Coordination through Information Design in the Presence of Hype[*]

Preliminary

Christopher Teh[†]

This version: March 18, 2024 (Latest)

Abstract

I study information design in binary-action supermodular games. Agents determine their equilibrium behaviour by introspection, which rests on an initial belief about others' behaviour called *hype*. I show that all outcomes are implemented by information structures which send "direct" recommendations to some agents to begin taking the high action (invest) early in the introspection process, i.e., as *leaders*, then sends "private email" recommendations to other agents to invest later on, i.e., as *followers*. I show that by reducing the cost of creating leaders, increasing hype increases aggregate investment and changes the optimal composition of leaders and followers to induce investment. Thus, optimal information structures can vary non-monotonically in hype.

Keywords: Information Design, Supermodular Games, Equilibrium Selection

[*]I thank my advisors Richard Holden, Keiichi Kawai and Anton Kolotilin for their continual guidance and support, and participants at the SAET 2023, ESAM Sydney 2023, AETW 2024, and local workshops at UNSW Sydney and the Toulouse School of Economics (TSE) for providing insightful comments. This paper has also benefited greatly from discussions with many local Academics and PhD students during my visit to TSE in the fall of 2023, of which I am indebted to.

[†]UNSW Sydney, chris.teh@unsw.edu.au

1 Introduction

In settings with strategic complementarities, how should a designer optimally disclose information to agents to facilitate coordination? This question has clear economic relevance. For instance, to designing banking stress tests to limit bank runs, media propaganda which determines citizen engagement in regime uprisings, and the information policy of crowdfunding platforms raising funds for projects. Here, the answer depends critically on the assumption of equilibrium selection, as evident by current analyses studying the extremes of designer-preferred (e.g., Bergemann and Morris, 2019) or adversarial (e.g., Mathevet et al., 2020) selection, which offer starkly different predictions. Yet lessons from behavioural economics suggest that the reasoning of real-world agents is rarely as extreme, and often influenced by payoff-irrelevant factors which the current literature is silent about. This paper considers information design in *Binary-Action Supermodular (BAS)* games under a selection criterion which not only incorporates such concerns, but also includes prior analyses from the literature as special cases.

In the model, a unit mass of symmetric agents each face a binary decision: invest or not invest. Agents' payoffs depend on a unknown state-of-the-world and the aggregate investment. An agent's incentive to invest over not is non-decreasing in the aggregate investment. The information designer prefers a larger aggregate investment on each state, and she chooses the information structure of agents to maximize her expected payoff, given the underlying Bayes-Nash equilibrium (BNE) selection criterion.

The key assumptions I make revolve around equilibrium selection. Given an information structure, one first defines a sequence of best-responses for agents that coincides with Level-K thinking (Crawford et al., 2013): a L1 best-response given an "initially believed" aggregate investment, a L2 best-response assuming other agents follow L1 strategies, and so on. If agents only switch from not investing to investing along the sequence, then I call the limit the *monotone introspective equilibrium*. The BNE selected is the monotone introspective equilibrium supported by the largest initial belief subject to an upper bound, called the *level of hype*. An increase in hype leads to a larger BNE selected, and the largest and smallest BNE, studied in the literature,[1] are selected under largest and smallest hype respectively.

[1]This is as prior works often consider a designer with monotone preferences. There, designer-

I first consider optimizing over *public* information structures. I show that the optimal public information structure has a simple threshold structure. Ordering states in decreasing order of the cost (agent's payoff from investing over not on the state given the level of hype) to benefit (designer's benefit from agents investing over not) ratio, all agents invest only on sufficiently highly ordered states. Increasing hype not only increases agents' incentives to invest on every state, but also increases agents' incentives to invest on different states at *different rates*. The former implies the designer benefits from higher levels of hype. The latter implies that changes in hype can change the order over states, which leads to possible non-monotonicities in the optimal information structure.

I then consider optimizing over all information structures. Fixing the level of hype, I observe that under a monotone introspective equilibrium of an information structure, any signal that induces an agent to invest "assigns" agents to one of two roles. Agents can either be a *leader* who begins investing at the earliest level of introspection, or a *follower* who begins investing at a later level. Every information structure *implements* a family of conditional distributions over (masses of) leaders and followers on each state. I call these *leader-follower outcomes*, which are my paper's analogue to outcomes in the standard approach to studying information design (Bergemann and Morris, 2016).

I show that implementable leader-follower outcomes are fully characterized by *introspective obedience* constraints. These can be understood in the context of the "hybrid" information structures which canonically implement all such outcomes. There, agents observe one of three kinds of recommendations. First, a "direct" recommendation to be a leader and so invest at all levels. Agents obey this if and only if *leader obedience* holds: the agent prefers to invest given both (i) the level of hype, and (ii) only other leaders invest. Second, a "private email" recommendation to be followers and so invest only at some later level. Agents obey this if and only if *follower obedience* holds: the agent prefers to invest given all other leaders invest and holding uniform beliefs over the mass of other followers who invest. Third, a direct recommendation not to invest. Agents obey this if and only if *downwards obedience* holds: the agent prefers to invest given both (i) the level of hype, and (ii) all other leaders and followers do not invest. Notably, the introspective obedience constraints slacken from increases in hype, and "collapse" to the standard

preferred and largest selection coincide, and adversarial selection and smallest selection coincide.

obedience constraints (Bergemann and Morris, 2016) under the highest hype, and those which characterise smallest equilibrium implementable outcomes (Morris et al., 2022a, Morris et al., 2022b) under the lowest hype.[2]

I use my implementation result to study the general information design problem. A key observation is that increasing hype reduces the "cost" of inducing investment via leaders (over followers). This induces the designer to switch to a leader-follower outcome with a greater mass of leaders. In turn, a leader- only or follower-only outcome is optimal only under extreme hype. I provide general necessary and sufficient conditions for when these occur, and find that for most values of hype, a mix of leaders and followers is strictly optimal.

I obtain a sharp characterization of optimal leader-follower outcomes in *threshold games*, where agents' payoffs are constant in the aggregate investment when they prefer not to invest (e.g., the Regime change game of Morris and Shin (2003)). Under an optimal leader-follower outcome, all agents invest only on state ordered high enough, where the order depends on the cost of inducing investment from an "efficient" composition of leaders and followers on the state, relative to the designer's benefit from investment. Changes in hype affect the ordering by changing the efficient composition. When payoffs are monotone in states, the ordering does not change, so increasing hype increases the set of states on which all agents invest, and the mass of leaders who invest on each state.

I extend the model in three ways, while restricting the designer to public information. First, I allow hype to either be chosen by the designer, or indirectly determined by the information structure. I provide sufficient conditions for a simple upper censorship information structures to be optimal. Second, I allow agents to obtain exogenous private information about the state. I find that an increase in hype also benefits the designer through a *bandwagon effect*, i.e., where a greater set of agents "discard" their private information and begin investing earlier. Third, I consider when hype is uncertain to the designer, and draw a connection to general *Bayesian Persuasion* problems (Kamenica and Gentzkow, 2011).

My paper contributes to the vast literature on information design, particularly to the study of BAS games. A non-exhaustive list of preceding works include, for designer-preferred selection, Arieli and Babichenko (2019) (binary states), Candogan and Drakopoulos (2020) (linear networks), and Taneva and Mathevet (2023)

[2] Refer to Section 4 for a precise statement, and comparison to prior results.

(constrained information design), for adversarial selection, Goldstein and Huang (2016) and Inostroza and Pavan (2023) (public information with exogenous private information in global games), Li et al. (2022) (general information structures with no private information in global games) and Mathevet et al. (2020) (through applying the belief-based approach).[3] Relative to these, I study information design in a general setting under various degree of equilibrium adversariality.

My paper also relates to the literature on implementation through information design in games. Aumann (1974) and Bergemann and Morris (2016) characterise the outcomes partially implementable, i.e., under some form of equilibrium selection, in games without and with common state uncertainty respectively, the weakest form of implementation. Building on the earlier insights of Rubinstein (1989) and Carlsson and van Damme (1993), Oyama and Takahashi (2020) and Morris and Ui (2005) provide necessary and sufficient conditions, respectively, for unique implementation, the strongest form of implementation, of equilibria for complete information BAS games.[4] Morris et al. (2022a) extends the construction of Oyama and Takahashi (2020) to fully characterize smallest and unique implementation in finite-player incomplete information BAS games, and Morris et al. (2022b) extends these to having a continuum of players.[5] I build upon these results by providing conditions for "intermediate" implementation under a given level of hype.

Monotone introspective equilibrium is a special case of the *introspective equilibrium* solution concept introduced in Kets and Sandroni (2020). The solution concept has been applied in many contexts recently, for instance, in the study of valuing coordination games (Kets et al., 2022), organisational design (Kets, 2021), raising investment (Akerlof and Holden, 2019), competition in markets with network externalities (Akerlof et al., 2023), and supply chain coordination (Akerlof and Holden, 2023). All these papers focus on a complete-information environment and consider the design of various non-informational aspects of the game. I focus on an incomplete information setting, and consider optimal information design.

The interaction between leaders and followers in coordinating investments takes

[3]Also related are papers studying unique implementation in supermodular environments through minimum cost incentives which allow for information design (Hoshino, 2022, Halac et al., 2021, Halac et al., 2022, Morris et al., 2022c)

[4]To be precise, Morris and Ui (2005) provided sufficient conditions for the equilibrium a complete information game to be fully implemented in an ϵ-elaboration of the game by an information structure. Oyama and Takahashi (2020) shows that the condition is necessary in BAS games.

[5]Both papers also study adversarial information design in general BAS games.

center-stage in my model. More generally, Andreoni (1998) studies (among others) the problem of a charity choosing an optimal mix of leading contributors to achieve a charitable fundraising goal. Morris and Shin (2006) studies how the IMF (leader), through funding injections, can coordinate the efforts of interested parties (followers) to mitigate financial crises. Akerlof and Holden (2019) considers how investors with large block or network capital (leaders) can facilitate capital assembly (also see Akerlof and Holden, 2016). Deb et al. (2023) and Ellman and Fabi (2022) study how (strategically timed) contributions by early funders (leaders) can be used to coordinate other funders' contributions (followers) in crowdfunding. A novel contribution of this paper is in examining the designer's incentives to optimally create leaders and followers through offering differential information, and how this trade-off is shaped by the degree of equilibrium adversariality.

The rest of the paper is organised as follows. Section 2 introduces the model. Section 3 studies the benchmark where the designer is restricted to public information. Section 4 considers implementation in the general setting. Section 5 studies optimal information design in the general setting. Section 6 considers extensions. All proofs are relegated to the Appendix A-C, and the Online Appendix.

2 The Model

2.1 Statement of the Problem

Base Game There are two kinds of players: a designer and a unit mass of symmetric agents $i \in [0,1]$. Each agent faces a binary decision $a_i \in \{0,1\}$, where I refer to $a_i = 1$ and $a_i = 0$ as *invest* and *not invest* respectively. Meanwhile, the designer chooses the information structure (defined later), which provides agents with information about an unknown state of the world $\theta \in \Theta \equiv [\underline{\theta}, \overline{\theta}]$, where $-\infty < \underline{\theta} < \overline{\theta} < \infty$ Players share a common prior $F \in \Delta(\Theta)$, where I assume F has full support, and is continuously differentiable with density f.[6]

An agent's payoff from not investing is zero. Meanwhile, an agent's payoff from investing, and the designer's payoff, depends both on the state θ and the aggregate investment $A \in [0,1]$. I denote the agent's payoff from investing by $D(A,\theta)$, and the designer's payoff by $v(A,\theta)$. Both $D(A,\theta)$ and $v(A,\theta)$ are assumed

[6] $\Delta(X)$ denotes the set of Borel probability measures over X, equipped with the weak topology.

to be upper-semicontinuous in (A, θ).

I focus on an environment in which agents' investments are complementary, and the designer aims to foster investment by agents. That is, I assume that for all $\theta \in \Theta$, $D(A, \theta)$ and $v(A, \theta)$ are non-decreasing in $A \in [0, 1]$. To simplify exposition, I further assume that the designer strictly prefers for all agents to invest over not, i.e., $v(0, \theta) = 0 < v(1, \theta)$, and that there exists a measurable subset $\overline{\Theta} \subseteq \Theta$ with $\int_{\overline{\Theta}} dF(\theta) > 0$ such that for all $\theta \in \overline{\Theta}$, $D(0, \theta) > 0$ holds.[7]

Information Structure An *information structure* is a pair $\mathcal{S} \equiv (S, (\pi(\cdot|\theta))_{\theta \in \Theta})$.[8] $S \subseteq \mathbb{R}$ is a non-empty Borel set of signals observed by agents. Meanwhile, for each $\theta \in \Theta$, $\pi(\cdot|\theta) \in \Delta(\Delta(S))$ is a distribution over signal distributions for agents $\mu \in \Delta(S)$ (see timing below), under which $\pi(Y|\theta)$ is measurable in θ for each $Y \in \mathcal{B}(\Delta(S))$, where $\mathcal{B}(X)$ denotes the Borel sigma-algebra of X.

Given an information structure $\mathcal{S}$, let π denote the induced joint distribution over $S \times \Delta(S) \times \Theta$, i.e., so $\pi(X \times Y \times Z) \equiv \int_Z \int_Y \int_X d\mu(s) d\pi(\mu|\theta) f(\theta) d\theta$ for each $X \in \mathcal{B}(S)$, $Y \in \mathcal{B}(\Delta(S))$ and $Z \in \mathcal{B}(\Theta)$. Further let $\pi(\cdot|s)$ denote any version of the regular conditional probability on $\Delta(S) \times \Theta$ given $s \in S$. Finally, let the marginal distribution over the signal space S and the product $\Delta(S) \times \Theta$ be denoted by π_S.

Timing First, the designer commits to an information structure $\mathcal{S}$. Next, the state $\theta \in \Theta$ is drawn according to F, and a signal distribution $\mu \in \Delta(S)$ is drawn according to $\pi(\cdot|\theta)$. Signals are allocated *anonymously* across agents, so each agent then privately observes an independent draw of a signal from μ. By an appropriate "large of large numbers" (Sun, 2006), μ also represents the empirical distribution over signal realizations, and so an agent who observes signal $s \in S$ develops a posterior belief over (the distribution over) other agents' signals and the state $(\mu, \theta) \in \Delta(S) \times \Theta$, given by $\pi(\cdot|s)$. Each agent then independently chooses $a_i \in \{0, 1\}$. Given the aggregate investment and state, players obtain their payoffs.

Solution Concept An information structure $\mathcal{S}$ induces a Bayesian game between agents. A pure-strategy for agents in this game is a measurable function $\alpha_i : S \rightarrow$

[7]The former assumption simply makes the orderings over θ in Sections 3 and 5 easier to state. The latter assumption implies that regardless of the degree of adversariality of equilibrium selection, the designer's information design problem is non-trivial.

[8]I follow the modelling approach of Morris et al. (2022b)

$\{0,1\}$. I focus on symmetric pure strategies throughout, and so drop the subscript.

Conditional on distribution μ being drawn, let $A(\alpha|\mu) \equiv \mu(\{s \in S : \alpha(s) = 1\})$ denote the measure of agents who invest under pure strategy α. A (symmetric) *Bayes-Nash Equilibrium (BNE)* of $\mathcal{S}$ is a pure strategy α^* under which no agent has a unilateral incentive to deviate given their beliefs about other agents signals, $\pi(\cdot|s)$:

$$\alpha^*(s) \in \arg\max_{a\in\{0,1\}} \left\{ \mathbb{I}_{a=1} \int_{\Delta(S)\times\Theta} D(A(\alpha^*|\mu), \theta) d\pi(\mu, \theta|s) \right\}, \quad \forall s \in S$$

The Information Design Problem The designer chooses an information structure $\mathcal{S}$ to maximize her expected payoff, subject to agents' behaviours coinciding with the designer's anticipated BNE of $\mathcal{S}$. Here, I assume that the designer's anticipated BNE is her most-preferred *h-monotone introspective equilibrium* of $\mathcal{S}$ subject to $h \leq H$, where $H \in [0,1]$ is referred to as the *maximum level of hype*. I defer the discussion of such concepts to Section 2.2 next. For now, I denote this equilibrium by $\alpha^{\mathcal{S},H}$, so that the designer's problem is

$$V^*(H) \equiv \sup_{\mathcal{S}} \int_\Theta \int_{\Delta(S)} v(A(\alpha^{\mathcal{S},H}|\mu), \theta) d\pi(\mu|\theta) f(\theta) d\theta \tag{1}$$

I refer to any solution of (1) as an *optimal* information structure for the designer.

2.2 Monotone Introspective Equilibrium

I begin by introducing the notion of a monotone introspective equilibrium.

