order beliefs, if he puts a sufficiently high probability on that second-order investor hierarchy of his opponent. By proceeding in this way, we generate the lowest possible first-order beliefs overall and thus generate joint investment with certainty at the lowest possible state distribution μ. When joint investment can no longer be sustained, and N has to be played with positive probability, we set the first-order beliefs of the highest-order investor hierarchies to zero, by virtue of lemma 5.

For an arbitrary $m < \infty$, suppose the designer sends m messages in total in the maximization within. Let us index each hierarchy of agent 1 by $l = 1, \ldots, L$ and each of agent 2 by $r = 1, \ldots, R$. By lemmas 4 and 5, the optimal minimal distributions will have either $L = R = m/2$ if m is even, or $L = (m + 1)/2$ and $R = (m - 1)/2$ if m is odd.

Given a commonly known $\mu := \text{Prob}(\theta = 2)$, denote an optimal minimal distribution e^*_μ by

e^*_μ	$t_{2,1}$	$\cdots$	$t_{2,r}$	$\cdots$	$t_{2,R}$
$t_{1,1}$	A_{11}	$\cdots$	A_{1r}	$\cdots$	A_{1R}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$t_{1,l}$	A_{l1}	$\cdots$	A_{lr}	$\cdots$	A_{lR}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$t_{1,L}$	A_{L1}	$\cdots$	A_{Lr}	$\cdots$	A_{LR}

where all entries A_{lr} are positive and $\Sigma_{l,r} A_{lr} = 1$. Define $A^1_l := \Sigma_r A_{lr}$ and $A^2_r := \Sigma_l A_{lr}$. Let $\mu^l_1 := \beta^*_1(\theta = 2|t_{1,l})$ and $\mu^r_2 := \beta^*_2(\theta = 2|t_{2,r})$ denote the agents' first-order beliefs at the respective hierarchies.

For $\mu > 2/3$, the optimal minimal distribution is given by table B1, where $\mu^1_i = \mu$ for $i = 1, 2$. The designer simply lets agents act under common knowledge of μ, as investment is uniquely optimal in this range.

For $\mu \leq 2/3$, the following system yields the optimal minimal distributions e^*_μ such that all agents invest at all t_i:

$$\begin{cases} \dfrac{\sum_r A_{1r}\mu^{1r}}{A^1_1} > \dfrac{2}{3}, & \text{(B1)} \\ \dfrac{\sum_l A_{lr}\mu^{lr}}{A^2_r} + \dfrac{1}{3}\dfrac{\sum_{l=1}^{r} A_{lr}}{A^2_r} > \dfrac{2}{3} \quad \text{for } r = 1, \ldots, R, & \text{(B2)} \\ \dfrac{\sum_r A_{lr}\mu^{lr}}{A^1_l} + \dfrac{1}{3}\dfrac{\sum_{r=1}^{l-1} A_{lr}}{A^1_l} > \dfrac{2}{3} \quad \text{for } l = 2, \ldots, L, & \text{(B3)} \\ \sum_{l,r} A_{lr}\mu^{lr} = \mu. & \text{(B4)} \end{cases}$$

where $\mu^{lr} := \beta^*(\theta = 2|t_{1,l}, t_{2,r})$ denotes the designer's belief that $\theta = 2$, given the profile of agents' hierarchies $(t_{1,l}, t_{2,r})$. We have used proposition 2 and lemma 4 to construct this system (eqq. [B1]–[B4]). Indeed, we have incorporated

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

TABLE B1
OPTIMAL MINIMAL DISTRIBUTIONS FOR $\mu \geq 2/3$

e^*_μ	$t_{2,1}$
$t_{1,1}$	1

conditions (7)–(9) from proposition 2: all agents' first-order beliefs should be a $\tau(t_{-i}|t_i)$-weighted average of the designer's beliefs. For example, in condition (B1),

$$\mu_1^1 := \frac{\sum_r A_{1r}\mu^{1r}}{A_1^1}$$

represents agent 1's (first-order) belief that $\theta = 2$ at $t_{1,1}$. Likewise, the first expressions on the left-hand side of conditions (B2) and (B3) represent first-order beliefs μ_2^r and μ_1^l, respectively. Condition (B4) captures condition (7) from proposition 2.

