Informed Information Design

Frédéric Koessler, Vasiliki Skreta

Informed Information Design*

Frédéric KOESSLER † Vasiliki SKRETA ‡

8 February 2022

Abstract

A designer is privately informed about the state and chooses an information-disclosure mechanism to influence the decisions of multiple agents playing a game. We define *interim-optimal mechanisms*, a subset of incentive-compatible mechanisms that are optimal in the sense that the informed designer cannot credibly find an alternative mechanism that strictly improves his interim payoff. We prove that an interim-optimal mechanism exists and that every interim-optimal mechanism is a perfect Bayesian equilibrium outcome of the informed-designer game. An ex-ante optimal mechanism may not be interim optimal, but it is when it is ex-post optimal. Likewise, the unraveling outcome in disclosure games is interim optimal. We provide a belief-based characterization of interim-optimal mechanisms and compare them with ex-ante optimal ones in common economic environments. In settings with strategic complements and binary actions, every ex-ante optimal mechanism is interim optimal. We compare interim optimality to other solutions of informed-principal problems.

KEYWORDS: interim information design, Bayesian persuasion, informed principal, disclosure games, unraveling, neutral optimum, strong neologism proofness, core mechanism, verifiable types.

JEL CLASSIFICATION: C72; D82.

*We thank the Editor and four anonymous referees for excellent comments. We also thank Ricardo Alonso, Elchanan Ben-Porath, Francoise Forges, Sergiu Hart, Andres Salamanca, and Joel Sobel for useful feedback as well as seminar participants at Bonn-Berlin Micro Theory Seminar, Brown, Concordia, HEC, Israel Theory seminar, Paris Game Theory Seminar, Paris School of Economics, Rice University, the Workshop in Dynamic Games in Quimper, VSET, Wisconsin, and Yale. Frederic Koessler acknowledges the support of the ANR through the program Investissements d'Avenir (ANR-17-EURE-001) and under grant ANR StratCom (ANR-19-CE26-0010-01). Vasiliki Skreta acknowledges funding by the European Research Council (ERC) consolidator grant 682417 "Frontiers In Design." Alkis Georgiadis-Harris provided excellent research assistance.

†Paris School of Economics – CNRS, 48 boulevard Jourdan, 75014 Paris, France; frederic.koessler@psemail.eu.

‡UT Austin, UCL and CEPR. vskreta@gmail.com.

1 Introduction

Decisions ranging from voting, career, and investment depend crucially on the information agents have. In the large and influential literature on Bayesian persuasion and information design, an uninformed designer optimally commits to a disclosure rule.[1] The purpose of the designer is to achieve a certain goal: a seller tries to persuade buyers that their product is good, a politician encourages voters to vote for them, and a pharmaceutical company tries to convince a doctor to prescribe their medicine. Often, however, parties selecting the informativeness of a procedure (details on a product brochure, scope and breadth of an investment opportunities study, dimensions on which to test a new vehicle) have private information that shapes their preferences about which procedure to choose, which in turn, affects inferences and the ultimate nature of information that is disclosed.

In this paper, we study *interim* information design, which is akin to an informed-principal problem (Myerson, 1983). We take the same interim perspective as the influential works on disclosure games[2] but enlarge the choice set of the informed party: instead of restricting attention to deterministic evidence, the designer in our setting can choose any mapping from the state to a distribution of signals. We define *interim-optimal* mechanisms, a subset of incentive-compatible mechanisms that are optimal in the sense that the informed designer cannot credibly find an alternative mechanism that strictly improves his interim payoff. We prove that an interim-optimal mechanism exists and that every interim-optimal mechanism is a perfect Bayesian equilibrium outcome of the informed-designer game. We compare interim-optimal mechanisms with ex-ante optimal mechanisms, which arise when the designer chooses the mechanism ex ante, before observing the state (the "commitment solution", Kamenica and Gentzkow, 2011, Bergemann and Morris, 2016, Taneva, 2019).[3] Finally, we show that the unraveling outcome in Milgrom (1981) and Grossman (1981) is interim optimal, and thus, it remains a perfect Bayesian equilibrium outcome even when the designer can deviate to any disclosure mechanism.