Definition 1. *Given an information structure $\mathcal{S}$, and $h \in [0,1]$, let $(\alpha^{\mathcal{S},h,k})_{k=1}^\infty$ be the sequence of pure-strategies defined as follows*

1. *If $k = 1$, then*

$$\alpha^{\mathcal{S},h,1}(s) \equiv \begin{cases} 1, & \int_{\Delta(S)\times\Theta} D(h,\theta) d\pi(\mu,\theta|s) \geq 0 \\ 0, & \textit{otherwise} \end{cases}, \quad \forall s \in S \tag{2}$$

2. *If* $k > 1$, *then*

$$\alpha^{\mathcal{S},h,k}(s) \equiv \begin{cases} 1, & \int_{\Delta(S)\times\Theta} D(A(\alpha^{\mathcal{S},h,k-1}|\mu),\theta)d\pi(\mu,\theta|s) \geq 0 \\ 0, & \text{otherwise} \end{cases}, \quad \forall s \in S \tag{3}$$

If the sequence $(\alpha^{\mathcal{S},h,k})_{k=1}^{\infty}$ *is (point-wise) non-decreasing in* k*, then I refer to the limit* $\alpha \equiv \lim_{k\to\infty} \alpha^{\mathcal{S},h,k}$ *as the* h*-monotone introspective equilibrium of* $\mathcal{S}$*.*

Definition 1 follows from the following introspection process of agents. Fix a state θ and profile of signals observed by agents. Each agent starts with an "initial" belief that the aggregate investment is h,[9] and so forms a "L1" best-response to this belief, yielding $\alpha^{\mathcal{S},h,1}$ defined in (2). Then, agents revise their beliefs through introspection, forming a "L2" best-response assuming other agents follow their L1 best-responses. This yields $\alpha^{\mathcal{S},h,2}$ defined in (3). Repeating the introspection process then yields the sequence of best-responses $(\alpha^{\mathcal{S},h,k})_{k=1}^{\infty}$. If agents only switch from not investing to investing along the sequence, then the limit exists and is referred to as the h-monotone introspective equilibrium of $\mathcal{S}$.

With an abuse of notation, I let $(\alpha^{\mathcal{S},H,k})_{k=1}^{\infty}$ denote the pure strategy sequence associated to the largest h-monotone introspective equilibrium of $\mathcal{S}$ satisfying $h \leq H$, and denote its limit by $\alpha^{\mathcal{S},H}$. Furthermore, I refer to an agent as *investing at Lk* under signal $s \in S$ if $\alpha^{\mathcal{S},H,k}(s) = 1$, and as *investing under* $\mathcal{S}$ if $\alpha^{\mathcal{S},H}(s) = 1$.

Properties of Monotone Introspective Equilibria As I prove in the Appendix (Theorem 5), every monotone introspective equilibrium is a BNE. That is, monotone introspective equilibrium is a form of equilibrium selection. In particular, for a given information structure $\mathcal{S}$, through applying the argument of Milgrom and Roberts (1990) and breaking ties in favour of investment, the smallest BNE is the monotone introspective equilibrium for $h = 0$. Following the argument in Morris et al. (2022b), the smallest BNE exists, so $\alpha^{\mathcal{S},H}$ is well-defined for each $H \in [0,1]$, and the designer's problem under adversarial equilibrium selection coincides with the designer in our problem expecting the lowest level of hype. By

[9] In the interpretation of Kets and Sandroni (2020), each agent independently observes an impulse to invest and not to invest with probabilities h and $1-h$ respectively. Initially, all agents are believed to blindly follow their impulses, so the aggregate investment is h.

a similar argument, for $H = 1$, the designer's problem coincides with that under designer-preferred equilibrium selection.

More important are the "intermediate" monotone introspective equilibria which are neither the smallest nor largest BNE. For example, if $D(A, \theta) = \mathbb{I}_{A \geq \theta - \epsilon} - 1/2$ for $\epsilon > 0$ small, then under full disclosure (so $S = \Theta$), there exists a continuum of monotone introspective equilibria $(\alpha^{\mathcal{S},h})_{h \in [0,1]}$ which involves $\alpha^{\mathcal{S},h}(s) = 1$ if $h \geq s - \epsilon$, and $\alpha^{\mathcal{S},h}(s) = 0$ if $h < s - \epsilon$. How these equilibria shape the designer's problem for intermediate levels of hype play a big role in the following analysis.

Remark 1. *Monotone introspective equilibrium is a refinement of the Introspective equilibrium solution concept introduced in Kets and Sandroni (2020). There, the sequence $(\alpha^{\mathcal{S},h,k})_{k=1}^{\infty}$ is only required to converge. The additional monotonicity requirement has no effect when the designer only uses public information (Section 3), but is otherwise desirable for two reasons. First, verifying whether an information structure has a monotone introspective equilibrium reduces to verifying that if an agent invests at L1 under a signal, then the agent also invests at L2 (see Lemma 4, Appendix A). This adds significant tractability to the analysis. Second, monotonicity rules out introspective equilibria which rely on overly optimistic initial beliefs to induce investment. Notice that an agent invests at L1 if the anticipated investment of other agents h is large, and at L2 if the actual investment of agents (at L1) is large. Thus, the designer only anticipates equilibria selected where actual investment is at least as large as the anticipated investment.*

2.3 An application to crowdfunding

I now interpret the model in the context of crowdfunding. I will provide a concrete interpretation of "hype", and an example of how my results can be applied.

Consider a group of funders (agents) deciding whether to invest in the project of an entrepreneur. The project is hosted on a crowdfunding platform (designer), and yields a positive return to funders if and only if enough funds are raised. Funders incur a cost from investing (e.g., due to a *flexible funding* model, as in Indiegogo, or opportunity costs from foregone investment opportunities), and cannot observe others' behaviours (e.g., due to a short funding window).[10] The platform earns commissions from investments, and aims to maximize profits.

[10] Beyond the short funding window, the static assumption, applied in several studies of crowdfunding (e.g. Strausz, 2017, Ellman and Hurkens, 2019, Chang, 2020) is not entirely unreasonable. Empirical evidence suggests that early investment is a key determinant for project funding success

Funders' investments depend on two sources of uncertainty. First, on beliefs about project characteristics (e.g. product features), which affects their benefit from investing if the project is funded. What funders learn is determined by the platform, who mandates what and how information about project characteristics is communicated to funders, i.e., the information structure.

Second, on beliefs about whether other funders will invest. Uncertainty arises here due to the anonymity and "small-scale" of funders, typical in crowdfunding. Here, funders' reasoning often depends on *pre-launch* hype around the project. For instance, on the project awareness created through marketing and social media posts, entrepreneur / platform reputation, and past performance by the platform. Notably, platforms have limited control over hype. While entrepreneurs are strongly recommended to create hype (e.g., Kickstarter, 2023, Indiegogo, 2023), the recommendation is non-binding. Also, hype can be dependent on past performance or prior trends inherited by the current platform manager. These constraints are captured by the exogenous upper bound H.

By the above, the platform's problem is to choose funders' information (structure) to maximize investment, holding fixed the hype around the project. In practice, one often observes a simple "hybrid" disclosure method by platforms: a big announcement to induce some funders to invest, and follow-up private recommendations to other funders to invest. For example, Kickstarter's newsletters, widely circulated among funders, are reserved for new projects, while their targeted recommendations involve projects already with a sizable backing.

My results shed light on the preceding phenomenon. Theorem 2 shows that all equilibrium outcomes under introspection can be implemented by a similar hybrid information structure. Furthermore, Theorem 3 states that unless hype is sufficiently extreme, both "parts" of the hybrid information structure are *necessary* for the platform to maximize investments. Finally, Section 5.3 explicitly characterises the optimal information structure for a generalized funding setting, while Corollary 3 shows that higher levels of hype imply that more agents are leaders, i.e., a more wide-spread initial annoucement.

My results also provide a rationale for *why* platforms prefer "hyped-up" projects. Higher levels of hype allow the platform to exploit funders' greater incentives

(e.g. Colombo et al., 2015, Crosetto and Regner, 2018). Also, platforms themselves may also have an incentive to limit funders' information about investments (Ellman and Fabi, 2022).

to invest "early" in the project to more widely recommend investment into the project. Also, higher levels of hype generate a bandwagon effect, drowning out funders' private information (if any), which can inhibit investment (Section 6.2).

3 Benchmark: Public Information

I first characterise the designer's optimal information structure under the restriction to *public* information structures, i.e., when all agents observe the same signal drawn.[11] The goal is to abstract away from how hype affects the *type* of information provided by the designer, e.g., private vs public, and draw focus to how hype affects the set of states on which agents invest, i.e., the "reach" of persuasion.

Excessive assurance I first pin down agents' monotone introspective equilibrium behaviour. Fix a public information structure $\mathcal{S}$, and a level of hype $H \in [0,1]$. Take any signal $s \in S$, and suppose that agents invest at L1 upon observing s. With symmetric strategies and public information, the agent believes that all other agents invest under signal s. Thus,

$$\underbrace{\int_{\Delta(S)\times\Theta} D(A(\alpha^{\mathcal{S},H,1}|\mu),\theta)d\pi(\mu,\theta|s)}_{\text{Payoff from investing at L2+}} = \int_{\Theta} D(1,\theta)d\pi(\theta|s) \geq \underbrace{\int_{\Theta} D(H,\theta)d\pi(\theta|s)}_{\text{Payoff from investing at L1}} \geq 0$$

so the agent invests at L2. By induction, the agent invests at all higher levels. Reversing the argument, if the agent does not invest at L1, then anticipating no other agents to invest at L1, the agent does not invest at L2+. This yields the following:

Lemma 1. *Fix a public information structure $\mathcal{S}$ and $H \in [0,1]$. Then, for all signals $s \in S$, an agent invests if and only if he invests at L1.*

Lemma 1 highlights the key feature of public information: *excessive assurance*. Every agent who invests in the monotone introspective equilibrium is a *leader*, i.e., invests at all levels. Consequently, *every* agent offers every other agent assurance to continue investing at all higher levels. In turn, an agent's decision of whether to invest *only* depends on first-order beliefs and the level of hype. An implication of this is that public information is often sub-optimal: when hype H is not too

[11] Formally, for each $\theta \in \Theta$, $\pi(\cdot|\theta)$ is a distribution over Dirac measures $\{\delta_s\}_{s\in S}$.

large, any agent who invests at L1, has a strict incentive to invest at higher levels. Intuitively, the designer would then benefit from "using" the excessive assurance built to induce a subset agents to begin investing at later levels, but under lower first-order beliefs. This can only be done via a *private* information structure.

Optimal information structures I now solve the public information design problem. Because agents' behaviours are symmetric, and whether an agent invests only depends on whether she invests at L1 (Lemma 1), the designer's problem is mathematically equivalent to a single-agent problem, where (i) the agent(s) decide between investing or not, obtaining payoffs on state θ of $D(H,\theta)$ and 0 respectively, and (ii) the designer's payoff when the agent invests and does not invest are $v(1,\theta)$ and 0 respectively. Following the standard approach of Bergemann and Morris (2016), the principal's problem reduces to choosing an *outcome*, i.e., a Θ-measurable map $q:\Theta\to[0,1]$, to solve

$$\max_q \int_\Theta v(1,\theta)q(\theta)f(\theta)d\theta \tag{4}$$

$$\text{s.t.} \quad \int_\Theta D(H,\theta)q(\theta)f(\theta)d\theta \geq 0 \tag{5}$$

where $q(\theta)$ is the probability agents invest on state θ, and (5) captures' agents' *upwards-obedience* constraint, i.e., concerning deviating from investing to not.[12]

The problem above can be solved via an "efficiency argument" (verified formally in the Appendix). Define the following order $x_L^H:\Theta\to\Theta$ over states

$$x_L^H(\theta) < x_L^H(\theta') \iff \frac{-D(H,\theta')}{v(1,\theta')} \leq \frac{-D(H,\theta)}{v(1,\theta)} \tag{6}$$

$\frac{-D(H,\theta)}{v(1,\theta)}$ captures the "cost" to agents for invest at L1 on θ, $-D(H,\theta)$, relative to the designer's benefit from having agents invest, $v(1,\theta)$. By a standard marginal reasoning, it is optimal for the agent to only have agents invest on the states with the lowest ratios. Since $\underline{x}_L^H$ orders states in decreasing order of their ratios $\frac{-D(H,\theta)}{v(1,\theta)}$,

[12] As the designer's payoff is monotone in investment, I avoid stating the "downwards obedience" constraint, i.e., concerning deviating from not investing to investing, here.

this translates into having agents invest if and only if $x_L^H(\theta) \geq x_L^H$, where

$$x_L^H \equiv \min \left\{ x \in \Theta : \int_{\theta : x_L^H(\theta) \geq x} D(H, \theta) f(\theta) d\theta \geq 0 \right\} \tag{7}$$

Theorem 1. *Under an optimal public information structure, all agents invest on all states with $x_L^H(\theta) \geq x_L^H$, and all agents do not invest on all states with $x_L^H(\theta) < x_L^H$.*

Theorem 1 shows that where hype plays a role in the public information design problem is in determining the relative cost of inducing investment on each state, i.e., the ratio $\frac{-D(H,\theta)}{v(1,\theta)}$. By doing so, hype determines the ordering in (6), and so which states the designer optimally induces investment on. Notably, the ordering can vary non-monotonically from changes in hype, resulting in non-monotone changes in agents' behaviour under an optimal information structure. An exception to this is when $D(A, \theta)$ and $v(1, \theta)$ are non-decreasing in θ (for all $A \in [0, 1]$). Here, an increase in H monotonically decreases the threshold state $x_L^H(\theta)$, and so increases the set of states on which agents invest.

Corollary 1. *Suppose that for all $A \in [0, 1]$, $D(A, \theta)$ and $v(1, \theta)$ are non-decreasing in θ. Then, an increase in hype H increases the subset of states on which agents invest under an optimal information structure.*

4 Implementation

I now study the general problem, focusing on simplifying the designer's problem here. I first show that agents behaviours in the monotone introspective equilibrium of any information structure can be fully described by a smaller class of *leader-follower* outcomes, the introspective equilibrium analogue to outcomes in the standard information design setting (Bergemann and Morris, 2016). I then fully characterise the leader-follower outcomes implementable by some information structure in Theorem 2, and discuss the connection to existing work.

Leader-follower outcomes Given an information structure $\mathcal{S}$, call an agent who observes a signal $s \in S$ a *leader* if he invests at the start of the introspection process, i.e., $\alpha^{\mathcal{S},H,1}(s) = 1$. Meanwhile, call an agent a *follower* if he invests only after enough rounds of introspection, i.e., there exists a $1 < \overline{k} < \infty$ such that $\alpha^{\mathcal{S},H,\overline{k}}(s) = 1$ if and only if $k \geq \overline{k}$.

Every information structure with a monotone introspective equilibrium under $H \in [0,1]$ *implements* a distribution over leaders, followers on each state. Formally, let $A^{\mathcal{S},H}(\mu) \equiv (A_L^{\mathcal{S},H}(\mu), A_F^{\mathcal{S},H}(\mu))$ denote the measurable map defined by

$$\mu \mapsto (\underbrace{\mu(\{s \in S : \alpha^{\mathcal{S},H,1}(s) = 1\})}_{\equiv A_L^{\mathcal{S},H}(\mu)}, \underbrace{\mu(\{s \in S : \alpha^{\mathcal{S},H,1}(s) = 0 < 1 = \alpha^{\mathcal{S},H}(s)\})}_{\equiv A_F^{\mathcal{S},H}(\mu)}) \tag{8}$$

Here, $A^{\mathcal{S},H}(\mu)$ associates, to each signal distribution μ, the masses of leaders and followers induced under $\mathcal{S}$. The set of all feasible pairs is given by

$$\mathcal{A} \equiv \{(A_L, A_F) \in [0,1]^2 : A_L + A_F \leq 1\}$$

Notice then that $\mathcal{S}$ *implements* a measurable mapping $\sigma : \Theta \to \Delta(\mathcal{A})$ via[13]

$$\sigma(W|\theta) \equiv \pi((A^{\mathcal{S},H})^{-1}(W)|\theta), \quad \forall \theta \in \Theta, \ \forall W \in \mathcal{B}(\mathcal{A}) \tag{9}$$

This leads us to the following definition.

Definition 2. *Call $\sigma : \Theta \to \Delta(\mathcal{A})$ a **leader-follower outcome**, and say that it is H-**implementable** if there exists an information structure $\mathcal{S}$ such that (9) holds.*

Leader-follower outcomes capture the minimal information required to describe agents' introspective equilibrium behaviour under some information structure, and are useful for two main reasons. First, they are also easy to work with. In particular, the set of implementable leader-follower outcomes are fully characterised by a small set of linear *obedience* constraints (Theorem 2). Second, they are sufficient to fully describe the designer's payoff, given by

$$V(\sigma) = \int_\Theta \int_\mathcal{A} v(A_L + A_F, \theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \tag{10}$$

Implementation I now characterise implementable leader-follower outcomes, beginning with the relevant incentive constraints for agents in different roles.

Definition 3. *Given $H \in [0,1]$, say that a leader-follower outcome σ satisfies*

[13] That is, $\sigma(\cdot|\theta)$ is the *pushforward (measure)* of $\pi(\cdot|\theta)$ onto $\mathcal{A}$ through $A^{\mathcal{S},h}$.

1. ***Leader obedience** if*

$$\int_{\Theta}\int_{\mathcal{A}} A_L D(H,\theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \geq 0 \quad \textit{and} \tag{11}$$

$$\int_{\Theta}\int_{\mathcal{A}} A_L D(A_L,\theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \geq 0 \tag{12}$$

2. ***Follower obedience** if*

$$\int_{\Theta}\int_{\mathcal{A}} \left(\int_{A_L}^{A_L+A_F} D(i,\theta) di\right) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \geq 0 \tag{13}$$

3. ***Downwards Obedience** if*

$$\int_{\Theta}\int_{\mathcal{A}} (1 - A_L - A_F) D(H,\theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta < 0 \tag{14}$$

$$\int_{\Theta}\int_{\mathcal{A}} (1 - A_L - A_F) D(A_L + A_F,\theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta < 0 \tag{15}$$

*Further say that σ is **upper obedient** (under H) if it satisfies leader and follower obedience, and **obedient** (under H) if it is upper obedient and satisfies downwards obedience.*

Given an information structure $\mathcal{S}$ which H-implements a leader-follower outcome σ, equations (11) and (12) are obtained by aggregating agents' expected payoffs from investing at L1 and L2, respectively, across all signals on which they are a leader, and so must be positive. Equation (13) is (approximately) equal to that obtained from aggregating agents' payoffs from investing at the lowest level in which they invest, across all signals under which they are a follower, and so must be positive. Finally, (14) and (15) is obtained by aggregating agents' payoffs from investing across all signals under which they do not invest, and so are all negative.