Inequalities (B1)–(B3) are the investment constraints inherited from the equilibrium characterization in equation (14) and lemma 4. As discussed above, some agent must have a hierarchy at which he is the first-order investor. Without loss, choose $t_{1,1}$ to play that role and require $\mu_1^1 > 2/3$ (this is condition [B1]). Choose $t_{2,1}$ to be the second-order investor, who invests on the basis of second-order beliefs, that is, $t_{2,1} \in \rho_2^2$. This is equivalent to $\mu_2^1 + (1/3)\beta_2^*(t_{1,1}|t_{2,1}) > 2/3$, where

$$\begin{aligned} \mu_2^1 &:= \frac{\sum_l A_{l1}\mu^{l1}}{A_1^2}, \\ \beta_2^*(t_{1,1}|t_{2,1}) &:= \frac{A_{11}}{A_1^2}. \end{aligned} \tag{B5}$$

This is precisely condition (B2) for $r = 1$, where (B5) is agent 2's second-order belief that $t_1 \in \rho_1^1$. Given that agent 2 invests on the basis of second-order beliefs at $t_{2,1}$, agent 1 invests on the basis of third-order beliefs if $t_{1,2} \in \rho_1^3$, which is equivalent to condition (B3) for $l = 2$, and so on. In conclusion, the system (B1)–(B4) describes the most efficient way of inducing joint investment with probability 1. If the system has a solution, it must be e^*_μ.

Proof of Claim 1

Only if.—By way of contradiction, suppose that e^*_μ induces joint investment with certainty at $\mu \leq 1/2$. Then, e^*_μ must solve system (B1)–(B4). Adding up inequalities (B1) and (B3) for $l = 2, \ldots, L$, we obtain

$$\sum_{l,r} A_{lr}\mu^{lr} > \frac{2}{3}\sum_l A_l^1 - \frac{1}{3}\sum_{l=2}^{L}\sum_{r=1}^{l-1} A_{lr},$$

which can be rewritten as

$$\mu > \frac{2}{3} - \frac{1}{3}\sum_{l=2}^{L}\sum_{r=1}^{l-1} A_{lr}, \tag{B6}$$

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

using equation (B4). Adding up equations (B2) for $r = 1, \ldots, R$, we similarly obtain

$$\mu > \frac{2}{3} - \frac{1}{3}\sum_{r}\sum_{l=1}^{r} A_{lr}. \tag{B7}$$

Note that $\Sigma_{l=2}^{L}\Sigma_{r=1}^{l-1}A_{lr} + \Sigma_{r}\Sigma_{l=1}^{r}A_{lr} = \Sigma_{r}^{R}\Sigma_{l}^{L}A_{lr} = 1$. Therefore, adding up equations (B6) and (B7), we get $2\mu > 2 \cdot (2/3) - (1/3)$, which implies that $\mu > 1/2$.

If.—Suppose that $\mu > 1/2$. For $\mu > 2/3$, e_{μ}^{*} in table B1 induces joint investment with certainty. For $\mu \in (1/2, 2/3)$, it can be easily checked that e_{μ}^{*} in table 3 is a solution to the system (B1)–(B4) with $R = 1$ and $L = 2$ and hence induces joint investment with certainty. QED

From claim 1, we know that for $\mu \leq 1/2$, the designer can no longer ensure that I will be played at all hierarchies. By lemma 5, the hierarchies at which N is played must have first-order beliefs of zero. To obtain the optimal minimal distributions for $\mu \leq 1/2$, we set

$$\mu_{2}^{R} := \frac{\sum_{l} A_{lR}\mu^{lR}}{A_{R}^{2}} = 0 \tag{B8}$$

and replace condition (B2) for $r = R$ with it. We then find the smallest μ for which the system has a solution and repeat this procedure by setting $\mu_{1}^{L} = 0$ and replacing condition (B3) for $l = L$ with it, and so on.

Proof of Claim 2

Consider the case when the last K types of agent 1 and the last K types of agent 2 do not invest.[18] This means that the system becomes

$$\begin{cases} \sum_{r} A_{1r}\mu^{1r} > \frac{2}{3}A_{1}^{1}, & \text{(B9)} \\ \sum_{l} A_{lr}\mu^{lr} > \frac{2}{3}A_{r}^{2} - \frac{1}{3}\sum_{l=1}^{r} A_{lr} \quad \text{for } r = 1, \ldots, R-K, & \text{(B10)} \\ \sum_{r} A_{lr}\mu^{lr} > \frac{2}{3}A_{l}^{1} - \frac{1}{3}\sum_{r=1}^{l-1} A_{lr} \quad \text{for } l = 2, \ldots, L-K, & \text{(B11)} \\ \sum_{l} A_{lr}\mu^{lr} = 0 \quad \text{for } r = R-K+1, \ldots, R, & \text{(B12)} \\ \sum_{r} A_{lr}\mu^{lr} = 0 \quad \text{for } l = L-K+1 \ldots, L, & \text{(B13)} \\ \sum_{l,r} A_{lr}\mu^{lr} = \mu, & \text{(B14)} \end{cases}$$

where the first-order beliefs of the last K noninvesting types of both agents have been set equal to zero. Adding up each side of the conditions corresponding to agent 1 ([B9], [B11], and [B13]) and using equation (B14), we get

[18] Note that the only options are that the last K types of both agents do not invest or the last K types of agent 1 and the last $K-1$ types of agent 2 do not invest.