Two key differences exist between the standard information-design setting and ours. First, the designer's interim incentives differ from his ex-ante ones. For example, a high-quality seller prefers information to be disclosed, but a low-quality one does not. Second, the *choice* of the information-disclosure policy can reveal information to the agents. For example, customers can update their expected valuation for a product if the seller designs product-testing procedures that have a low probability of uncovering bad characteristics, or if some product features are not tested at all. Interim information design is thus not a constrained optimization problem but a game that shares

[1] See, for example, Kamenica and Gentzkow (2011), Bergemann and Morris (2016), Taneva (2019), and Mathevet et al. (2020). See Bergemann and Morris (2019), Kamenica (2019), and Forges (2020) for surveys of the literature.

[2] See, for example, Milgrom (1981), Grossman (1981), Okuno-Fujiwara et al. (1990), Seidmann and Winter (1997), Sher (2011), Hagenbach et al. (2014), Hart et al. (2017), and Ben-Porath et al. (2019).

[3] Other papers that relax the commitment assumption of the standard information-design paradigm in different ways than us include Lipnowski et al. (2022), Lipnowski and Ravid (2020), and references therein. In Lipnowski et al. (2022), the designer is uninformed and chooses an experiment ex ante, but can ex-post lie when the signal realization is "bad." Lipnowski and Ravid (2020) study cheap-talk communication (rather than commitment to a disclosure rule) by an informed party that has state-independent preferences over actions.

features with disclosure games (cf. Milgrom, 1981) and informed-principal problems (cf. Myerson, 1983).

We consider the following setting. There are $n+1$ players: the designer and n agents. Players' payoffs depend on the state of the world and on the profile of actions chosen. The informed-designer game proceeds as follows: the designer observes the state (which is distributed according to a common prior) and can design any information-disclosure mechanism, namely, a mapping from the set of states to the set of distributions over signals $\Delta(X)$, where $x \in X$ is a profile of signals and each agent i observes component x_i. The information-disclosure mechanism coincides with a Blackwell experiment when there is one agent. After information is disclosed, agents interact in a game, whose continuation equilibrium outcomes depend on the information released by the designer. The set of states and the set of actions for each player can be arbitrary finite sets, and we impose no assumption on players' payoff functions. In the baseline model of Section 2, the designer knows the state. In Section 8, we extend our model by allowing the designer to be imperfectly informed about the state.

We follow Myerson (1983) in the formulation of the informed-designer game with the key difference being that here the designer's information is verifiable. The setting is a common value one in Maskin and Tirole's (1992) terminology because the state of the world can affect all players' payoffs. Our setup differs from the usual formulations of informed-principal problems in two ways. First, in contrast to the setting in Maskin and Tirole (1992), and the majority of work on informed-principal problems,[4] the principal has no truth-telling constraints because the state of the world is verifiable. In other words, the mediator implementing the mechanism is omniscient in the language of Forges (1993). Second, the experiment's outputs are signals rather than contractually enforceable outcomes,[5] which makes our game closer to an informed-principal setting with moral hazard.[6]

We identify a set of mechanisms, namely, interim-optimal mechanisms, that always exist (Theorem 1) and constitute (perfect Bayesian) equilibria of the informed-designer game (Theorem 2). We say a mechanism is interim optimal if it is incentive compatible given the prior and there does not exist a belief q and another incentive-compatible mechanism given q such that the expected payoff of every designer type in the support of q is strictly larger than that generated by the original mechanism. A necessary condition for a mechanism to be interim optimal is that each type gets at least his ex-post optimal payoff, namely, the best outcome when agents know the designer's type. But, of course, groups of designer types can also get together and profitably deviate by selecting an alternative "blocking" mechanism that results in an incentive-compatible *allocation* (a vector of interim payoffs for the designer) given a belief q that assigns strictly positive probability *only* to designer types that *strictly* benefit from this deviation. In other words, an interim-optimal mechanism is immune to alternative mechanisms and beliefs when we impose credibility constraints on beliefs, analogous to those imposed by the notions of core in Myerson (1983) and neologism proofness in Farrell (1993). Combined with Theorem 2, this property implies that interim op-

[4]To the best of our knowledge, the exceptions are De Clippel and Minelli (2004), who assume that types are verifiable, and Koessler and Skreta (2019), who allow for general evidence structures.