By the above, a necessary condition for σ to be H-implementable is for it to be obedient. Theorem 2 below establishes that the converse also holds. Let $T : \mathcal{A} \to [0,1]$ denote the map defined via $T(A_L, A_F) \equiv A_L + A_F$ for all $(A_L, A_F) \in \mathcal{A}$. Call a measurable map $p : \Theta \to \Delta[0,1]$ an *outcome*, where $p(A|\theta)$ captures the probability of having afraction A of agents who invest on a state. Further say that p is *implemented* by a leader-follower outcome σ, written as $p = p_\sigma$, if

$$p(\tilde{W}|\theta) \equiv \sigma(T^{-1}(\tilde{W})|\theta), \quad \forall \theta \in \Theta,\ \tilde{W} \in \mathcal{B}([0,1])$$

In terms of describing the *aggregate* investment of agents, leader-follower outcomes which implement the same outcome are identical. Theorem 2 then states that the outcome implemented by any obedient leader-follower outcome is approximately the outcome of an implementable leader-follower outcome.

Theorem 2. *Take any leader-follower outcome σ and any $H \in [0,1]$.*

1. *If σ is H-implementable, then it is obedient under H.*

2. *If σ is obedient under h, then there exists a sequence of H-implementable leader-follower outcomes $(\sigma^n)_{n=1}^{\infty}$ such that $(p_{\sigma^n})_{n=1}^{\infty}$ converges to p_σ.*

The proof of sufficiency is constructive. I sketch the rough idea here, relegating details to Appendix B. Take an obedient leader-follower outcome σ under H. When state θ is realized, draw (A_L, A_F) with probability $\sigma(A_L, A_F|\theta)$. Then, recommendations to invest are sent out to agents as follows.

First, a random mass of A_L agents observe a direct recommendation to be a leader, i.e., to invest at all levels $L1+$. Conditional on A_L being drawn, the agent is a leader with probability A_L. Thus, her (unconditional) payoff from investing at L1 on state θ is $A_L D(H, \theta)$. Aggregating over all draws of (A_L, θ) yields (11), which is positive, so the agent invests at L1. Since all leaders invest at L1, any such agent's payoff from investing at L2 on state θ is (at least) $A_L D(A_L, \theta)$. Aggregating over all draws of (A_L, θ) yields (12), and so the agent invests at L2 (and above).

Next, a random mass of A_F other agents is drawn. Each agent observes a private recommendation to invest at $Lk+$ as a follower, where $k > 1$. ks are distributed such that conditional on observing k, an agent "knows" that leaders invest, but holds (approximately) uniform beliefs over the mass of other followers $i \in [0, A_F]$ having observed a signal $k' < k$, i.e., that have already begun investing. Thus, the agent's payoff from investing at Lk conditional on pair $((A_L, A_F), \theta)$ is (at least) $\int_{A_L}^{A_L+A_F} D(i, \theta) di$. Aggregating over all pairs yields (13), so the agent obeys the recommendation to invest at Lk (and above).

Finally, the last $1-A_L-A_F$ agents observe "nothing", i.e., are recommended not to invest. Any such agent's expected payoff from investing at L1 (L2+) is at most equal to the LHS of equation (14) (equation (15)). So, the agent does not invest.

Notice that if downwards obedience is violated, then under the information structure constructed above, agents may invest when recommended not to invest. If so, one may construct another information structure which prescribes such agents to instead invest. Hence, for the purpose of identifying *optimal* leader-follower outcomes, downwards obedience can be ignored.

Corollary 2. *Take any upper obedient leader-follower outcome σ. Then, there exists an obedient leader-follower outcome $\tilde{\sigma}$ that the designer weakly prefers over σ.*

Connecting Theorem 2 to the literature Bergemann and Morris (2016) show that outcomes implementable when the designer can freely select the equilibrium, i.e., those which are *partially implementable*, are characterised by *obedience*. There, an agent must be willing to follow her action recommendation, assuming other agents *simultaneously* obey theirs. A related (but different) requirement applies to leaders and non-investing agents. If a leader-follower outcome is implementable, then (i) a leader must prefer to invest given all other *leaders* invest, and (ii) a non-investing agent must prefer not to invest given all other leaders and followers invest.

Next, Morris et al. (2022a) and Morris et al. (2022b) show that for outcomes implementable under *smallest* equilibrium selection (adversarial selection under monotone preferences), one replaces "upwards" obedience, i.e., involving deviating from investing to not, with *sequential obedience*.[14] With symmetric agents, agents must prefer to invest under (approximately) uniform beliefs over the mass of other agents investing. A related (but different) requirement applies here to followers: followers must be prefer to invest given (i) uniform beliefs over other followers investing, and (ii) also anticipating all leaders to invest.

Hence, introspective obedience is, in general, *not* equivalent to simply imposing both standard and/or sequential obedience. The distinction arises as introspective obedience (also) accounts for how leaders' investments motivate followers to invest, but not the other way around. Furthermore, there is the additional requirement for leaders to be willing to invest given the level of hype H.

However, there are two special cases where the concepts effectively coincide. First, when hype is sufficiently low, Theorem 3 later shows that a follower-only outcome is optimal. Provided hype is low so (14) holds if (15) holds, a follower-only outcome is implementable if and only if follower-obedience (which is then sequential obedience) and (15) (which is downwards obedience) holds. Hence, im-

[14] Morris et al. (2022a) focus on a finite agent setting. Morris et al. (2022b), like me, considers a continuum agent setting. The class of information structures described in this paper draws heavy inspiration from those constructed in Morris et al. (2022a).

plementable leader-follower outcomes and adversarial equilibrium implementable outcomes coincide. Second, when hype is sufficiently high, Theorem 3 later shows that a leader-only outcome where agents strictly prefer not to invest on all states on which they do not invest is optimal. Provided hype is high so (11) holds if (12) holds, such a leader-only outcome is implementable if and only if L2 leader obedience in (12) (which is then the standard obedience constraint for investing) and (15) (which is downwards obedience) holds. Hence, the relevant implementable leader-follower outcomes and partially implementable outcomes coincide.

5 Information Design

By Theorem 2 and Corollary 2, for a given $H \in [0, 1]$, the designer solves

$$\max_{\sigma:\sigma \text{ is upper obedient under } H} \int_{\Theta} \int_{\mathcal{A}} v(A_L + A_F, \theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta$$

Call any leader-follower outcome that solves this problem *optimal*. Label the designer's first-best payoff, i.e., where agents always invest, as $V^{\text{FB}} \equiv \int_{\Theta} v(1, \theta) f(\theta) d\theta$.

To draw a tighter connection to the analysis for public information, Section 5.1 begins by imposing an additional assumption to guarantee the existence of an optimal leader-follower which perfectly coordinates agents' investments. Section 5.2 then provides necessary and sufficient conditions for single-stage outcomes to be optimal. Section 5.3 characterises an optimal leader-follower outcome when agents payoffs have a threshold property and the comparative statics on hype.

5.1 Perfect coordination

Denote the threshold aggregate action under which agents switch from preferring not to invest to preferring to invest by $\underline{A}(\theta)$. That is, $\underline{A}(\theta) \equiv \sup\{A \in [0, 1] : D(A, \theta) \leq 0\}$ if the set is non-empty, and $\underline{A}(\theta) = 0$ otherwise.

Assumption 1. *For all $\theta \in \Theta$, $v(A, \theta)$ is convex on $[0, \underline{A}(\theta)]$*

Assumption 1 has the following key implication.[15]

[15] A related result is obtained in Morris et al. (2022a), who shows that with sufficiently symmetric agents and smallest equilibrium selection, an asymmetric version of Assumption 1 implies the designer to optimally perfectly coordinate agents' investments.

Lemma 2. *Suppose that Assumption 1 holds. Then, there exists an optimal leader-follower outcome σ^* that satisfies perfect coordination: for all states $\theta \in \Theta$, if $(A_L, A_F) \in \text{supp}(\sigma^*(\cdot|\theta))$, then either $A_L + A_F = 0$ (no agents invest), or $A_L + A_F = 1$ (all agents invest).*

The intuition behind Lemma 2 is as follows. First, suppose the designer's and agent's payoffs are aligned on $((A_L, A_F), \theta)$, so $D(A_L + A_F, \theta) \geq 0$ holds. Then, increasing A_F to $1 - A_L$ benefits the designer, while maintaining followers' incentives to invest. Next, suppose the designer's and agent's payoffs are misaligned, so $D(A_L + A_F, \theta) < 0$ holds. Then, the supermodularity of agents' payoffs and Assumption 1 implies that, respectively, the agents' and designer's payoffs are convex in $A \in [0, \underline{A}(\theta)]$. Hence, a mean-preserving spread of $A_L + A_F$ over the perfectly coordinated pairs $((0,0),\theta), ((A_L, 1), \theta)$ and $((\underline{A}(\theta), 1 - \underline{A}(\theta)), \theta)$ benefits the designer, while maintaining agents' investment incentive.

Lemma 2 reduces the designer's problem to choosing, for each state $\theta \in \Theta$, (i) the probability of all agents investing $q(\theta) \in [0, 1]$, and (ii) conditional on agents investing, the distribution over leaders investing $\sigma_C(\cdot|\theta) \in \Delta([0, 1])$, where $\sigma_C(A_L|\theta)$ is the probability of having A_L leaders and $1 - A_L$ followers invest. Abusing terminology, I simply refer to the pair $(q, \sigma_C) \equiv \{(q(\theta), \sigma_C(\cdot|\theta))\}_{\theta \in \Theta}$ as a (perfectly coordinated) leader-follower outcome moving forward, such that the designer solves.

$$\max_{q, \sigma_C} \int_\Theta v(1, \theta) q(\theta) f(\theta) d\theta \tag{16}$$

$$\text{s.t.} \quad \int_\Theta \int_0^1 A_L D(H, \theta) d\sigma_C(A_L|\theta) q(\theta) f(\theta) d\theta \geq 0 \tag{17}$$

$$\text{and} \quad \int_\Theta \int_0^1 A_L D(A_L, \theta) d\sigma_C(A_L|\theta) q(\theta) f(\theta) d\theta \geq 0 \tag{18}$$

$$\text{and} \quad \int_\Theta \int_0^1 \left(\int_{A_L}^1 D(i, \theta) di \right) d\sigma_C(A_L|\theta) q(\theta) f(\theta) d\theta \geq 0 \tag{19}$$

where (17) - (19) are the leader and follower obedience constraints respectively.

5.2 Single-role benchmarks

I begin by considering the designer's payoff under the benchmark case(s) where the designer assigns all agents who invest to the same role on each state, and provide necessary and sufficient conditions for such cases to be optimal.

Leader-only outcome Call a leader-follower outcome (q, σ_C) *leader-only* if $\sigma_C = \sigma_L$, where σ_L assigns a mass of one to $A_L = 1$ on all states $\theta \in \Theta$, i.e., assigns all agents to the role of a leader. These are important as they are precisely the outcomes implemented by public information structures, and optimal in the extreme case where the designer's preferred equilibrium is selected (which coincides with having $H = 1$). It then follows that Theorem 1 in Section 3 characterises agents' investment behaviours under an *optimal* leader-only outcome (q_L^H, σ_L), where

$$q_L^H(\theta) \equiv \begin{cases} 1, & x_L^H(\theta) \geq \underline{x}_L^H \\ 0, & x_L^H(\theta) < \underline{x}_L^H \end{cases}$$

Follower-only outcome Call a leader-follower outcome (q, σ_C) *follower-only* if $\sigma_C = \sigma_F$, where σ_F assigns a mass of one to $A_L = 0$ on all states $\theta \in \Theta$, i.e., assigns all agents to the role of a follower. These are important as they are precisely the outcomes implemented by "fully-private" information structures, where all agents possess uniform uncertainty about the mass of other agents investing when deciding whether to begin investing, and are optimal in the extreme case where the adversarial equilibrium is selected (which coincides with having $H = 0$). Here, the designer faces the problem of maximizing (16) subject to the single follower-obedience constraint in (19). Thus, following the intuition of the leader-only case, by ordering states $x_F : \Theta \to \Theta$ as follows

$$x_F(\theta) < x_F(\theta') \iff \frac{-\int_0^1 D(i, \theta')di}{v(1, \theta')} \leq \frac{-\int_0^1 D(i, \theta)di}{v(1, \theta)} \tag{20}$$

and letting

$$\underline{x}_F \equiv \min \left\{ x \in \Theta : \int_{\theta : x_F(\theta) \geq x} \int_0^1 D(i, \theta) di f(\theta) d\theta \geq 0 \right\}$$

An optimal follower-only outcome (q_F, σ_F) is defined by

$$q_F(\theta) \equiv \begin{cases} 1, & x_F(\theta) \geq \underline{x}_F \\ 0, & x_F(\theta) < \underline{x}_F \end{cases}$$

Optimality of leader-only and follower-only outcomes Define

$$
\begin{aligned}
\underline{H} &\equiv \sup\{H : F(\{\theta : D(H,\theta) = D(0,\theta)\}) = 1\} \\
\underline{H}^* &\equiv \sup\{H : F(\{\theta : D(H,\theta) = D(0,\theta) \text{ and } x_F(\theta) \geq \underline{x}_F\} | x_F(\theta) \geq \underline{x}_F) = 1\} \\
\overline{H}^* &\equiv \inf\left\{H : \begin{array}{c} \int_\Theta v(1,\theta) q_L^H(\theta) f(\theta) d\theta < V^{\text{FB}}, \quad \text{and} \\ F(\{\theta : D(H,\theta) = D(1,\theta) \text{ and } x_L^H(\theta) \geq \underline{x}_L^H\} | x_L^H(\theta) \geq \underline{x}_L^H) = 1 \end{array}\right\} \\
\overline{H} &\equiv \min\{H : F(\{\theta : D(H,\theta) = D(1,\theta)\} = 1\}
\end{aligned}
$$

where $0 \leq \underline{H} \leq \underline{H}^* \leq \overline{H}^* \leq \overline{H} \leq 1$. $\underline{H}$ captures the largest level of hype under which inducing agents to invest at L1 (as a leader) is harder than inducing agents to invest as a follower. $\underline{H}^*$ has a similar interpretation, but restricted to the support of the optimal follower-only outcome. Meanwhile, $\overline{H}$ and $\overline{H}^*$ have a similar interpretation to their smaller counterparts, but with respect to the smallest level of hype under which inducing investment from a follower is harder than a leader.

Given the preceding definitions, it is intuitive that a follower-only outcome is optimal if $H \leq \underline{H}$, a leader-only outcome is optimal if $H \geq \overline{H}$, and neither of the two are optimal for intermediate H. Theorem 3 below confirms this to be true.

Theorem 3. *Suppose that Assumption 1 holds, and $\int_\Theta v(1,\theta) q_F(\theta) f(\theta) d\theta < V^{FB}$. Then,*

1. *If $H \leq \underline{H}$, then the follower-only outcome is optimal*

2. *If $H \geq \overline{H}$, then the leader-only outcome is optimal*

3. *If $H \in (\underline{H}^*, \overline{H}^*)$, then any optimal leader-follower outcome is neither a follower-only or leader-only outcome.*

Theorem 3 illustrates how changes in hype affect the type of information offered by the designer. By increasing H, the designer moves from using a "fully private" follower-only outcome, i.e., where agents are uniformly uncertain about when others invest, to a hybrid outcome where leaders and followers have different private information, to a leader-only outcome which is implementable by a public information structure. Perhaps notably, the conditions under which extremes of fully private and public information are only optimal are often very stringent. For instance, in the case of Figure 1, a follower-only outcome is optimal if and only if $H = 0$, while a leader-only outcome is optimal if and only if she can induce all agents to invest with probability one.

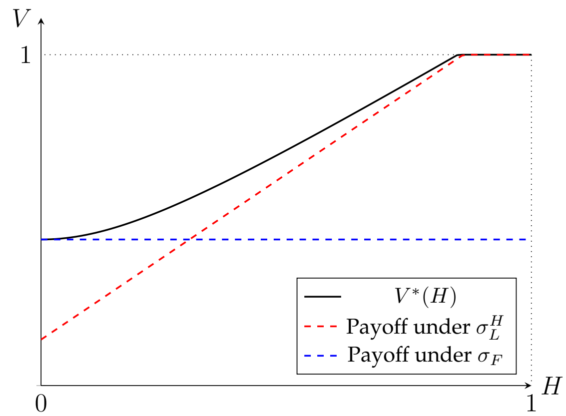


Figure 1: The designer's payoff under (i) an optimal leader-follower outcome $V^*(H)$, (ii) an optimal leader-only outcome and (iii) an optimal follower-only outcome, when F is uniform on $[0,1]$, $D(A,\theta) = -2 + \frac{9}{4}\mathbb{I}_{A \geq \max\{1-\frac{5}{4}\theta, 0\}}$ and $v(A,\theta) = \mathbb{I}_{A \geq \max\{1-\frac{5}{4}\theta, 0\}}$.