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

$$\mu > \frac{2}{3}\sum_{l=1}^{L-K} A_l^1 - \frac{1}{3}\sum_{l=2}^{L-K}\sum_{r=1}^{l-1} A_{lr} = \frac{2}{3}\sum_{l=1}^{L-K}\sum_{r=1}^{R-K} A_{lr} + \frac{2}{3}\sum_{l=1}^{L-K}\sum_{r=R-K+1}^{R} A_{lr} - \frac{1}{3}\sum_{l=2}^{L-K}\sum_{r=1}^{l-1} A_{lr} \tag{B15}$$

and, correspondingly, for agent 2 (adding up each side of conditions [B10] and [B12]),

$$\mu > \frac{2}{3}\sum_{r=1}^{R-K} A_r^2 - \frac{1}{3}\sum_{r=1}^{R-K}\sum_{l=1}^{r} A_{lr} = \frac{2}{3}\sum_{l=1}^{L-K}\sum_{r=1}^{R-K} A_{lr} + \frac{2}{3}\sum_{l=L-K+1}^{L}\sum_{r=1}^{R-K} A_{lr} - \frac{1}{3}\sum_{r=1}^{R-K}\sum_{l=1}^{r} A_{lr}. \tag{B16}$$

Summing up the above two equations, we obtain

$$2\mu > \sum_{l=1}^{L-K}\sum_{r=1}^{R-K} A_{lr} + \frac{2}{3}\sum_{l=1}^{L-K}\sum_{r=R-K+1}^{R} A_{lr} + \frac{2}{3}\sum_{l=L-K+1}^{L}\sum_{r=1}^{R-K} A_{lr}, \tag{B17}$$

where we have used that

$$\frac{1}{3}\sum_{l=2}^{L-K}\sum_{r=1}^{l-1} A_{lr} + \frac{1}{3}\sum_{r=1}^{R-K}\sum_{l=1}^{r} A_{lr} = \frac{1}{3}\sum_{l=1}^{L-K}\sum_{r=1}^{R-K} A_{lr}.$$

Note that inequality (B17) can be written as

$$2\mu > \Pr(\text{both invest}) + \frac{2}{3}\Pr(\text{only 1 invests}) + \frac{2}{3}\Pr(\text{only 2 invests}). \tag{B18}$$

Multiplying both sides by $v(\mathrm{I},\mathrm{I}) > 0$, we obtain

$$2\mu v(\mathrm{I},\mathrm{I}) > v(\mathrm{I},\mathrm{I})\left(\Pr(\text{both invest}) + \frac{2}{3}\Pr(\text{only 1 invests}) + \frac{2}{3}\Pr(\text{only 2 invests})\right). \tag{B19}$$

Note that the left-hand side of this inequality is $(\mathrm{cav}\tilde{w}^*)(\mu)$ for $\mu \le 1/2$. Let us denote the right-hand side of (B19) by RHS(B19). The designer's expected value at any optimal $e_\mu^* \in \mathcal{C}^{\mathrm{M}}$ with m messages can be written as

$$\begin{aligned}\mathbb{E}_{e_\mu^*}[v] = {} & v(\mathrm{I},\mathrm{I})\Pr(\text{both invest}) + v(\mathrm{I},\mathrm{N})\Pr(\text{only 1 invests}) \\ & + v(\mathrm{N},\mathrm{I})\Pr(\text{only 2 invests}).\end{aligned} \tag{B20}$$

Then,

$$\begin{aligned}\mathrm{RHS(B19)} - \mathbb{E}_{e_\mu^*}[v] = {} & \left(\frac{2}{3}v(\mathrm{I},\mathrm{I}) - v(\mathrm{N},\mathrm{I})\right)\Pr(\text{only 1 invests}) \\ & + \left(\frac{2}{3}v(\mathrm{I},\mathrm{I}) - v(\mathrm{N},\mathrm{I})\right)\Pr(\text{only 2 invests}).\end{aligned} \tag{B21}$$

Hence, if $(2/3)v(\mathrm{I},\mathrm{I}) - v(\mathrm{N},\mathrm{I}) \ge 0$, then $\mathrm{RHS(B19)} \ge \mathbb{E}_{e_\mu^*}[v]$. This, in turn, implies that

$$(\mathrm{cav}\tilde{w}^*)(\mu) > \mathbb{E}_{e_\mu^*}[v]$$

for $\mu \le 1/2$ and for any optimal $e_\mu^* \in \mathcal{C}^{\mathrm{M}}$. QED