[5]Our formulation straightforwardly adjusts to accommodate contractible actions, and we do so in Section 8.

[6]See, for example, Wagner et al. (2015) and Mekonnen (2018).

timality is a refinement of perfect Bayesian equilibrium that selects outcomes that are optimal from the informed designer's point of view.

Interim-optimal mechanisms are a tractable class and can be characterized using state-of-the-art techniques. In Proposition 2, we provide a characterization of interim-optimal allocations via the belief-based approach of Kamenica and Gentzkow (2011). We do so in a single-agent setting in which the designer's preferences are state independent, namely, the setting that Lipnowski and Ravid (2020) coin as *transparent motives*.[7] We have already observed that a necessary condition for an allocation to be interim optimal is that each designer type gets a payoff weakly higher than that from full disclosure. Proposition 3 shows this condition is also sufficient when the designer's value function is quasiconvex in beliefs, a property satisfied in many leading economic environments.[8]

Proposition 3, in conjunction with Theorem 2, imply full disclosure (i.e., the "unraveling" outcome, which is the unique equilibrium outcome in the leading games on evidence disclosure) is interim optimal and thus a perfect Bayesian equilibrium outcome even if the informed party can choose arbitrary mechanisms. We leverage Propositions 2 and 3 to build a constrained information-design program that characterizes the designer ex-ante preferred interim-optimal mechanism and allocation, which we call the *interim-optimal solution*. We illustrate this solution in simple examples. We also discuss cases in which the interim optimality constraints do not bind, and hence, the concavification of the designer's value function results in an interim-optimal allocation.

In Section 6, we study interim optimality in multi-agent binary-action settings that include as special cases several environments in the information-design literature, such as the settings in Alonso and Câmara (2016), Arieli and Babichenko (2019), Perez-Richet (2014), and Chan et al. (2019), as well as some parameterized examples in Bergemann and Morris (2019), Taneva (2019), and Mathevet et al. (2020). Proposition 4 establishes that an ex-ante optimal mechanism is interim optimal and thus a perfect Bayesian equilibrium in leading environments in the information-design literature in which agents' actions are strategic complements. Proposition 4 then implies that in a large class of binary-action settings the usual ex-ante commitment assumption in the information-design literature is without loss: the ex-ante optimal mechanism is interim optimal and thus also optimally selected at the interim stage. In Section 6.2, we provide a complete characterization of interim optimality in a binary parametrized setting analogous to the corresponding one in Taneva (2019). We illustrate that ex-ante optimal allocations fail to be interim optimal if actions exhibit strong strategic substitutability, and we discuss how interim optimality compares to ex-ante optimality.

[7] Lipnowski and Ravid (2020) show how the belief-based approach can be used to characterize equilibrium payoffs in cheap-talk sender-receiver games with transparent motives. In contrast to Lipnowski and Ravid (2020), we allow the designer to choose any experiment rather than communicate via cheap-talk messages.

[8] Quasiconvexity of the designer's value function naturally arises in many economic environments, such as settings in which a salesperson discloses information about the quality of the good with the goal to sell more products (as in Grossman, 1981), a manager seeks to motivate the worker to exert maximal effort and the worker exerts higher effort with an increase in the likelihood that the project is promising (as in the application "motivating through strategic disclosure" in Dworczak and Martini, 2019), a job candidate wants to get hired, a politician wants to win office, and so forth. The value function V is also quasiconvex in the investment-recommendation application in Dworczak and Martini (2019) and in the think-tank and broker applications in Lipnowski and Ravid (2020).