5.3 Threshold Games

I now specialize the analysis to the class of *threshold games*, captured by Assumption 2 below, to sharpen the understanding of how hype affects the designer's provision of information.

Assumption 2. *For all $\theta \in \Theta$, $D(A,\theta)$ is constant in $A \in [0, \underline{A}(\theta))$.*

Threshold games capture a general funding setting (e.g. Section 2.3). Agents decide whether to invest into a project. θ captures the agents' common uncertainty about project fundamentals. Launching the project requires enough investment: $A \geq \underline{A}(\theta)$. An agent who invests always incurs an cost of $D(0,\theta) < 0$. Meanwhile, an agent earns a non-negative yield when the project is successfully funded, which can be increasing in A due to positive network effects. These span several games of interest in the literature, one example being the *regime change* game below.

Example 1. *Suppose there exists $c > 0$, and upper semi-continuous functions $\underline{A} : \Theta \to [0,1]$, $B : \Theta \to (c, \infty)$, $W : \Theta \to (0,\infty)$, so that*

$$D(A,\theta) \equiv \begin{cases} B(\theta) - c, & A \geq \underline{A}(\theta) \\ -c, & A < \underline{A}(\theta) \end{cases}, \quad v(A,\theta) \equiv \begin{cases} W(\theta), & A \geq \underline{A}(\theta) \\ 0, & A < \underline{A}(\theta) \end{cases}$$

This captures a canonical regime change setting (e.g. Morris and Shin, 2003), studied recently in Goldstein and Huang (2016), Li et al. (2023), Inostroza and Pavan (2023) and

Taneva and Mathevet (2023). Here, $W(\theta) > 0$ captures the idea that the designer strictly prefers for the regime to be maintained ($A \geq \underline{A}(\theta)$).

I will now characterise an optimal leader-follower outcome in this environment. I show this in two steps: first constructing an upper bound on the designer's payoff across all games using an efficiency argument, and then constructing an upper obedient leader-follower outcome which achieves the upper bound.

An upper bound Given Lemma 2, a relaxation of the designer's problem involves optimizing over leader-follower outcomes subject to ensuring that the *joint* incentives of leaders and followers to begin investing is weakly positive, i.e.,

$$\max_{(q,\sigma_C)} \int_\Theta v(1,\theta)q(\theta)f(\theta)d\theta$$
$$\text{s.t.} \quad \int_\Theta \int_0^1 \Bigg(\underbrace{A_L D(H,\theta) + \int_{A_L}^1 D(i,\theta)di}_{\text{Benefit to } A_L \text{ leaders and } 1-A_L \text{ followers investing on state } \theta \text{ (X)}} \Bigg) d\sigma_C(A_L|\theta)f(\theta)d\theta \geq 0$$

Observe that for all $\theta \in \Theta$, term (X) is maximized at $A_L = H$. That is, conditional on having all agents invest, it is always efficient to have a fraction of H leaders. If so, then by substituting $\sigma_C(\cdot|\theta) = \delta_H$ into the expression above, it follows that the relaxed problem reduces to the public information design problem described by equations (4) - (5) in Section 3, replacing $D(H,\theta)$ with $HD(H,\theta) + \int_H^1 D(i,\theta)di$ in the constraint (5). In turn, reordering states in a similar manner, having all agents invest only on states with a high enough order solves the relaxed problem.

Lemma 3. *Given $H \in [0,1]$, let $x^H : \Theta \to \Theta$ denote the order over states such that*

$$x^H(\theta) < x^H(\theta') \iff \frac{-(HD(H,\theta') + \int_H^1 D(i,\theta')di)}{v(1,\theta')} \leq \frac{-(HD(H,\theta) + \int_H^1 D(i,\theta)di)}{v(1,\theta)} \quad (21)$$

and define

$$\underline{x}^H \equiv \min\left\{x \in \Theta : \underbrace{\int_{\theta: x^H(\theta)\geq x} \left(HD(H,\theta) + \int_H^1 D(i,\theta)di\right) f(\theta)d\theta}_{\equiv \mathcal{D}(x)} \geq 0\right\}$$

Then, an upper bound on the designer's payoff under an optimal leader follower outcome is

$$\overline{V}(H) \equiv \int_{\theta : x^H(\theta) \geq \underline{x}^H} v(1,\theta) f(\theta) d\theta$$

Achieving the upper bound Denote the set of states on which leaders prefer to invest by $\overline{\Theta}^H \equiv \{\theta : D(H,\theta) \geq 0\}$. Notice that under a solution to the relaxed problem, all agents invest on all states $\theta \in \overline{\Theta}^H$. Meanwhile, for all states $\theta \in \Theta \backslash \overline{\Theta}^H$,

$$HD(H,\theta) + \int_H^1 D(i,\theta) di = \underline{A}(\theta) D(0,\theta) + \int_{\underline{A}(\theta)}^1 D(i,\theta) di$$

Hence, the total benefit to agents from investing is

$$\begin{aligned}\mathcal{D}(\underline{x}^H) = & \underbrace{\int_{\overline{\Theta}^H} HD(H,\theta) f(\theta) d\theta}_{\text{(a) Leaders' benefit from investing at L1:}\geq 0} + \underbrace{\int_{\theta : x^H(\theta) \geq \underline{x}^H} \left(\int_{\max\{H, \underline{A}(\theta)\}}^1 D(i,\theta) di \right) f(\theta) d\theta}_{\text{(b) Followers' benefit from investing:}\geq 0} \\ & + \underbrace{\int_{\theta \in \Theta \backslash \overline{\Theta}^H : x^H(\theta) \geq \underline{x}^H} \underline{A}(\theta) D(0,\theta) f(\theta) d\theta}_{\text{Cost of investing (c):}\leq 0} \end{aligned} \tag{22}$$

To construct an optimal leader-follower outcome which mimics agents' behaviours under the solution to the relaxed problem, one must "split" the cost of investment (c) among leaders and followers in a way that satisfies upper introspective obedience. I now show that this can be done. Define

$$\overline{x}^H \equiv \min \left\{ x \in [\underline{x}^H, 1] \backslash x^H(\overline{\Theta}^H) : \begin{array}{c} \int_{\overline{\Theta}^H} HD(H,\theta) f(\theta) d\theta \\ + \int_{\theta \notin \overline{\Theta}^H : x^H(\theta) \in [x,1]} \underline{A}(\theta) D(0,\theta) f(\theta) d\theta \end{array} \geq 0 \right\}$$

and let (q^H, σ_C^H) be defined by

$$q^H(\theta) = \begin{cases} 1, & x^H(\theta) \geq \underline{x}^H \\ 0, & x^H(\theta) < \underline{x}^H \end{cases}, \quad \sigma_C^H(\cdot|\theta) \equiv \begin{cases} \delta_H, & \theta \in \overline{\Theta}^H \\ \delta_{\underline{A}(\theta)}, & x^H(\theta) \in [\overline{x}^H, 1] \backslash x^H(\overline{\Theta}^H) \\ \delta_0, & x^H(\theta) \in [0, \overline{x}^H] \backslash x^H(\overline{\Theta}^H) \end{cases} \tag{23}$$

Then, observe that under (q^H, σ_C^H),

$$\underbrace{\int_{\theta \in \overline{\Theta}^H} HD(H,\theta)f(\theta)d\theta + \int_{\theta \notin \overline{\Theta}^H : x^H(\theta) \in [\overline{x}^H, 1]} \underline{A}(\theta)D(0,\theta)f(\theta)d\theta \geq 0}_{\text{Leader obedience under } (q^H, \sigma_C^H) \text{ from investing at L1 (equation (17))}} \tag{24}$$

$$\underbrace{\int_{\theta \in \overline{\Theta}^H} \int_H^1 D(i,\theta)di f(\theta)d\theta + \int_{\theta \notin \overline{\Theta}^H : x^H(\theta) \in [\underline{x}^H, \overline{x}^H]} \left(\underline{A}(\theta)D(0,\theta) + \int_{\underline{A}(\theta)}^1 D(i,\theta)di \right) f(\theta)d\theta \geq 0}_{\text{Follower obedience under } (q^H, \sigma_C^H) \text{ (equation (19))}} \tag{25}$$

It can also be verified that the second line of leader-obedience under (q^H, σ_C^H), i.e., equation (18), is positive. Hence, (q^H, σ_C^H) satisfies upper-obedience, and so the designer achieves the upper bound. In particular, the optimal leader-follower outcome, after reordering states, has a simple cut-off structure.

Theorem 4. *Suppose that Assumptions 1 - 2 hold. Then, (q^H, σ_C^H) defined in (23) is an optimal leader-follower outcome, under which for all states $\theta \in \Theta$,*

1. *All agents invest on θ if $x^H(\theta) \geq \underline{x}^H$*
2. *No agents invest on θ if $x^H(\theta) < \underline{x}^H$*

Remark 2. *Following the construction above, one finds that for "upper" threshold games, i.e., where for all $\theta \in \Theta$, $D(A, \theta)$ is constant in $A \in [\underline{A}(\theta), 1]$, (i) the upper bound $\overline{V}(H)$ can also be achieved, and (ii) all agents invest under an optimal leader-follower outcome on states satisfying $x^H(\theta) \geq \underline{x}^H$. These include the class of games in Example 1.*

Comparative Statics Theorem 4 pinpoints exactly where hype H affects the optimal disclosure of information. By changing the cost of inducing investment from leaders, changes in H affect the change in the efficient *composition* of leaders and followers to induce investment on a state (which is to have H leaders and $1 - H$ followers). This changes the relative cost of inducing investment across states, i.e., the ranking in (21), which then leads to a change in the subset of states on which agents invest under an optimal information structure.

The above has two implications. First, unless a leader-only outcome is optimal, the ordering in (21) may *not* be the same as the ordering over states under an optimal public information structure (equation (7)). Hence, a designer who has access to private information induces investment across different states than that with public information. Second, as the ordering in (21) can be non-monotone in H, changes in hype can lead to non-monotone affects on agents' aggregate investment. Thus, to deliver monotone comparative statics in hype, I impose a further monotonicity assumption on payoffs to avoid this possibility.

Assumption 3. *For all $A \in [0,1]$, $D(A,\theta)$ and $v(1,\theta)$ are non-decreasing in θ.*

Under Assumption 3, the ordering over states x^H coincides with the value of the state itself (see Figure 2). Hence, letting

$$\theta^H \equiv \min \overline{\Theta}^H, \quad \overline{\theta}^H \equiv \overline{x}^H, \quad \underline{\theta}^H \equiv \underline{x}^H,$$

It follows that $0 \leq \underline{\theta}^H \leq \overline{\theta}^H \leq \theta^H \leq 1$, and so the optimal leader-follower outcome (q^H, σ_C^H) can be written as

$$q^H(\theta) = \begin{cases} 1, & \theta \geq \underline{\theta}^H \\ 0, & \theta < \underline{\theta}^H \end{cases}, \quad \sigma_C^H(\cdot|\theta) \equiv \begin{cases} \delta_H, & \theta \in [\theta^H, 1] \\ \delta_{\underline{A}(\theta)},, & \theta \in [\overline{\theta}^H, \theta^H) \\ \delta_0, & \theta \in [0, \overline{\theta}^H) \end{cases} \tag{26}$$

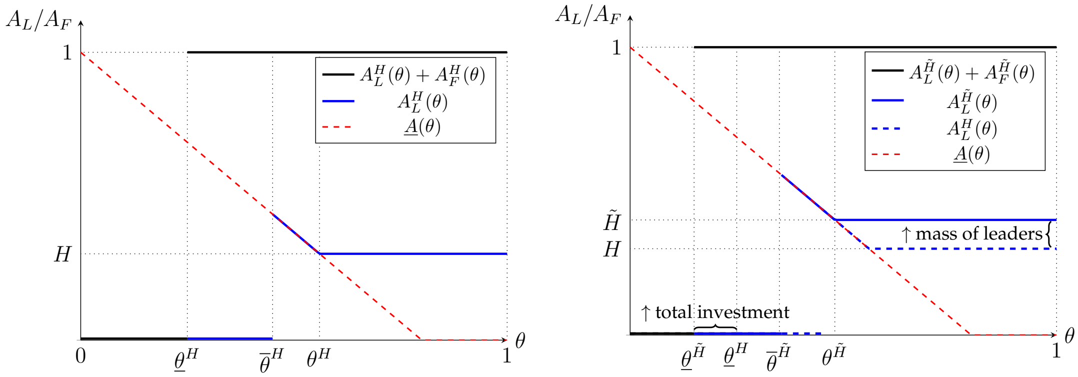
Corollary 3 below states how changes in hype affect (q^H, σ_C^H).

Corollary 3. *Suppose that Assumptions 1 - 3. Then, under the optimal leader-follower outcome (q^H, σ_C^H), an increase in hype H*

1. *Weakly increases the set of states on which agents invest.*

2. *Weakly increases the mass of leaders who invest on all states.*

As Claim 1 is easily verified, I focus on Claim 2. An increase in H decreases the upper threshold θ^H, and so increases the set $\overline{\Theta}^H = [\theta^H, 1]$. This increases the "budget" allocated to leaders (term (a) in (22)), and so allows more of the cost of investment (term (c)) to be allocated to leaders. As such, the middle threshold $\overline{\theta}^H$ falls. Finally, the lower threshold $\underline{\theta}^H$ falls by Claim 1. Hence, by (26), the mass of leaders induced on any state then increases either from 0 to $\underline{A}(\theta)$, or from $\underline{A}(\theta)$ to $H \geq \underline{A}(\theta)$.



(a) Low hype (H)

(b) High hype $(\tilde{H} > H)$

Figure 2: $A_L^H(\theta)$ and $A_F^H(\theta)$ are the mass of leaders and followers, respectively, drawn on state θ under the optimal leader-follower outcome σ_M^H, in the regime change game of Figure 1.

6 Extensions

The main analysis of Sections 4 and 5 extended the public information benchmark of Section 3 by providing the designer with greater flexibility over information structures. This section considers a complementary pursuit: limiting the designer to public information, but enriching the environment studied. I provide a summary of key insights here, leaving additional details to the Online Appendix.

6.1 Endogenizing Hype

I first consider when the level is hype is endogenously determined by the designer. There are two possibilities. First, suppose that the designer only chooses the public information structure. However, conditional on the public signal s realized, the level of hype is $H(q(s))$, where $H(\cdot)$ is exogenously determined, and $q(s) \in \Delta(\Theta)$ is agents' first-order belief given s. This is the case of *indirect hype generation*. It captures, in reduced form, a third-party generating hype in response to information disclosed. Given $(s, q(s))$, agents then play the monotone introspective equilibrium under hype $H(q(s))$.

Second, suppose that the designer commits to both a public information structure, *and* the level of hype H conditional on the realized signal. The cost of generating hype H is $C(H)$, where $C(\cdot)$ is strictly increasing and satisfies $C(0) = 0$. This is the case of *direct hype generation*, which captures an environment where the designer has much control over payoff-irrelevant factors.

Both cases are solved through a similar approach. First, one identifies the set of beliefs under which all agents invest, conditional on the level of hype (in the direct case, this is determined optimally by the designer). Then, one finds that it is sufficient to optimize over *collective* information structures, comprising of a randomization over (i) a "collective" component supported on beliefs under which agents invest, and (ii) a fully informative component which fully discloses the state, on which no agents invest (and in the direct case, the designer generates no hype).

I offer two insights to this problem. First, I provide sufficient conditions under which there exists an optimal "upper censorship" information structure in either case. That is, the collective component comprises of a single belief involving all high enough states, which is identified via an "efficiency argument" as in Section 3. In the indirect case, this holds when (i) the designer's payoff and agents' payoffs are monotone in states and $H(q)$ is non-decreasing with respect to first-order stochastic dominance in beliefs q, and (ii) agents' payoff is concave in the aggregate action and linear in the state, and $H(q)$ is concave in beliefs. In the direct case, this holds when (i) the designer's payoff and agents' payoffs are monotone in states, and (ii) the agents' payoff is concave in actions and linear in the state and the cost of generating hype $C(H)$ is convex in H.

Second, the extent of control the designer has over hype can affect the multiplicity of optimal signals. For example, suppose an agent's payoff is strictly concave in the aggregate action and linear in the state, and (i) in the indirect case, hype $H(q)$ is concave in beliefs, or (ii) in the direct case, the cost of generating hype $C(H)$ is convex in H. Then, there exists an collective information structure which *pools* agents' signals which induce investment. Whether such a pooling property is *necessary* for optimality differs across cases. In the indirect case, all optimal information structures have the pooling property. In contrast, in the direct case, there always exists an optimal information structure which is *pairwise*, i.e., where the support of any belief under which agents invest contains at most two states.

6.2 The Bandwagon Effect of Hype

Next, I consider the environment of Inostroza and Pavan (2023), by allowing agents to receive private information which cannot be manipulated.

Suppose that after the state θ and public signal s are drawn, each agent observes a private signal $t \in T \subset \mathbb{R}$. Private signals are independently drawn across agents according to $\rho(\cdot|\theta) \in \Delta(T)$, which has full support and is log-supermodular in (t, θ). I assume throughout that Assumption 2 holds, $v(A, \theta) = 0$ holds for all $A \in [0, \underline{A}(\theta)]$, and that the designer cannot elicit agents' private signals. The definition of a monotone introspective equilibrium is extended such that agents' strategies now depend both on public and private signals observed.