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

References

Alonso, Ricardo, and Odilon Câmara. 2016. "Persuading Voters." *A.E.R.* 106 (11): 3590–605.
Angeletos, George-Marios, and Alessandro Pavan. 2007. "Efficient Use of Information and Social Value of Information." *Econometrica* 75 (4): 1103–42.
Aumann, Robert J. 1976. "Agreeing to Disagree." *Ann. Statis.* 4 (6): 1236–39.
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." *J. Econ. Theory* 144 (3): 1320–32.
Bergemann, Dirk, Benjamin Brooks, and Stephen Morris. 2017. "First-Price Auctions with General Information Structures: Implications for Bidding and Revenue." *Econometrica* 85 (1): 107–43.
Bergemann, Dirk, and Stephen Morris. 2016. "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games." *Theoretical Econ.* 11:487–522.
———. 2019. "Information Design: A Unified Perspective." *J. Econ. Literature* 57 (1): 44–95.
Best, James, and Daniel Quigley. 2018. "Persuasion for the Long-Run." Working paper, Oxford Univ.
Brandenburger, Adam, and Eddie Dekel. 1993. "Hierarchies of Beliefs and Common Knowledge." *J. Econ. Theory* 59 (1): 189–98.
Brocas, Isabelle, and Juan D. Carrillo. 2007. "Influence through Ignorance." *RAND J. Econ.* 38 (4): 931–47.
Carlsson, Hans, and Eric van Damme. 1993. "Global Games and Equilibrium Selection." *Econometrica* 61 (5): 989–1018.
Chan, Jimmy, Seher Gupta, Fei Li, and Yun Wang. 2019. "Pivotal Persuasion." *J. Econ. Theory* 180:178–202.
Dekel, Eddie, Drew Fudenberg, and Stephen Morris. 2007. "Interim Correlated Rationalizability." *Theoretical Econ.* 2:15–40.
Ely, Jeffrey. 2017. "Beeps." *A.E.R.* 107 (1): 31–53.
Ely, Jeffrey, and Marcin Peski. 2006. "Hierarchies of Belief and Interim Rationalizability." *Theoretical Econ.* 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): 213–33.
Hoshino, Tetsuya. 2018. "Using Strategic Uncertainty to Persuade Multiple Agents." Working paper, Pennsylvania State Univ.
Inostroza, Nicolas, and Alessandro Pavan. 2017. "Persuasion in Global Games with Application to Stress Testing." Working Paper, Northwestern Univ.
Kajii, Atsushi, and Stephen Morris. 1997. "The Robustness of Equilibria to Incomplete Information." *Econometrica* 65 (6): 1283–309.
Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian Persuasion." *A.E.R.* 101 (6): 2590–615.
Laclau, Marie, and Ludovic Renou. 2016. "Public Persuasion." Working paper.
Liu, Qingmin. 2009. "On Redundant Types and Bayesian Formulation of Incomplete Information." *J. Econ. Theory* 144 (5): 2115–45.
———. 2015. "Correlation and Common Priors in Games with Incomplete Information." *J. Econ. Theory* 157:49–75.
Mathevet, Laurent, David Pearce, and Ennio Stacchetti. 2018. "Reputation and Information Design." Working paper, New York Univ.
Mertens, Jean-François, and Shmuel Zamir. 1985. "Formulation of Bayesian Analysis for Games with Incomplete Information." *Internat. J. Game Theory* 14 (1): 1–29.

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

Morris, Stephen, and Hyun Song Shin. 2002. "Social Value of Public Information." *A.E.R.* 92 (5): 1521–34.

———. 2003. "Global Games: Theory and Applications." In *Advances in Economics and Econometrics, the Eighth World Congress*, edited by Mathias Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky, 56–114. Cambridge: Cambridge Univ. Press.

Rayo, Luis, and Ilya Segal. 2010. "Optimal Information Disclosure." *J.P.E.* 118 (5): 949–87.

Rockafellar, R. Tyrrell. 1970. *Convex Analysis.* 2nd ed. Princeton, NJ: Princeton Univ. Press.

Rubinstein, Ariel. 1989. "The Electronic Mail Game: Strategic Behavior under 'Almost Common Knowledge'." *A.E.R.* 79 (3): 385–91.

Taneva, Ina. 2019. "Information Design." *American Econ. J.: Microeconomics* 11 (4): 151–85.

Vives, Xavier. 1988. "Aggregation of Information in Large Cournot Markets." *Econometrica* 56 (4): 851–76.

Yildiz, Muhamet. 2015. "Invariance to Representation of Information." *Games and Econ. Behavior* 94:142–56.

Zhang, Jun, and Junjie Zhou. 2016. "Information Disclosure in Contests: A Bayesian Persuasion Approach." *Econ. J.* 126:2197–217.

This content downloaded from 129.215.019.157 on March 05, 2020 06:17:08 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).