Related literature Conceptually and in terms of motivation, interim optimality relates closely to the axiomatic notion of neutral optimum defined in Myerson (1983). It also relates to other notions in the informed-principal literature that identify specific subsets of incentive-compatible mechanisms. Interim optimality relates most closely to the notion of mechanisms with no weak objection defined in De Clippel and Minelli (2004) in a single-agent setting with double-sided verifiable information and no moral hazard. The set of interim-optimal allocations is a subset of core allocations (Myerson, 1983); see Proposition 5. On the other hand, it is a superset of the set of strong neologism-proof allocations (Mylovanov and Tröger, 2012, 2014; Wagner et al., 2015); see Proposition 6. The difference in the sets of core (Myerson, 1983) and that of interim-optimal allocations stems from the beliefs that can accompany alternative mechanism proposals. The set of strong neologism-proof allocations and that of strong unconstrained Pareto optimal allocations (Maskin and Tirole, 1990) both differ from the set of interim-optimal allocations because both concepts allow types that only weakly (rather than strictly) benefit to block. Within the context of information design, the set of core may contain allocations that are not perfect Bayesian (and even Nash) equilibrium allocations (see Example 2). At the same time, strong neologism-proof and strong unconstrained Pareto optimal allocations may fail to exist in our setting (see Example 5). Thus, in general, the concepts of core, strong neologism proofness, and strong unconstrained Pareto optimality are not suitable for the informed-designer problem.[9]

A small set of papers studies information design at the interim stage. In a pioneering paper, Perez-Richet (2014) studies equilibrium refinements and constrained-information policies in a single-agent setting with binary actions and states, and state-independent utilities for the designer.[10] Hedlund (2017) considers a binary-state sender-receiver game in which the sender's type is his belief about the payoff-relevant state. The sender chooses a Blackwell experiment whose outputs only depend on the payoff-relevant state. Thus, Hedlund's (2017) model is casted as a signaling game. By contrast, we allow for general mechanisms, mappings from *both* the state and the designer's private information to a probability distribution over signals. Hedlund (2017) shows equilibrium outcomes that satisfy D1 are either fully disclosing (the experiment fully reveals the state) or fully separating (the choice of experiment reveals the state). In Chen and Zhang (2020), the principal's type is binary and it indexes the distribution of the buyer's values. As in Hedlund (2017), the designer in Chen and Zhang (2020) (the seller) offers experiments that provide information about the buyer's value. They show private information hurts the seller. The aforementioned papers examine single-agent settings. Eliaz and Serrano (2014) analyze a setting where an informed planner discloses information about the state to two interacting agents who play state-dependent versions of a prisoner's dilemma game. The planner knows the state and

[9]The notions of Rothschild-Stiglizt-Wilson allocation of Maskin and Tirole (1992) and that of assured allocation of Balkenborg and Makris (2015) do not seem to have analogous versions in our setting. Both concepts are defined in settings in which all decisions are contractible, there are transfers and types are ordered.

[10]Without putting some reasonable restrictions on beliefs, perfect Bayesian equilibrium has very little predictive power in such settings because off-path beliefs can be chosen in a way that completely cancels out the information revealed by off-path experiments. Hence, the worst action for the designer can be induced after every deviation by the designer. This observation generalizes to some information-design settings with state-independent preferences for the designer, where the ex-ante optimal mechanism is a perfect Bayesian equilibrium mechanism even when it is strictly dominated by the ex-post optimal mechanism for some designer types (see Zapechelnyuk, 2022).

sends each agent a private signal that consists of any subset of states that contains the true one.