Allowing for private information does not qualitatively affect optimal information disclosure, regardless of the level of hype. Similar to no private information, there exists an optimal "upper censorship" information structure which pools agents' recommendations to invest, while fully disclosing states on which (all) agents do not invest. Additionally, the states on which agents invest/do not invest can be identified via an efficiency argument analogous to Section 3.

Meanwhile, private information affect the mechanics behind how investment is achieved. As agents can now be asymmetric prior to investment, unless hype is sufficiently high, the designer's optimal information structure will induce a mix of leaders (with higher private signals) and followers (with lower private signals).

Additionally, the designer benefits from an increase in hype, not only by raising leaders' incentives to invest, but also through the corresponding *bandwagon effect* on agents. To see this, taking any (public) signal realized s, notice that the smallest type of agent which invests at L1 is

$$\underline{t}^H(s) \equiv \min\{t \in T : \int_\Theta D(H, \theta)\rho(t|\theta)d\pi(\theta|s) \geq 0\}$$

which is *decreasing* in H. That is, agents previously too pessimistic to invest early, due to their private information, are now "won over" by the increase in hype. Such agents now mimic their higher-signal counterparts, and invest earlier in the introspection process. In doing so, they provide greater assurance to all other followers to invest, leading to a further bandwagoning effect where all followers have stronger incentives to invest. The designer capitalizes on this by then pooling more states into agents' (public) recommendation to invest.

Note that the bandwagon effect identified here is distinct to that for informational cascades (Bikhchandani et al., 1992, Banerjee, 1992). In those models, sequentially arriving agents, who observe prior agents actions, mimic prior agents' behaviour under the belief that prior agents are collectively more informed then the agent himself. Here, the bandwagon effect arises due to the agent's greater optimism about coordinating her investments with others, driven by higher hype.

6.3 Uncertainty about Hype

Finally, I consider when the designer must commit to an information structure in advance of the realization of hype (which may be uncertain). Formally, suppose now that the designer first chooses the public information structure $\mathcal{S}$. After the signal and state are drawn, a level of hype $H \in [0, 1]$ is drawn according to a distribution $G(\cdot|\theta) \in \Delta[0, 1]$ with strictly positive density $g(\cdot|\theta) > 0$. Agents then play the monotone introspective equilibrium corresponding to H.

My main observation is that under public information, "equilibrium uncertainty" reduces to "individual uncertainty". Recall by Section 3 that under public information, agents monotone introspective equilibrium behaviours only depend on hype H, and first-order beliefs. Thus, if H is known, then the designer's problem is isomorphic to a single-agent problem, where the designer's payoff from the agent investing and not investing are $v(1, \theta)$ and $v(0, \theta)$ respectively, and the agent invests if and only if given signal s, $\int_\Theta D(H, \theta) d\pi(\theta|s) \geq 0$ holds.[16] Likewise, when H is "privately" observed by agents, then the designer's problem is isomorphic to the problem of persuading a single-agent with non-elicitable private information.

The preceding connection allows one to use results from the rich *Bayesian Persuasion* literature (Kamenica and Gentzkow, 2011) to study the problem with equilibrium uncertainty. Here, a key question that can be addressed using existing tools is of the *assortative properties* of an optimal information structure. That is, connecting the values of H drawn under which agents invest, to the state θ. To do so, following Kolotilin (2018), the designer's problem is (also) isomorphic to a single-agent problem without private information, where the agent's "action" is $H \in [0, 1]$, and the designer's and agent's payoffs on (H, θ) are, respectively,

$$V(H, \theta) \equiv \int_H^1 v(1, \theta) g(H'|\theta) dH', \quad U(H, \theta) \equiv \int_H^1 D(H', \theta) g(H'|\theta) dH'$$

Under certain parametric restrictions, the problem above can be addressed using

[16] Thus, Theorem 1 characterises the optimal information structure for the single-agent game.

existing results. First, when G is state-independent and $D(A,\theta)$ is linear in θ, so the agent's optimal action only depends on the posterior mean (e.g., Gentzkow and Kamenica, 2016, Dworczak and Martini, 2019, Kolotilin et al., 2022). Second, when D is smooth in (A,θ), and $D(0,\theta) \leq 0 \leq D(1,\theta)$ holds, so the agent's optimal action is pinned down by a *first-order condition* (Kolotilin, 2018, Kolotilin et al., 2023).[17] I illustrate this observation with a simple example.

Example 2. *Suppose $D(A,\theta) = A - \theta$ and $v(A,\theta) = A$ (e.g. the investment game of Bergemann and Morris (2016)), F and $G(\cdot|\theta) = G(\cdot)$ have full support and densities f and g respectively, and g is log-concave. Then, a lower-censorship information structure is optimal (Theorem 2, Kolotilin et al., 2022). Thus, under an optimal information structure, an agent invests if and only if (i) the state favours investment enough, i.e., is not too large, and (ii) the realized level of hype is high enough. Hence, there is a (weak) positive assortment between maximum adversariality and the state.*

Appendix A: Further Details

Every Monotone Introspective Equilibrium is a Bayes-Nash equilibrium

Theorem 5. *For all information structures $\mathcal{S}$ and levels of hype $h \in [0,1]$, if $\alpha^{\mathcal{S},h}$ is a monotone introspective equilibrium of $\mathcal{S}$, then it is a Bayes-Nash equilibrium of $\mathcal{S}$*

Proof of Theorem 5 Take any information structure $\mathcal{S}$, and suppose that $\alpha^{\mathcal{S},h}$ is a monotone introspective equilibrium of $\mathcal{S}$. Fix any signal observed $s \in S$. I focus on the case where $\alpha^{\mathcal{S},h}(s) = 1$, noting that the case with $\alpha^{\mathcal{S},h}(s) = 0$ is proven similarly. Since the agent invests in the introspective equilibrium, there exists a $\overline{k} \geq 2$ such that for all $k \geq \overline{k}$, $\alpha^{\mathcal{S},h,k}(s) = 1$. That is,

$$\int_{\Delta(S)\times\Theta} D(A(\alpha^{\mathcal{S},h,k-1}|\mu),\theta)d\pi(\mu,\theta|s) \geq 0$$

Now, notice that $(D(A(\alpha^{\mathcal{S},h,k-1}|\mu),\theta))_{k\geq\overline{k}}$ is a sequence of measurable functions defined on $\Delta(S)\times\Theta$, which converges monotonically point-wise to $D(A(\alpha^{\mathcal{S},h}|\mu),\theta)$, and is bounded above and below by the integrable functions $D(1,\theta)$ and $D(0,\theta)$

[17] For a multi-agent application of the first-order approach, see Smolin and Yamashita (2023).

respectively. Thus, by the Dominated Convergence Theorem,

$$\int_{\Delta(S)\times\Theta} D(A(\alpha^{\mathcal{S},h}|\mu),\theta)d\pi(\mu,\theta|s) = \lim_{k\to\infty}\int_{\Delta(S)\times\Theta} D(A(\alpha^{\mathcal{S},h,k-1}|\mu),\theta)d\pi(\mu,\theta|s) \geq 0$$

and so, $\alpha^{\mathcal{S},h}(s) = 1$ is a best-response for the agent observing signal s. □

Lemma 4 (Contagion Lemma). *Take any information structure $\mathcal{S}$, level of hype h, and any $k > 1$. If every agent who invests at L$k-1$ also invests at Lk, i.e., the following holds*

$$\alpha^{\mathcal{S},h,k-1}(s) = 1 \Rightarrow \alpha^{\mathcal{S},h,k}(s) = 1, \quad \forall s \in S \tag{27}$$

then the sequence $(\alpha^{\mathcal{S},h,k'})_{k'\geq k}$ is (point-wise) non-decreasing in k'.

Proof of Lemma 4 I will prove that (27) implies the following claim (which yields the Lemma): for all $k' \geq k$ and $s \in S$, $\alpha^{\mathcal{S},h,k'-1}(s) = 1$ implies $\alpha^{\mathcal{S},h,k'}(s) = 1$. I proceed by induction. The base case for $k' = k$ holds by (27). Meanwhile, suppose that the induction hypothesis holds for all $\tilde{k} < k'$ for some $k' > k$. Take any agent and signal $s \in S$, and suppose that $\alpha^{\mathcal{S},h,k'-1}(s) = 1$. Then, since $\alpha^{\mathcal{S},h,k'-2}(s) \leq \alpha^{\mathcal{S},h,k'-1}(s)$ for all signals $s \in S$, $D(A(\alpha^{\mathcal{S},h,k'-2}|\mu),\theta) \leq D(A(\alpha^{\mathcal{S},h,k'-1}|\mu),\theta)$ holds. Hence,

$$\int_{\Delta(S)\times\Theta} D(A(\alpha^{\mathcal{S},h,k'-1}|\mu),\theta)d\pi(\mu,\theta|s) \geq \int_{\Delta(S)\times\Theta} D(A(\alpha^{\mathcal{S},h,k'-2}|\mu),\theta)d\pi(\mu,\theta|s) \geq 0$$

Thus, agent i invests at Lk' under s, so the induction hypothesis holds. □

Appendix B: Main Text Proofs

Sections B1 and B2 prove the sufficiency and necessity components, respectively, of Theorem 2. Section B3 contains the remainder of the proofs.

B1: Proof of Sufficiency: Theorem 2

Let $\overline{\sigma}$ denote the leader-follower outcome associated with the full disclosure information structure, i.e., such that $\overline{\sigma}(\cdot|\theta) = \delta_{(1,0)}$ whenever $D(H,\theta) \geq 0$, and

$\overline{\sigma}(\cdot|\theta) = \delta_{(0,0)}$ whenever $D(H,\theta) < 0$. Since $\int_{\overline{\Theta}} D(0,\theta)f(\theta)d\theta > 0$, $\overline{\sigma}$ satisfies introspective obedience strictly, i.e., where the associated inequalities are all strict.

Now, given a leader-follower outcome σ, define $\sigma^\epsilon \equiv (1-\epsilon)\sigma + \epsilon\overline{\sigma}$ for all $\epsilon \in (0,1]$. Notice that if σ satisfies introspective obedience, then σ^ϵ satisfies introspective obedience strictly. Our main goal is to prove the following.

Lemma 5. *Take any leader-follower outcome σ which satisfies introspective obedience under H. Then, for all $\tilde{\epsilon} \in (0,1]$, there exists a H-implementable outcome $\tilde{\sigma}$ under which $p_{\tilde{\sigma}} = p_{\sigma^\epsilon}$.*

Before proving Lemma 5, let me discuss how it can be applied to prove sufficiency. Taking any leader-follower outcome σ which satisfies Bayes-plausibility introspective obedience, consider the sequence $(\sigma^{1/n})_{n=1}^\infty$ and $(\tilde{\sigma}^{1/n})_{n=1}^\infty$. Since $p_{\sigma^{1/n}} = p_{\tilde{\sigma}}$ for all $n \geq 1$, and $\lim_{n\to\infty} p_{\sigma^{1/n}} = p_\sigma$ by construction of $\sigma^{1/n}$ and the properties of the push-forward, $\lim_{n\to\infty} p_{\tilde{\sigma}^{1/n}} = p_\sigma$ holds. This yields Claim 2 in Theorem 2.

Proof of Lemma 5 Take any leader-follower outcome σ which satisfies introspective obedience, and any $\tilde{\epsilon} > 0$. Throughout, I write σ rather than $\sigma^{\tilde{\epsilon}}$ to simplify exposition. I break the proof into four parts. Part I establishes two useful observations. Part II constructs the candidate information structure for the problem. Part III proves several properties of agents' strategies under the candidate information structure. Part IV applies the results of Parts I-III, to prove that there exists a monotone introspective equilibrium under the candidate information structure constructed, and the corresponding outcome implemented is equal to p_σ.

Part I: Preliminary Definitions Given an $\epsilon \in (0,\tilde{\epsilon})$, let $\tilde{\sigma}$ be defined by setting $\tilde{\sigma}(\cdot|\theta) = \sigma$ for all $\theta \in \Theta\backslash\overline{\Theta}$, while $\tilde{\sigma}(\cdot|\theta) = \frac{\sigma(\cdot|\theta)-\epsilon\delta_{(0,1)}}{1-\epsilon}$ for all $\theta \in \overline{\Theta}$, where $\epsilon > 0$ sufficiently small such that all Leader-obedience and Downwards-obedience constraints continue to hold under $\tilde{\sigma}$: this is possible as all constraints are assumed to be strict. There are two properties of $\tilde{\sigma}$ relevant to the task:

Property 1. *For each $n \in \mathbb{N}$, define*

$$\tilde{D}_n((A_L, A_F), \theta) \equiv \frac{1}{n}\Bigg[(\lceil nA_L \rceil - nA_L)D(A_L, \theta) + \sum_{i=\lceil nA_L \rceil}^{\lfloor n(A_F+A_L) \rfloor - 1} D(\frac{i}{n}, \theta) + [n(A_F + A_L) - \lfloor n(A_F + A_L) \rfloor] D(\frac{\lfloor n(A_F + A_L) \rfloor}{n}, \theta)\Bigg]$$

where the first term does not appear if $A_L = 0$, and the last term does not appear if $A_L + A_F = 1$. Then, there exists a sufficiently large $N \geq 1$ such that

$$\int_\Theta \int_\mathcal{A} \tilde{D}_N((A_L, A_F), \theta) d\tilde{\sigma}(A_L, A_F|\theta) f(\theta) d\theta > 0 \tag{28}$$

Proof. Fixing a triple $((A_L, A_F), \theta)$, notice that $\mathbb{I}_{i \in [A_L, A_L + A_F]} D(i, \theta)$ is piecewise-monotone. Therefore, it is Riemann-integrable, such that its integral over $[0, 1]$ can be approximated by functions of the form of $\tilde{D}_n((A_L, A_F), \theta)$, i.e.,

$$\int_0^{A_F} D(A_L + i, \theta) di = \int_0^1 \mathbb{I}_{[A_L, A_L + A_F]} D(i, \theta) di = \lim_{n \to \infty} \tilde{D}_n((A_L, A_F), \theta)$$

Combined with the fact that

$$\int_\Theta \int_\mathcal{A} \int_0^{A_f} D(A_L + i, \theta) di d\tilde{\sigma}(A_L, A_F|\theta) f(\theta) d\theta = \int_\Theta \int_\mathcal{A} \int_0^{A_f} D(A_L + i, \theta) di d\sigma(A_L, A_F|\theta) f(\theta) d\theta > 0$$

and, by the Dominated convergence theorem,

$$\lim_{n \to \infty} \int_\Theta \int_\mathcal{A} \tilde{D}_n((A_L, A_F), \theta) d\tilde{\sigma}(A_L, A_F|\theta) f(\theta) d\theta = \int_\Theta \int_\mathcal{A} \int_0^{A_f} D(A_L + i, \theta) di d\tilde{\sigma}(A_L, A_F|\theta) f(\theta) d\theta$$

there must then exist $N \geq 1$ large enough such that (28) holds. □

Property 2. *There exists $\eta \in (0, 1)$ sufficiently small such that*

$$\frac{\epsilon}{N - 1} \int_{\overline{\Theta}} D(0, \theta) f(\theta) dF\theta + \eta \int_{\Theta \setminus \overline{\Theta}} D(0, \theta) f(\theta) d\theta \geq 0 \tag{29}$$

and

$$\int_{\Theta}\int_{\mathcal{A}}\frac{1}{N}\left[\begin{array}{c}(1-\eta)^{-\lceil NA_L\rceil}(\lceil NA_L\rceil-NA_L)D(A_L,\theta)\\+\sum_{i=\lceil NA_L\rceil}^{\lfloor N(A_F+A_L)\rfloor-1}(1-\eta)^{-i-1}D(\frac{i}{N},\theta)\\+(1-\eta)^{-\lfloor N(A_F+A_L)\rfloor-1}[N(A_F+A_L)-\lfloor N(A_F+A_L)\rfloor]D(\frac{\lfloor N(A_F+A_L)\rfloor}{N},\theta)\end{array}\right]d\tilde{\sigma}(A_L,A_F|\theta)f(\theta)d\theta\geq 0 \tag{30}$$

Proof. Follows from Property 1 and the fact that $\int_{\overline{\Theta}} D(0,\theta)f(\theta)d\theta > 0$ holds. □

Part 2: Constructing the information structure Consider the information structure $\mathcal{S}$ with $S = \mathbb{Z}_+ \cup \{\infty\}$, and $\pi(\cdot|\theta)$ obtained from the following procedure

1. **Step 1:** Draw (A_L, A_F) according to $\sigma(\cdot|\theta)$
2. **Step 2:** Independently, draw a $z \in \mathbb{Z}_+$ with probability $\eta(1-\eta)^z$, and $y \in \{0,1\}$ with probability ϵ.
3. **Step 3a:** If $\theta \in \overline{\Theta}$, $(A_L, A_F) = (1,0)$ and $y = 1$, then draw $\mu_1 \in \Delta(S)$, where

$$\mu_1(s) \equiv \begin{cases}\frac{1}{N-1}, & s \in \{1,...,N-1\}\\ 0, & \text{otherwise}\end{cases}$$

4. **Step 3b:** If $\theta \notin \overline{\Theta}$, and either $(A_L, A_F) \neq (1,0)$ or $y \neq 1$, then draw $\mu_{(z,A_L,A_F)} \in \Delta(S)$, where

$$\mu_{(z,A_L,A_F)}(s) \equiv \begin{cases}A_L, & s = 0\\ \frac{\lceil NA_L\rceil}{N} - A_L, & s = \lceil NA_L\rceil + z\\ \frac{1}{N}, & \lceil NA_L\rceil + z < s \leq \lfloor N(A_L+A_F)\rfloor + z\\ A_L + A_F - \frac{\lfloor N(A_L+A_F)\rfloor}{N}, & s = \lfloor N(A_L+A_F)\rfloor + z + 1\\ 1 - A_L - A_F, & s = \infty\\ 0, & \text{otherwise}\end{cases}$$

Observe that by the construction above, conditional on drawing (A_L, A_F), the measure of agents who observe $s < \infty$ is, with probability one, $A_L + A_F$. Hence, letting $\tilde{p}$ denote the outcome which describes agents' state-wise behaviours whenever they invest if and only if they observe a signal $s < \infty$, $\tilde{p} = p_\sigma$.