Clearly, when the designer has access to all experiments, the optimum can be achieved by choosing and committing to the experiment before knowing the state, which is not generally true if the set of experiments is restricted. Degan and Li (2021) make this point in a binary-state, binary-action setting in which the informed sender chooses the signal's precision. Alonso and Câmara (2018) focus on whether the designer can benefit from having private information prior to offering an experiment in a setting in which the designer may have access to a limited set of experiments. By contrast, in this paper, we consider an interim information-design setting with an arbitrary number of states, actions, agents, and general payoffs, in which the informed designer can choose any disclosure mechanism.

The rest of the paper is structured as follows. Section 2 describes the setting and defines ex-ante and ex-post optimal mechanisms. In Section 3, we define interim-optimal mechanisms and prove that they exist. Section 4 formulates the informed-designer game and establishes that interim-optimal mechanisms are perfect Bayesian equilibrium outcomes of that game. Section 5 provides a belief-based characterization of interim optimality. In Section 6, we study interim optimality in multi-agent settings. In Section 7, we compare interim optimality to other leading concepts in the informed-principal literature. In Section 8, we present the general model in which the designer can be imperfectly informed about the state. Section 9 concludes. We prove Theorem 1 and Theorem 2 directly for the general setting of Section 8 in Appendix A.1 and Appendix A.2, respectively.

2 Model

We consider an incomplete-information environment with $n+1$ players. Player 0 is the *information designer* who interacts with n players called *agents*. We denote by $I = \{1, \dots, n\}$ the set of agents. Each agent $i \in I$ has a non-empty and finite set of actions A_i. Let $A = \prod_{i \in I} A_i$ be the set of action profiles.

The designer is privately informed about the state of the world that affects players' payoffs.[11] Let T be the non-empty and finite set of states, which is the set of types of the designer. The common prior $p \in \Delta(T)$ is assumed to have full support. For every action profile $a \in A$ and type $t \in T$, the utility of the designer is $u_0(a, t)$ and the utility of agent i is $u_i(a, t)$. Following the terminology of Myerson (1982, 1983), the setting above is called a *Bayesian incentive problem* and is denoted by

$$\Gamma = ((A_i)_{i \in I}, (u_i)_{i=0}^n, T, p).$$

Γ also corresponds to the basic game as defined in Bergemann and Morris (2019).

A *direct mechanism* is a mapping $\mu : T \to \Delta(A)$, where $\mu(a_1, \dots, a_n \mid t)$ is the probability that the mechanism privately recommends a_i to each agent i when the actual type of the designer is t.

[11] To keep the exposition focused in the main text we present definitions and results for this baseline setting in which the designer is perfectly informed about the state but all definitions readily extend to the general setting of Section 8.

Let $q \in \Delta(T)$ be a common belief for the agents. The mechanism μ is *q-incentive compatible* (q-IC) iff for each agent i obedience (following every recommendation a_i) is optimal if all the other agents are obedient; that is,

$$\sum_{a_{-i} \in A_{-i}} \sum_{t \in T} q(t)\mu(a \mid t)[u_i(a,t) - u_i((a_i', a_{-i}), t)] \geq 0, \text{ for every } a_i \text{ and } a_i' \text{ in } A_i.$$

The mechanism μ is *incentive compatible* (IC) if it is p-IC, so it is incentive compatible for the prior. The set of IC mechanisms is the set of Bayes correlated equilibrium mechanisms (Bergemann and Morris, 2016, Taneva, 2019).[12]

A mechanism μ is *ex-post incentive compatible*[13] iff for every $i \in I$ and $t \in T$, we have

$$\sum_{a_{-i} \in A_{-i}} \mu(a \mid t)[u_i(a,t) - u_i((a_i', a_{-i}), t)] \geq 0, \text{ for every } a_i \text{ and } a_i' \text{ in } A_i.$$

An ex-post IC mechanism satisfies the agents' obedience constraints when they know the state and it is q-IC for *every* $q \in \Delta(T)$. It maps every t to a correlated equilibrium (Aumann, 1974) of the n-player normal form game $(I, (A_i)_{i \in I}, (u_i(\cdot, t))_{i \in I})$. An ex-post IC mechanism always exists in our environment because the set of correlated equilibria is non-empty and the mediator is omniscient: the state that is inputed in the mechanism is verifiable as in Kamenica and Gentzkow (2011), Bergemann and Morris (2016), and Taneva (2019).[14]

Let

$$U_0(\mu \mid t) = \sum_{a \in A} \mu(a \mid t) u_0(a,t),$$

denote the interim expected utility of the designer at state t from mechanism μ when agents are obedient. We call *allocation* the corresponding vector of designer utilities $(U_0(\mu \mid t))_{t \in T}$. Let $\boldsymbol{U}(q) \subseteq \mathbb{R}^T$ be the set of q-IC allocations for the designer:

$$\boldsymbol{U}(q) := \{U \in \mathbb{R}^T : U = (U_0(\mu \mid t))_{t \in T} \text{ and } \mu \text{ is } q\text{-IC}\}.$$

We now formally define the concepts of ex-ante optimal and ex-post optimal mechanisms.

Definition 1 A mechanism μ is *ex-ante optimal* (EAO) iff μ is incentive compatible and for every other incentive-compatible mechanism ν, we have

$$\sum_{t \in T} p(t) U_0(\mu \mid t) \geq \sum_{t \in T} p(t) U_0(\nu \mid t).$$

An EAO mechanism corresponds to a solution of the standard information-design problem (Kamenica and Gentzkow, 2011, Bergemann and Morris, 2019, Taneva, 2019) and is the natural optimal solution for an uninformed designer.

[12] And it is the set of Bayesian solutions (Forges, 1993, 2006).

[13] Such a mechanism is called *safe* in Myerson (1983) and *full-information* incentive compatible in Maskin and Tirole (1990).

[14] An ex-post IC mechanism also exists in the private-value environments with unverifiable types of Maskin and Tirole (1990) and of Mylovanov and Tröger (2014). In the general model of Myerson (1983), the designer's types are unverifiable and an ex-post IC mechanism may *not exist* because a mechanism that is ex-post IC for the agents may not satisfy the designer's truth-telling constraints.

Definition 2 A mechanism μ is *ex-post optimal* (EPO) iff μ is ex-post incentive compatible and for every other ex-post incentive-compatible mechanism ν, we have

$$U_0(\mu \mid t) \geq U_0(\nu \mid t), \text{ for every } t \in T.$$

The EPO allocation is the best correlated equilibrium allocation for the designer when t is commonly known and it is the natural optimal solution for the designer under complete information.

In the next section, we present the notion of interim-optimal mechanisms and explain why it is a natural optimal solution when the designer is privately informed about t.

3 Interim-optimal mechanisms

Our goal is to identify a tractable subset of IC mechanisms (i.e., Bayes correlated equilibria) that are the best an informed designer can credibly select, in the sense that a designer type cannot credibly deviate and increase his payoff at the interim stage. Because each designer type is able to implement any ex-post IC mechanism by fully revealing the state, our notion of interim optimality requires that a mechanism guarantees each designer type a utility weakly higher than his EPO utility. By an analogous token, a group of designer types should not be able to select an alternative mechanism that strictly benefits all members of the group by revealing that the state belongs to that group. The set of interim-optimal mechanisms that we define next is robust to such "blockings" in the sense we now make precise:

Definition 3 A mechanism $\mu : T \to \Delta(A)$ is *interim optimal* (IO) iff μ is incentive-compatible and no mechanism ν and belief q exist such that ν is q-IC and $U_0(\nu \mid t) > U_0(\mu \mid t)$ for every $t \in \text{supp}[q]$.

In other words, an IC mechanism μ is IO iff no group of designer types can find a belief q that assigns strictly positive probability only to types in that group and some IC mechanism ν given belief q that strictly benefits every member of the group. The notion of interim optimality is strong for two reasons. First, the continuation equilibrium of any alternative mechanism ν (given belief q) is *optimally* selected. This property makes interim optimality comparable to ex-ante and ex-post optimality. In particular, if only one possible designer type exists ($|T| = 1$), the definition of interim optimality coincides with the definition of ex-ante and ex-post optimality.[15] Second, the designer is able to choose any belief q that satisfies the following credibility requirement: if the designer selects an alternative mechanism ν, the agents assign positive probability *only* to designer types who strictly benefit from the alternative mechanism ν.