Part 3: Properties of the information structure To prove that $\tilde{p}$ is indeed the outcome implemented by $\mathcal{S}$, I first prove five properties about $\mathcal{S}$.

Claim 1: An agent invest at all levels under signal $s = 0$

Proof. Consider when an agent observes $s = 0$. By Part 2, conditional on (A_L, A_F) being drawn on state θ, an agent has probability of $(1-\epsilon)A_L$ of observing 0 (which coincides with the mass of other agents who have observed the same signal) if $\theta \in \overline{\Theta}$ and $(A_L, A_F) = (1, 0)$, and A_L if $\theta \notin \overline{\Theta}$. Thus, the unconditional payoff of the agent from investing at L1 is simply the LHS of (11) under $\tilde{\sigma}$, which is positive. Furthermore, provided that (at least) a measure of A_L other agents are investing at L1 whenever an agent observes a signal of 0, the unconditional payoff of the agent from investing at L2 is at least (12) under $\tilde{\sigma}$, which is positive. From here, a straightforward induction argument implies that all agents who observe a signal of 0 invest at all levels □

Claim 2: An agent invests at all levels under signals $s \in \{1, ..., N-1\}$

Proof. Take any signal $s \in \{1, ..., N-1\}$. Conditional on (A_L, A_F) drawn on state θ, Part 2 implies that on any state $\theta \in \overline{\Theta}$, the agent observes signal s with probability of at least $\frac{\epsilon}{N-1}$, i.e., when $y = 1$ is drawn. Meanwhile, if $\theta \notin \overline{\Theta}$, the probability that signal $s = 0$ is observed is at most η. Hence, the agent's unconditional expected payoff from investing at any level is at least (29), which is positive. Therefore, any such agent invests at all levels. □

Claim 3: All agents do not invest at all levels under signal ∞.

Proof. I prove this by inducting on k. Consider the base case of $k = 1$. Fix (A_L, A_F) drawn on state θ. By Part 2, the probability that the agent observes $s = \infty$ is exactly $(1-\epsilon)(1 - A_L - A_F)$ if $\theta \in \overline{\Theta}$ and $(A_L, A_F) = (1, 0)$, and $(1 - A_L - A_F)$ if $\theta \notin \overline{\Theta}$. Therefore, the agent's unconditional payoff from investing at L1 is (14) under $\tilde{\sigma}$, which is strictly negative, so the agent does not invest at L1. Now, suppose that the induction hypothesis is true for all levels $k' < k$ for some $k \geq 2$. Consider the agent's incentive to invest at level k. By the induction hypothesis, the upper bound on the measure of agents who invest at Lk-1 conditional on (A_L, A_F) being drawn exactly $1 - A_L - A_F$. Consequently, the agent's unconditional payoff from investing at Lk is at most the LHS of (15) under $\tilde{\sigma}$, which is strictly negative. Hence, the agent does not invest at Lk, which proves the induction hypothesis. □

Claim 4: For all signals $s \in \{1, 2, 3, ..., \}$ observed, all agents take action 1 at all levels greater than or equal to s.

Proof. I prove the claim by induction on $s \in \{1, 2, ..., \}$. Claim 2 implies that the base case, i.e., for $s \leq N-1$ holds. Now, suppose that the induction hypothesis holds for all $s < k$ for some $k \geq N$. Fix $s = k$ being observed by an agent when $((A_L, A_F), \theta)$ and $z \in \mathbb{Z}_+$ is drawn. Then, a mass of at least $\max\{\frac{k-z-1}{N}, A_L\}$ other agents must have observed a signal strictly smaller k, and so are investing at Lk-1. Hence, the agent's payoff from investment is (at least) $D(\max\{\frac{k-z-1}{N}, A_L\}, \theta)$.

Consider then the probability of observing k when $((A_L, A_F), \theta)$ is drawn. If $z = k - \lceil N(A_L + A_F)\rceil$ is drawn, then the agent observes k with probability $\frac{1}{N}(N(A_L + A_F) - \lceil N(A_L + A_F)\rceil)$. If $k - \lceil N(A_L + A_F)\rceil < z < k - \lceil NA_L \rceil$, then the agent observes k with probability $\frac{1}{N}$. Finally, if $z = k - \lceil NA_L\rceil$, then the agent observes k with probability $\frac{1}{N}(\lceil NA_L\rceil - A_L)$. Thus, conditional on $((A_L, A_F), \theta)$ being drawn, the agent's unconditional payoff from investing is

$$\frac{(1-\epsilon)}{N}\left[\begin{array}{c} \eta(1-\eta)^{k-\lceil N(A_L+A_F)\rceil}(\lceil NA_L\rceil - NA_L)D(A_L,\theta) \\ +\sum_{z=k-\lceil NA_L\rceil-1}^{k-\lfloor N(A_L+A_F)\rfloor}(1-\eta)^z D(\frac{n-z-1}{N},\theta) \\ +(1-\eta)^{k-\lceil N(A_L+A_F)\rceil}[N(A_F+A_L) - \lfloor N(A_F+A_L)\rfloor]D(\frac{\lfloor N(A_F+A_L)\rfloor}{N},\theta) \end{array}\right]$$

$$= \frac{(1-\epsilon)\eta(1-\eta)^k}{N}\left[\begin{array}{c} (1-\eta)^{-\lceil NA_L\rceil}(\lceil NA_L\rceil - NA_L)D(A_L,\theta) \\ +\sum_{i=\lceil NA_L\rceil}^{\lfloor N(A_F+A_L)\rfloor-1}(1-\eta)^{-i-1}D(\frac{i}{N},\theta) \\ +(1-\eta)^{-\lfloor N(A_F+A_L)\rfloor-1}[N(A_F+A_L) - \lfloor N(A_F+A_L)\rfloor]D(\frac{\lfloor N(A_F+A_L)\rfloor}{N},\theta) \end{array}\right]$$

Integrating over all triples $((A_L, A_F), \theta)$ then yields (30), which is positive. Therefore, the agent invests at Lk upon observing k. From here, noticing that the induction hypothesis implies at least $k - z - 1$ agents invest at L$k' - 1$ for all $k' \geq k$, the agent invests at all levels $k' \geq k$ upon observing k. □

Claim 5: If an agent invests at L1 under some $N \leq s < \infty$, then for all such signals, the agent invests at all levels L1+.

Proof. First, following the logic of Claim 4, observe that an agent's payoff from investing at L1 under a signal $N \leq s < \infty$,

$$(1-\epsilon)\frac{\eta(1-\eta)^s}{N}\int_\Theta\int_\mathcal{A}\left[\begin{array}{c}(1-\eta)^{-\lceil NA_L\rceil}(\lceil NA_L\rceil - NA_L)D(h,\theta)\\ +\sum_{i=\lceil NA_L\rceil}^{\lfloor N(A_F+A_L)\rfloor-1}(1-\eta)^{-i-1}D(h,\theta)\\ +(1-\eta)^{-\lfloor N(A_F+A_L)\rfloor-1}[N(A_F+A_L)-\lfloor N(A_F+A_L)\rfloor]D(h,\theta)\end{array}\right]d\tilde{\sigma}(A_L,A_F|\theta)f(\theta)d\theta \tag{31}$$

Now, suppose that an agent invests at L1 under one such signal, so (31) is positive. Notice that the sign of (31) does not depend on s, so agents must be investing at L1 under all signals $N \leq s < \infty$. Combined with Claims 1-4, this implies an agent invests if and only if he observes $s < \infty$.

Next, consider an agent's investment decision at L2 upon observing some $s \in \{N, N+1, ..., \}$ drawn. Since the agent anticipates a mass of $A_L + A_F$ investing when (A_L, A_F) is drawn, the agent's unconditional expected payoff from investing at L2 is given by

$$(1-\epsilon)\frac{\eta(1-\eta)^s}{N}\int_\Theta\int_\mathcal{A}\left[\begin{array}{c}(1-\eta)^{-\lceil NA_L\rceil}(\lceil NA_L\rceil - NA_L)D(A_L+A_F,\theta)\\ +\sum_{i=\lceil NA_L\rceil}^{\lfloor N(A_F+A_L)\rfloor-1}(1-\eta)^{-i-1}D(A_L+A_F,\theta)\\ +(1-\eta)^{-\lfloor N(A_F+A_L)\rfloor-1}[N(A_F+A_L)-\lfloor N(A_F+A_L)\rfloor]D(A_L+A_F,\theta)\end{array}\right]d\tilde{\sigma}(A_L,A_F|\theta)f(\theta)d\theta$$

$$\geq (1-\epsilon)\frac{\eta(1-\eta)^s}{N}\int_\Theta\int_\mathcal{A}\left[\begin{array}{c}(1-\eta)^{-\lceil NA_L\rceil}(\lceil NA_L\rceil - NA_L)D(A_L,\theta)\\ +\sum_{i=\lceil NA_L\rceil}^{\lfloor N(A_F+A_L)\rfloor-1}(1-\eta)^{-i-1}D(\frac{i}{N},\theta)\\ +(1-\eta)^{-\lfloor N(A_F+A_L)\rfloor-1}[N(A_F+A_L)-\lfloor N(A_F+A_L)\rfloor]D(\frac{\lfloor N(A_F+A_L)\rfloor}{N},\theta)\end{array}\right]d\tilde{\sigma}(A_L,A_F|\theta)f(\theta)d\theta \geq 0$$

Hence, the agent invests at L2. A straightforward induction argument then implies any such agent invests at all levels 2+. □

Part 4: Proving Lemma 5 I now prove that the information structure constructed in Part 2, $\mathcal{S}$, has a monotone introspective equilibrium, and in the monotone introspective equilibrium, an agent invests if and only if $s < \infty$. As noted at the end of Part 2, this would imply that $\tilde{p} = p_\sigma$, so Lemma 5 holds.

Here, there are two possibilities. First, suppose that there exists $s \geq N$ who invests at L1. By Claims 1, 2 and 5, all agents $s < \infty$ invest at all levels, and all agents $s = \infty$ do not invest at all levels. Further notice that agents' best-responses are non-decreasing, so the behaviour described above is indeed a monotone introspective equilibrium of $\mathcal{S}$.

Next, suppose that for all $s \geq N$, the agent does not invest at L1. By Claims 1-5, an agent invests at L1 if and only if $s \leq N-1$, and all such agents invest at L2 (and at all higher levels). Hence, by Lemma 4, $\mathcal{S}$ has a monotone introspective equilibrium. Applying Claims 1-5 then implies that in such an equilibrium, an agent invests if and only if $s < \infty$. $\square$

B2: Proof of Necessity: Theorem 2

Take any H-implementable leader-follower outcome σ, and let $\mathcal{S}$ be any information structure which implements it, where the index of the largest monotone introspective equilibrium subject to the upper bound H is denoted by h.

Leader obedience Let $S^1 \equiv \{s \in S : \alpha^{\mathcal{S},h,1}(s) = 1\}$. Then,

$$\begin{aligned}
\int_\Theta \int_\mathcal{A} A_L D(H,\theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta &\geq \int_\Theta \int_\mathcal{A} A_L D(h,\theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \\
&= \int_\Theta \int_{\Delta(S)} \mu(\{s \in S : \alpha^{\mathcal{S},h,1}(s) = 1\}) D(h,\theta) d\pi(\mu|\theta) f(\theta) d\theta \\
&= \int_\Theta \int_{\Delta(S)} \int_{S^1} D(h,\theta) d\mu(s) d\pi(\mu|\theta) f(\theta) d\theta \\
&= \int_{S^1} \underbrace{\int_{\Delta(S)\times\Theta} D(h,\theta) d\pi(\mu,\theta|s)}_{\geq 0 \text{ by } (2)} d\pi_S(s) \geq 0
\end{aligned}$$

Meanwhile,

$$\begin{aligned}
\int_\Theta \int_\mathcal{A} A_L D(A_L,\theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta &= \int_\Theta \int_{\Delta(S)} \mu(\{s \in S : \alpha^{\mathcal{S},h,1}(s) = 1\}) D(A^{\mathcal{S},h,1}(\mu),\theta) d\pi(\mu|\theta) f(\theta) d\theta \\
&= \int_\Theta \int_{\Delta(S)} \int_{S^1} D(A^{\mathcal{S},h,1}(\mu),\theta) d\mu(s) d\pi(\mu|\theta) f(\theta) d\theta \\
&= \int_{S^1} \underbrace{\int_{\Delta(S)\times\Theta} D(A^{\mathcal{S},h,1}(\mu),\theta) d\pi(\mu,\theta|s)}_{\geq 0 \text{ by } (3)} d\pi_S(s) \geq 0
\end{aligned}$$

Hence, σ satisfies leader obedience. $\square$

Follower obedience For each $k > 1$, defining $S^k \equiv \{s \in S : \alpha^{\mathcal{S},h,k}(s) = 1 > 0 = \alpha^{\mathcal{S},h,k-1}(s)\}$, and notice that

$$\begin{aligned}\int_{A_L^{\mathcal{S},h}(\mu)}^{A_L^{\mathcal{S},h}(\mu)+A_F^{\mathcal{S},h}(\mu)} D(i,\theta)di &= \sum_{k=2}^{\infty}\int_{A(\alpha^{\mathcal{S},h,k-1}|\mu)}^{A(\alpha^{\mathcal{S},h,k}|\mu)} D(i,\theta)di \\ &\geq \sum_{k=2}^{\infty} A(\alpha^{\mathcal{S},h,k-1}|\mu)D(A(\alpha^{\mathcal{S},h,k-1}|\mu),\theta) \\ &= \sum_{k=2}^{\infty}\int_{S^k} D(A(\alpha^{\mathcal{S},h,k-1}|\mu),\theta)d\mu(s)\end{aligned}$$

Therefore,

$$\begin{aligned}\int_{\Theta}\int_{\mathcal{A}}\left(\int_{A_L}^{A_L+A_F} D(i,\theta)di\right)d\sigma(A_L,A_F|\theta)f(\theta)d\theta &= \int_{\Theta}\int_{\Delta(S)}\int_{A_L^{\mathcal{S},h}(\mu)}^{A_L^{\mathcal{S},h}(\mu)+A_F^{\mathcal{S},h}(\mu)} D(i,\theta)did\pi(\mu|\theta)f(\theta)d\theta \\ &\geq \int_{\Theta}\int_{\Delta(S)}\sum_{k=2}^{\infty}\int_{S^k} D(A(\alpha^{\mathcal{S},h,k-1}|\mu),\theta)d\mu(s)d\pi(\mu|\theta)f(\theta)d\theta \\ &\geq \sum_{k=2}^{\infty}\int_{S^k}\Bigg(\underbrace{\int_{\Delta(S)\times\Theta} D(A(\alpha^{\mathcal{S},h,k-1}|\mu),\theta)d\pi(\mu,\theta|s)}_{\geq 0 \text{ by } (3)}\Bigg)d\pi_S(s) \geq 0\end{aligned}$$

where the change in order of summation/integrals in line 3 follows from Fubini's Theorem,[18] so Follower obedience holds. □

Downwards obedience Take any signal $s \in S$ under which $\alpha^{\mathcal{S},h}(s) = 0$. As $D(A,\theta)$ is non-decreasing in A for each θ, and $A_L^{\mathcal{S},h,k}(\mu)+A_F^{\mathcal{S},h,k}(\mu)$ is non-decreasing in k for each μ, $(D(A_L^{\mathcal{S},h,k}(\mu)+A_F^{\mathcal{S},h,k}(\mu),\theta))_{k\geq 2}$ is a sequence of ($\pi(\cdot|s)$-measurable) functions on $\Delta(S)\times\Theta$ that converges monotonically point-wise to $D(A_L^{\mathcal{S},h}(\mu) + A_F^{\mathcal{S},h}(\mu),\theta)$, and is bounded above and below by the integrable functions $D(1,\theta)$ and $D(0,\theta)$ respectively. Hence, by the Dominated Convergence Theorem,

$$\begin{aligned}&\int_{\Delta(S)\times\Theta} D(A_L^{\mathcal{S},h}(\mu)+A_F^{\mathcal{S},h}(\mu),\theta)d\pi(\mu,\theta|s) \\ &= \lim_{k\to\infty}\int_{\Delta(S)\times\Theta} D(A_L^{\mathcal{S},h,k}(\mu)+A_F^{\mathcal{S},h,k}(\mu),\theta)d\pi(\mu,\theta|s) \leq 0\end{aligned}$$

[18] Whenever I swap the order moving forward, I am appealing to Fubini's Theorem.

Meanwhile, at $k = 1$, having $\alpha^{\mathcal{S},h,1}(\mu) = 0$ implies $\int_{\Delta(S)\times\Theta} D(h,\theta)d\pi(\mu,\theta|s) < 0$. Hence, defining $S^3 \equiv \{s \in S : \alpha^{\mathcal{S},h}(s) = 0\}$ and noting that $1 - A_L^{\mathcal{S},h}(\mu) - A_F^{\mathcal{S},h}(\mu) = \int_{S^3} d\mu(s)$,

$$\begin{aligned}\int_\Theta \int_\mathcal{A} (1 - A_L - A_F) D(h,\theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta &= \int_\Theta \int_{\Delta(S)} \int_{S^3} d\mu(s) D(h,\theta) d\pi(\mu|\theta) f(\theta) d\theta \\ &= \int_{S^3} \left(\int_{\Delta(S)\times\Theta} D(h,\theta) d\pi(\mu,\theta|s) \right) d\pi_S(s) \leq 0\end{aligned}$$

while

$$\begin{aligned}\int_\Theta \int_\mathcal{A} (1 - A_L - A_F) D(A_L + A_F, \theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta &= \int_\Theta \int_{\Delta(S)} \int_{S^3} D(A_L^{\mathcal{S},h}(\mu) + A_F^{\mathcal{S},h}(\mu), \theta) d\mu(s) d\pi(\mu|\theta) f(\theta) d\theta \\ &= \int_{S^3} \left(\int_{\Delta(S)\times\Theta} D(A_L^{\mathcal{S},h}(\mu) + A_F^{\mathcal{S},h}(\mu), \theta) d\pi(\mu,\theta|s) \right) d\pi_S(s) \geq 0\end{aligned}$$

Hence, σ satisfies Downwards-obedience. □

B3: Other Proofs

Proof of Lemma 1 Follows from in-text discussion. □

Proof of Theorem 1 To ease notation, I assume states are ordered such that $\theta \geq \theta'$ if and only if $\frac{-D(H,\theta)}{v(1,\theta)} \leq \frac{-D(H,\theta')}{v(1,\theta')}$. Furthermore, I assume the designer cannot induce all agents to invest on all states (for otherwise the claim holds trivially), which implies $\int_{\theta^H}^{\overline{\theta}} D(H,\theta) f(\theta) d\theta = 0$.

Now, observe that the designer's problem can be reformulated as the *primal* linear programming problem: choose a measurable function $q : \Theta \to \mathbb{R}_+$ to solve

$$\begin{aligned}&\max_q \int_\Theta v(1,\theta) q(\theta) f(\theta) d\theta \\ &\text{s.t.} \quad \int_\Theta D(H,\theta) q(\theta) f(\theta) d\theta \geq 0 \quad \text{and} \quad \forall \theta \in \Theta,\ q(\theta) \leq 1\end{aligned}$$

The above admits the following *dual problem*: choose a $\lambda \geq 0$ and a measurable function $\phi : \Theta \to \mathbb{R}_+$ to solve

$$\min_{\lambda,\phi} \int_{\Theta} \phi(\theta) f(\theta) d\theta$$
$$\text{s.t.} \quad v(1,\theta) + \lambda D(H,\theta) \leq \phi(\theta), \quad \forall \theta \in \Theta$$

Given this, consider the primal-dual pair $(q^*, (\lambda^*, \phi^*))$ defined as follows

$$q^*(\theta) \equiv \begin{cases} 1, & \theta \geq \theta^H \\ 0, & \theta < \theta^H \end{cases}, \quad \lambda^* \equiv \frac{-D(H,\theta^H)}{v(1,\theta^H)}, \quad \phi^*(\theta) \equiv \max\{0, v(1,\theta) + \lambda^* D(H,\theta)\}$$

where q^* captures agents' behaviours under the information structure described in Theorem 1. By definition, q^* is feasible for the primal problem and (λ^*, ϕ^*) is feasible for the dual problem. Further note that $v(1,\theta) + \lambda^* D(H,\theta) \geq 0$ if and only if $\frac{-D(H,\theta)}{v(1,\theta)} \leq \lambda^* \equiv \frac{-D(H,\theta^H)}{v(1,\theta^H)}$, which holds if and only if $\theta \geq \theta^*$. Hence,

$$\underbrace{\int_{\Theta} \phi^*(\theta) f(\theta) d\theta}_{\text{Value of dual objective under } (\lambda^*,\phi^*)} = \underbrace{\int_{\theta^H}^{\overline{\theta}} v(1,\theta) f(\theta) d\theta}_{\text{Value of primal objective under } q^*} + \lambda^* \underbrace{\int_{\theta^H}^{\overline{\theta}} D(H,\theta) f(\theta) d\theta}_{=0}$$

and so by the Weak Duality Theorem, q^* solves the primal problem. □

Proof of Corollary 1 Follows from in-text discussion. □

Proof of Corollary 2 Take any leader-follower outcome σ satisfying upper introspective obedience under H. Let

$$\mathcal{R}_1 \equiv \{((A_L, A_F), \theta) : D(A_L + A_F, \theta) \geq 0\}$$
$$\mathcal{R}_2 \equiv \{((A_L, A_F), \theta) : ((A_L, A_F), \theta) \notin \mathcal{R}_1 \text{ and } D(h,\theta) \geq 0\}$$

and define, for each $\theta \in \Theta$, the transport map $\widetilde{T}_\theta : \mathcal{A} \to \mathcal{A}$ as follows:

$$\widetilde{T}_\theta(A_L, A_F) \equiv \begin{cases} (A_L, 1 - A_L), & A_L + A_F < 1 \text{ and } ((A_L, A_F), \theta) \in \mathcal{R}_1 \\ (1, 0), & A_L + A_F < 1 \text{ and } ((A_L, A_F), \theta) \in \mathcal{R}_2 \\ (A_L, A_F), & \text{otherwise} \end{cases}$$

Now, let $\tilde{\sigma}$ be defined as the leader-follower outcome under which $\tilde{\sigma}(\cdot|\theta)$ is the push-forward of $\sigma(\cdot|\theta)$ through $\widetilde{T}_\theta$ for all $\theta \in \Theta$. It is easily verified that $p_{\tilde{\sigma}}$ first-order stochastically dominates p_σ per state, so the designer weakly prefers $\tilde{\sigma}$ over σ. To see that $\tilde{\sigma}$ satisfies introspective obedience, for leader obedience,

$$\begin{aligned}\int_\Theta \int_\mathcal{A} A_L D(H,\theta) d\tilde{\sigma}(A_L, A_F|\theta) f(\theta) d\theta &= \left[\begin{array}{c} \int_\Theta \int_\mathcal{A} A_L D(H,\theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \\ + \int_{\mathcal{R}_2} (1-A_L) D(H,\theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \end{array} \right] \\ &\geq \int_\Theta \int_\mathcal{A} A_L D(H,\theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \geq 0\end{aligned}$$

and

$$\begin{aligned}\int_\Theta \int_\mathcal{A} A_L D(A_L,\theta) d\tilde{\sigma}(A_L, A_F|\theta) f(\theta) d\theta &= \left[\begin{array}{c} \int_{\mathcal{A}\times\Theta\setminus\mathcal{R}_2} A_L D(A_L,\theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \\ + \int_{\mathcal{R}_2} D(1,\theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \end{array} \right] \\ &\geq \left[\begin{array}{c} \int_{\mathcal{A}\times\Theta\setminus\mathcal{R}_2} A_L D(A_L,\theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \\ + \int_{\mathcal{R}_2} A_L D(A_L,\theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \end{array} \right] \\ &\geq 0\end{aligned}$$

Meanwhile, for follower obedience

$$\begin{aligned}\int_\Theta \int_\mathcal{A} \int_0^{A_F} D(A_L+i,\theta) di d\tilde{\sigma}(A_L, A_F|\theta) f(\theta) d\theta &= \left[\begin{array}{c} \int_{\mathcal{A}\times\Theta\setminus\mathcal{R}_1} \int_0^{A_F} D(A_L+i,\theta) di d\sigma(A_L, A_F|\theta) f(\theta) d\theta \\ + \int_{\mathcal{R}_1} \int_0^{1-A_L} D(A_L+i,\theta) di d\sigma(A_L, A_F|\theta) f(\theta) d\theta \end{array} \right] \\ &\geq \left[\begin{array}{c} \int_{\mathcal{A}\times\Theta\setminus\mathcal{R}_1} \int_0^{A_F} D(A_L+i,\theta) di d\sigma(A_L, A_F|\theta) f(\theta) d\theta \\ + \int_{\mathcal{R}_1} \int_0^{A_F} D(A_L+i,\theta) di d\sigma(A_L, A_F|\theta) f(\theta) d\theta \end{array} \right] \\ &= \int_\Theta \int_\mathcal{A} \int_0^{A_F} D(A_L+i,\theta) di d\sigma(A_L, A_F|\theta) f(\theta) d\theta \geq 0\end{aligned}$$

Finally, because, for each state θ, the transport map(s) $\tilde{T}_\theta$ move mass off pairs (A_L, A_F) under which either $D(H,\theta) \geq 0$ or $D(A_L + A_F, \theta) \geq 0$ hold, it follows that (14) and (15) both continue to hold under $\tilde{\theta}$, i.e., downwards obedience holds. □

Proof of Lemma 2 Take any leader-follower outcome σ satisfying upper introspective obedience under H. Following the proof of Corollary 2, I assume, without loss of generality, that for all $\theta \in \Theta$, if $(A_L, A_F) \in \text{supp}(\sigma(\cdot|\theta))$ and $A_L + A_F \in (0,1)$, then $D(H,\theta) < 0$ and $D(A_L + A_F, \theta) < 0$ both hold. Define the measurable set of such pairs for each state as

$$X(\theta) \equiv \{(A_L, A_F) : A_L + A_F \in (0,1),\ D(H,\theta) < 0,\ D(A_L + A_F, \theta) < 0\}$$

By the convexity of $\int_{A_L}^{A_L+x} D(i,\theta)di$ in x, observe

$$\begin{aligned}
\int_\Theta \int_{X(\theta)} \left(\int_{A_L}^{A_L+A_F} D(i,\theta)di \right) d\sigma(A_L, A_F|\theta) f(\theta) d\theta &\leq \int_\Theta \int_{X(\theta)} \frac{A_F}{\underline{A}(\theta) - A_L} \left(\int_{A_L}^{\underline{A}(\theta)} D(i,\theta)di \right) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \\
&\leq \int_\Theta \int_{X(\theta)} \frac{A_F}{\underline{A}(\theta) - A_L} \left(\int_{A_L}^{1} D(i,\theta)di \right) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \\
&+ \int_\Theta \int_{X(\theta)} \left(1 - \frac{A_F}{\underline{A}(\theta) - A_L} \right) \frac{A_L}{\underline{A}(\theta)} \left(\int_{\underline{A}(\theta)}^{1} D(i,\theta)di \right) d\sigma(A_L, A_F|\theta) f(\theta) d\theta
\end{aligned}$$

while

$$\int_\Theta \int_{X(\theta)} A_L D(A_L,\theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \leq \int_\Theta \int_{X(\theta)} \begin{pmatrix} \frac{A_F A_L}{\underline{A}(\theta) - A_L} D(A_L, \theta) \\ + \left[1 - \frac{A_F}{\underline{A}(\theta) - A_L}\right] \frac{A_L}{\underline{A}(\theta)} \underline{A}(\theta) D(\underline{A}(\theta), \theta) \end{pmatrix} d\sigma(A_L, A_F|\theta) f(\theta) d\theta$$

Now, consider the leader-follower outcome $\tilde{\sigma}$ defined as follows: for all $\theta \in \Theta$, $\tilde{\sigma}(\cdot|\theta) \equiv K_\theta \circ \sigma(\cdot|\theta)$, where $K_\theta : \mathcal{A} \to \Delta(\mathcal{A})$ is defined as follows:

1. For all $(A_L, A_F) \in X(\theta)$, $K_\theta(\cdot|A_L, A_F)$ has its mass concetrated on the pairs $(0,0), (\underline{A}(\theta), 1 - \underline{A}(\theta)), (A_L, 1 - A_L)$, and assigns probabilities as follows:

$$K_\theta(0,0|A_L, A_F) = \underbrace{\left[1 - \frac{A_F}{\underline{A}(\theta) - A_L}\right]\left[1 - \frac{A_L}{\underline{A}(\theta)}\right]}_{\equiv \kappa((A_L, A_F), \theta)},$$

$$K_\theta(A_L, 1 - A_L|A_L, A_F) = \frac{A_F}{\underline{A}(\theta) - A_L}, \quad K_\theta(\underline{A}(\theta), 1 - \underline{A}(\theta)|A_L, A_F) = \left[1 - \frac{A_F}{\underline{A}(\theta) - A_L}\right] \frac{A_L}{\underline{A}(\theta)}$$

   where if $A_L = \underline{A}(\theta)$, then $K_\theta(A_L, 1 - A_L|A_L, A_F = 1 - \kappa((A_L, A_F), \theta)$.

2. If $(A_L, A_F) \notin X(\theta)$, then $K_\theta(\cdot|A_L, A_F) = \delta_{(A_L, A_F)}$.

Clearly, $\tilde{\sigma}$ satisfies perfect coordination, while

$$\begin{aligned}
V(\tilde{\sigma}) &= V(\sigma) + \int_\Theta \int_{X(\theta)} \begin{pmatrix} \kappa((A_L, A_F), \theta) v(0,\theta) + (1 - \kappa((A_L, A_F), \theta)) v(1,\theta) \\ -v(A_L + A_F, \theta) \end{pmatrix} d\sigma(A_L, A_F|\theta) f(\theta) d\theta \\
&\geq V(\sigma) + \int_\Theta \int_{X(\theta)} \begin{pmatrix} \kappa((A_L, A_F), \theta) v(0,\theta) + (1 - \kappa((A_L, A_F), \theta)) v(\underline{A}(\theta), \theta) \\ -v(A_L + A_F, \theta) \end{pmatrix} d\sigma(A_L, A_F|\theta) f(\theta) d\theta \\
&\geq V(\sigma)
\end{aligned}$$

where the last inequality follows from the fact that $v(A, \theta)$ is convex on $[0, \underline{A}(\theta)]$ by Assumption 1, and that $(1 - \kappa((A_L, A_F), \theta))\underline{A}(\theta) = A_L + A_F$. Hence, the designer prefers $\tilde{\sigma}$ over σ.

Thus, to complete the proof of Lemma 2, it must be proven that $\tilde{\sigma}$ satisfies upper introspective obedience. For leader obedience,

$$\int_\Theta \int_\mathcal{A} A_L D(H,\theta) d\tilde{\sigma}(A_L, A_F|\theta) f(\theta) d\theta = \left[\begin{array}{c} \int_\Theta \int_\mathcal{A} A_L D(H,\theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \\ + \int_\Theta \int_{X(\theta)} \left(\frac{A_L A_F}{\underline{A}(\theta) - A_L} + \left[1 - \frac{A_F}{\underline{A}(\theta) - A_L} \right] \frac{A_L}{\underline{A}(\theta)} \underline{A}(\theta) - A_L \right) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \end{array} \right]$$
$$= \int_\Theta \int_\mathcal{A} A_L D(H,\theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \geq 0$$

and

$$\int_\Theta \int_\mathcal{A} A_L D(A_L,\theta) d\tilde{\sigma}(A_L, A_F|\theta) f(\theta) d\theta = \left[\begin{array}{c} \int_\Theta \int_\mathcal{A} A_L D(A_L,\theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \\ + \int_\Theta \int_{X(\theta)} \left(\begin{array}{c} \frac{A_F A_L}{\underline{A}(\theta) - A_L} D(A_L, \theta) \\ + \left[1 - \frac{A_F}{\underline{A}(\theta) - A_L} \right] \frac{A_L}{\underline{A}(\theta)} \underline{A}(\theta) D(\underline{A}(\theta), \theta) \\ - A_L D(A_L, \theta) \end{array} \right) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \end{array} \right]$$
$$\geq \int_\Theta \int_\mathcal{A} A_L D(A_L,\theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \geq 0$$

For follower obedience,

$$\int_\Theta \int_\mathcal{A} \left(\int_{A_L}^1 D(i,\theta) di \right) d\tilde{\sigma}(A_L, A_F|\theta) f(\theta) d\theta = \left[\begin{array}{c} \int_\Theta \int_\mathcal{A} \left(\int_{A_L}^1 D(i,\theta) di \right) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \\ + \int_\Theta \int_{X(\theta)} \frac{A_F}{\underline{A}(\theta) - A_L} \left(\int_{A_L}^1 D(i,\theta) di \right) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \\ + \int_\Theta \int_{X(\theta)} \left(1 - \frac{A_F}{\underline{A}(\theta) - A_L} \right) \frac{A_L}{\underline{A}(\theta)} \left(\int_{\underline{A}(\theta)}^1 D(i,\theta) di \right) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \\ - \int_\Theta \int_{X(\theta)} \left(\int_{A_L}^{A_L + A_F} D(i,\theta) di \right) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \end{array} \right]$$
$$\geq \int_\Theta \int_\mathcal{A} \left(\int_{A_L}^1 D(i,\theta) di \right) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \geq 0$$

Thus, $\tilde{\sigma}$ satisfies upper introspective obedience. □

Proof of Theorem 3 I split the proof into three parts.

Part 1 Suppose that $H \leq \underline{H}$. Let $\underline{T} : \mathcal{A} \to \mathcal{A}$ denote the transport map which pools the mass of all (A_L, A_F) onto $(0, 1)$ if $A_L + A_F = 1$, and leaves it untouched otherwise. Take any leader-follower outcome σ satisfying upper introspective obedience under H, and let $\tilde{\sigma}$ be defined as the leader-follower outcome where $\tilde{\sigma}(\cdot|\theta)$ is the push-forward of $\sigma(\cdot|\theta)$ through $\underline{T}$ for all $\theta \in \Theta$. Clearly, $\tilde{\sigma}$ implements the same outcome as σ, so the designer is indifferent between σ and $\tilde{\sigma}$. Meanwhile, to see that $\tilde{\sigma}$ satisfies upper introspective obedience (which, since $\tilde{\sigma}$ is a follower-only outcome, requires showing that follower-obedience holds), observe that

$$\int_\Theta \int_\mathcal{A} \left(\int_{A_L}^1 D(i,\theta)di \right) d\tilde{\sigma}(A_L, A_F|\theta) f(\theta) d\theta = \begin{bmatrix} \int_\Theta \int_\mathcal{A} \left(\int_{A_L}^1 D(i,\theta)di \right) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \\ + \int_\Theta \int_\mathcal{A} \left(\int_0^{A_L} D(i,\theta)di \right) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \end{bmatrix}$$

$$\geq \begin{bmatrix} \int_\Theta \int_\mathcal{A} \left(\int_{A_L}^1 D(i,\theta)di \right) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \\ + \underbrace{\int_\Theta \int_\mathcal{A} A_L D(0,\theta) d\sigma(A_L, A_F|\theta) f(\theta) d\theta}_{\text{Leader obedience under } \sigma} \end{bmatrix}$$

$$\geq \int_\Theta \int_\mathcal{A} \left(\int_{A_L}^1 D(i,\theta)di \right) d\sigma(A_L, A_F|\theta) f(\theta) d\theta \geq 0$$

This proves Part 1. □

Part 2 Suppose that $H \geq \overline{H}$. Here, consider the transport map $\overline{T} : \mathcal{A} \to \mathcal{A}$ which pools the mass of all (A_L, A_F) with $A_L + A_F = 1$ onto $(1, 0)$, and otherwise leaves it at $(0, 0)$. Taking any leader-follower outcome σ satisfying upper introspective obedience and letting $\tilde{\sigma}$ be defined such that $\tilde{\sigma}(\cdot|\theta)$ is the push-forward of $\sigma(\cdot|\theta)$ through $\overline{T}$ for all $\theta \in \Theta$, an analogous argument to Part 1 implies $\tilde{\sigma}$ satisfies upper introspective obedience and the designer is indifferent between σ and $\tilde{\sigma}$. That $\tilde{\sigma}$ is a leader-only outcome then yields the claim. □

Part 3 First, suppose $H > \underline{H}^*$. I prove that σ_F cannot be optimal for the designer. To begin, notice that

$$\{\theta : x_F(\theta) \geq \underline{x}_F\} = \overline{\Theta} \cup \underbrace{\{\theta : \theta \notin \overline{\Theta} \text{ and } x_F(\theta) \geq \underline{x}_F\}}_{\equiv \underline{\Theta}}$$

Now, let

$$\overline{\Theta}^H \equiv \{\theta : D(H,\theta) > D(0,\theta) \text{ and } x_F(\theta) \geq \underline{x}_F\}$$

and take any $\epsilon, \delta \in (0, 1)$ sufficiently small such that

$$\epsilon \underbrace{\int_{\overline{\Theta}} \int_0^H D(i,\theta) di f(\theta) d\theta}_{>0} + \delta \underbrace{\int_{\underline{\Theta}} \int_0^H D(i,\theta) di f(\theta) d\theta}_{<0} = 0$$

By upper semi-continuity of $D(A,\theta)$, for all $\theta \in \overline{\Theta}^H$, $\int_0^H D(i,\theta)di < HD(H,\theta)$, so

$$\epsilon \int_{\overline{\Theta}} HD(H,\theta)f(\theta)d\theta + \delta \int_{\underline{\Theta}} HD(H,\theta)f(\theta)d\theta > 0$$

and thus, there must exist a $\kappa > 0$ such that

$$\epsilon \int_{\overline{\Theta}} HD(H,\theta)f(\theta)d\theta + \delta \int_{\underline{\Theta}} HD(H,\theta)f(\theta)d\theta + \kappa \int_{\Theta\backslash(\overline{\Theta}\cup\underline{\Theta})} D(H,\theta)f(\theta)d\theta \geq 0$$

Now, consider the leader-follower outcome $\tilde{\sigma}$ defined as follows: for all $\theta \in \Theta$, $\tilde{\sigma}(\cdot|\theta) \equiv K_\theta \circ \sigma(\cdot|\theta)$, where $K_\theta : \mathcal{A} \to \Delta(\mathcal{A})$ is defined as follows:

1. For all $\theta \in \overline{\Theta}$ and $(A_L, A_F) = (0,1)$, $K_\theta(\cdot|A_L, A_F)$ has its mass concentrated on the pairs $(0,1), (H, 1-H)$, and assigns probabilities as follows: $K_\theta(0,1|A_L, A_F) = 1-\epsilon$ and $K_\theta(H, 1-H|A_L, A_F) = \epsilon$

2. For all $\theta \in \underline{\Theta}$ and $(A_L, A_F) = (0,1)$, $K_\theta(\cdot|A_L, A_F)$ has its mass concentrated on the pairs $(0,1), (H, 1-H)$, and assigns probabilities as follows: $K_\theta(0,1|A_L, A_F) = 1-\delta$ and $K_\theta(H, 1-H|A_L, A_F) = \delta$

3. For all $\theta \in \Theta\backslash(\overline{\Theta}\cup\underline{\Theta})$ and $(A_L, A_F) = (0,0)$, $K_\theta(\cdot|A_L, A_F)$ has its mass concentrated on the pairs $(0,0), (1,0)$, and assigns probabilities as follows: $K_\theta(0,0|A_L, A_F) = 1-\kappa$ and $K_\theta(1,0|A_L, A_F) = \kappa$

4. For all other triples $((A_L, A_F), \theta)$, let $K_\theta(\cdot|A_L, A_F) \equiv \delta_{(A_L, A_F)}$.

By construction, under $\tilde{\sigma}$, all agents invest on all states in $\{\theta : x_F(\theta) \geq \underline{x}_F\}$, and with probability $\kappa > 0$ on all other states. Hence, the designer strictly prefers $\tilde{\sigma}$ than σ_F. Finally,

$$\int_\Theta \int_{\mathcal{A}} \left(\int_{A_L}^1 D(i,\theta)di\right) d\tilde{\sigma}(A_L, A_F|\theta)f(\theta)d\theta = \begin{bmatrix} \int_\Theta \int_{\mathcal{A}} \left(\int_{A_L}^1 D(i,\theta)di\right) d\sigma_F(A_L, A_F|\theta)f(\theta)d\theta \\ -\left(\epsilon \int_{\overline{\Theta}} \int_0^H D(i,\theta)di f(\theta)d\theta + \delta \int_{\underline{\Theta}} \int_0^H D(i,\theta)di f(\theta)d\theta\right) \end{bmatrix}$$

$$= \int_\Theta \int_{\mathcal{A}} \left(\int_{A_L}^1 D(i,\theta)di\right) d\sigma_F(A_L, A_F|\theta)f(\theta)d\theta \geq 0$$

So $\tilde{\sigma}_F$ satisfies follower obedience. Meanwhile,

$$\int_\Theta \int_\mathcal{A} A_L D(H,\theta) d\tilde{\sigma}(A_L, A_F|\theta) f(\theta) d\theta = \begin{bmatrix} \epsilon \int_{\overline{\Theta}} HD(H,\theta)f(\theta)d\theta \\ +\delta \int_{\underline{\Theta}} HD(H,\theta)f(\theta)d\theta \\ +\kappa \int_{\Theta\setminus(\overline{\Theta}\cup\underline{\Theta})} D(H,\theta)f(\theta)d\theta \end{bmatrix}$$
$$\geq \int_\Theta \int_\mathcal{A} A_L D(H,\theta) d\sigma_F(A_L, A_F|\theta) f(\theta) d\theta \geq 0$$

and

$$\int_\Theta \int_\mathcal{A} A_L D(A_L,\theta) d\tilde{\sigma}(A_L, A_F|\theta) f(\theta) d\theta = \begin{bmatrix} \epsilon \int_{\overline{\Theta}} HD(H,\theta)f(\theta)d\theta \\ +\delta \int_{\underline{\Theta}} HD(H,\theta)f(\theta)d\theta \\ +\kappa \int_{\Theta\setminus(\overline{\Theta}\cup\underline{\Theta})} D(H,\theta)f(\theta)d\theta \end{bmatrix}$$
$$\geq \int_\Theta \int_\mathcal{A} A_L D(A_L,\theta) d\sigma_F(A_L, A_F|\theta) f(\theta) d\theta \geq 0$$

Hence, $\tilde{\sigma}_F$ satisfies upper introspective obedience. Thus, σ_F cannot be optimal for the designer.

To prove that $H < \overline{H}^*$ implies σ_L^H, one follows a similar schema to the above, but instead by marginally switching the spread over leaders and followers of σ_L^H from $(1,0)$ to $(1-\epsilon,\epsilon)$ for small $\epsilon > 0$ on a subset of states (with $D(H,\theta) < D(1,\theta)$). I omit repeating the argument for brevity. □

Proof of Lemma 3 Follows from in-text discussion. □

Proof of Corollary 3 Follows from in-text discussion. □

References

Akerlof, Robert and Richard Holden, "Movers and Shakers," *The Quarterly Journal of Economics*, 2016, *131* (4), 1849–1874.

_ **and** _ , "Capital Assembly," *The Journal of Law, Economics, and Organization*, 07 2019, *35* (3), 489–512.

_ **and** _ , "Coordinating Supply Chains," *Working paper*, 2023.

_ **,** _ **, and Luis Rayo**, "Network Externalities and Market Dominance," *Management Science*, 2023.

Andreoni, James, "Toward a Theory of Charitable Fund-Raising," *Journal of Political Economy*, 1998, *106* (6), 1186–1213.

Arieli, Itai and Yakov Babichenko, "Private Bayesian persuasion," *Journal of Economic Theory*, 2019, *182*, 185–217.

Aumann, Robert J., "Subjectivity and correlation in randomized strategies," *Journal of Mathematical Economics*, 1974, *1* (1), 67–96.

Banerjee, Abhijit V., "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics*, 1992, *107* (3), 797–817.

Bergemann, Dirk and Stephen Morris, "Bayes correlated equilibrium and the comparison of information structures in games," *Theoretical Economics*, 2016, *11* (2), 487–522.

_ and _ , "Information Design: A Unified Perspective," *Journal of Economic Literature*, March 2019, *57* (1), 44–95.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch, "A theory of fads, fashion, custom, and cultural change as informational cascades," *Journal of political Economy*, 1992, *100* (5), 992–1026.

Candogan, Ozan and Kimon Drakopoulos, "Optimal Signaling of Content Accuracy: Engagement vs. Misinformation," *Operations Research*, 2020, *68* (2). Published Online: 2 Mar 2020.

Carlsson, Hans and Eric van Damme, "Global Games and Equilibrium Selection," *Econometrica*, 1993, *61* (5), 989–1018.

Chang, Jen-Wen, "The Economics of Crowdfunding," *American Economic Journal: Microeconomics*, May 2020, *12* (2), 257–280.

Colombo, Massimo G., Chiara Franzoni, and Claudia Rossi-Lamastra, "Internal Social Capital and the Attraction of Early Contributions in Crowdfunding," *Entrepreneurship Theory and Practice*, 2015, *39* (1), 75–100.

Crawford, Vincent P., Miguel A. Costa-Gomes, and Nagore Iriberri, "Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications," *Journal of Economic Literature*, March 2013, *51* (1), 5–62.

Crosetto, Paolo and Tobias Regner, "It's Never Too Late: Funding Dynamics and Self-Pledges in Reward-Based Crowdfunding," *Research Policy*, 2018, *47* (8), 1463–1477.

Deb, Joyee, Aniko Oery, and Kevin R. Williams, "Aiming for the Goal: Contribution Dynamics of Crowdfunding," 2023.

Dworczak, P. and G. Martini, "The Simple Economics of Optimal Persuasion," *Journal of Political Economy*, 2019, *127*, 1993–2048.

Ellman, Matthew and Michele Fabi, "A Theory of Crowdfunding Dynamics," 2022.

— **and Sjaak Hurkens**, "Optimal crowdfunding design," *Journal of Economic Theory*, 2019, *184*, 104939.

Gentzkow, M. and E. Kamenica, "A Rothschild-Stiglitz Approach to Bayesian Persuasion," *American Economic Review: Papers and Proceedings*, 2016, *106* (5), 597–601.

Goldstein, Itay and Chong Huang, "Bayesian Persuasion in Coordination Games," *American Economic Review*, May 2016, *106* (5), 592–96.

Halac, Marina, Elliot Lipnowski, and Daniel Rappoport, "Rank Uncertainty in Organizations," *American Economic Review*, March 2021, *111* (3), 757–786.

— **, — , and —** , "Addressing Strategic Uncertainty with Incentives and Information," *AEA Papers and Proceedings*, May 2022, *112*, 431–37.

Hoshino, Tetsuya, "MULTI-AGENT PERSUASION: LEVERAGING STRATEGIC UNCERTAINTY," *International Economic Review*, 2022, *63* (2), 755–776.

Indiegogo, "Build Excitement for Your Crowdfunding Campaign with a Pre-Launch Page," *Retrieved from https://entrepreneur.indiegogo.com/education/article/how-to-build-pre-launch-page/, accessed 2023-11-13*, 2023.

Inostroza, Nicolas and Alessandro Pavan, "Adversarial Coordination and Public Information Design," *Working paper*, 2023.

Kamenica, Emir and Matthew Gentzkow, "Bayesian Persuasion," *The American economic review*, 2011, *101* (6), 2590–2615.

Kets, Willemien, "Organizational Design: Culture and Incentives," *Working paper*, 2021.

— **and Alvaro Sandroni**, "A Theory of Strategic Uncertainty and Cultural Diversity," *The Review of Economic Studies*, 08 2020, *88* (1), 287–333.

— **, Wouter Kager, and Alvaro Sandroni**, "The value of a coordination game," *Journal of Economic Theory*, 2022, *201*, 105419.

Kickstarter, "How to Launch a Product on Kickstarter: Building Your Community," *Retrieved from https://www.kickstarter.com/articles/how-to-launch-a-product-on-kickstarter-building-your-community*, *accessed 2023-11-13*, 2023.

Kolotilin, Anton, "Optimal information disclosure: A linear programming approach," *Theoretical Economics*, 2018, *13* (2), 607–635.

_ , Roberto Corrao, and Alexander Wolitzky, "Persuasion and Matching: Optimal Productive Transport," *Working paper*, 2023.

_ , Timofiy Mylovanov, and Andriy Zapechelnyuk, "Censorship as optimal persuasion," *Theoretical Economics*, 2022, *17* (2), 561–585.

Li, Fei, Yangbo Song, and Mofei Zhao, "Global Manipulation by Local Obfuscation," *Working paper*, 2022.

_ , _ , and _ , "Global manipulation by local obfuscation," *Journal of Economic Theory*, 2023, *207*, 105575.

Mathevet, Laurent, Jacopo Perego, and Ina Taneva, "On Information Design in Games," *Journal of Political Economy*, 2020, *128* (4), 1370–1404.

Milgrom, Paul and John Roberts, "Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities," *Econometrica*, 1990, *58* (6), 1255–1277.

Morris, Stephen and Hyun Song Shin, *Global games: Theory and applications*, United Kingdom: Cambridge University Press, January

_ and Hyun Song Shin, "Catalytic finance: When does it work?," *Journal of International Economics*, 2006, *70* (1), 161–177.

_ and Takashi Ui, "Generalized Potentials and Robust Sets of Equilibria," *Journal of Economic Theory*, 2005, *124* (1), 45–78.

_ , Daisuke Oyama, and Satoru Takahashi, "Implementation via Information Design in Binary-Action Supermodular Games," *Working paper*, 2022.

_ , _ , and _ , "Implementation via Information Design using Global Games," *Working paper*, 2022.

_ , _ , and _ , "On the Joint Design of Information and Transfers," *Working paper*, 2022.

Oyama, Daisuke and Satoru Takahashi, "Generalized Belief Operator and Robustness in Binary-Action Supermodular Games," *Econometrica*, 2020, *88* (2), 693–726.

Rubinstein, Ariel, "The Electronic Mail Game: Strategic Behavior under 'Almost Common Knowledge'," *American Economic Review*, 1989, *79* (3), 385–391.

Smolin, Alex and Takuro Yamashita, "Information Design in Games: Certification Approach," *Working paper*, 2023.

Strausz, Roland, "A Theory of Crowdfunding: A Mechanism Design Approach with Demand Uncertainty and Moral Hazard," *American Economic Review*, June 2017, *107* (6), 1430–76.

Sun, Yeneng, "The exact law of large numbers via Fubini extension and characterization of insurable risks," *Journal of Economic Theory*, 2006, *126* (1), 31–69.

Taneva, Ina and Laurent Mathevet, "Organized Information Transmission," *Working paper*, 2023.