The second property of IO mechanisms implies they are robust to evidence disclosure: if a mechanism is IO, no subset S of designer types exists such that all types in S strictly benefit from disclosing S to the agents, whatever the agents' consistent inference. The fact that interim optimality is a strong selection of IC mechanisms implies positive results are strong as well: Theorem 1 below shows that an IO mechanism always exists.

[15]More generally, in the generalized setting of Section 8, interim optimality coincides with ex-ante and ex-post optimality if the designer is uninformed, even with multiple states of the world.

When an EAO mechanism turns out to be IO (Proposition 1 and Proposition 4 below), the ex-ante commitment solution of standard information design is implementable as a perfect equilibrium of the informed-designer game (see the next section) and the designer cannot credibly select a better mechanism.[16] Likewise, when an EPO mechanism (the *unraveling* outcome) is IO (see Corollary 1), it satisfies the properties mentioned above. Theorem 2 in the following section shows an IO mechanism has, in addition, a solid game-theoretic foundation because it is a perfect Bayesian equilibrium of an informed-designer game.

As discussed in the introduction, the notion of interim optimality is closely related to other notions introduced in the informed-principal literature (see Section 7 for more details). In particular, the set of IO mechanisms is a subset of core mechanisms (Myerson, 1983). Although the set of core mechanisms is non-empty, it has two drawbacks: it includes mechanisms that are not equilibrium mechanisms of the informed-designer game (see Example 2 and the related comment on page 26) and it is not immune to payoff-irrelevant decompositions of types (see the related comment in Section 7.1). On the other hand, notions such as strong unconstrained Pareto optimal mechanisms (Maskin and Tirole, 1990) or strong neologism-proof mechanisms (Mylovanov and Tröger, 2012, 2014) do not always exist.

We proceed to illustrate the concept of interim optimality and compare it with that of ex-ante optimality in a series of simple examples.

Example 1 (State-independent preferences, one agent) Suppose that there is only one agent, two states $T = \{1, 0\}$ with prior $p(1) = p = \frac{1}{6}$, the set of actions of the agent is a subset of $\{\underline{a}^0, a^2, a^3, \bar{a}^0\}$, and the designer has state-independent utilities: $u_0(\underline{a}^0) = u_0(\bar{a}^0) = 0$, $u_0(a^2) = 2$, $u_0(a^3) = 3$. The three examples that follow differ in the set of actions of the agent. For brevity, we only describe the agent's optimal action as a function of the belief and do not provide the payoff details that lead to that function.

(i) Binary actions. In the first example, which is similar to the judge example in Kamenica and Gentzkow (2011), the agent has two possible actions, $A = \{\underline{a}^0, a^2\}$, and the optimal action as a function of his belief $q(1) = q$ is $\underline{a}^0$ if $q < 1/3$ and a^2 if $q > 1/3$. The designer's expected utility as a function of q is denoted by $V(\cdot)$ and is depicted in Figure 1. The EAO mechanism is obtained from the concavification of the value function V, denoted by cav V and depicted with the dashed line in Figure 1.[17] It splits the prior $p = \frac{1}{6}$ uniformly to the posteriors $q = 0$ and $q = \frac{1}{3}$, which induces the EAO allocation $U = (U(1), U(0)) = (2, \frac{2}{5})$. The EAO allocation is IO in this example: designer type $t = 1$ gets his first-best, so belief credibility in Definition 3 implies $q = 0$. The only q-IC utility allocation for type $t = 0$ is then 0, which is lower than $\frac{2}{5}$. More generally, Proposition 4 that follows implies an EAO allocation is IO in any single-agent setting with binary actions and state-independent preferences for the designer.[18]

[16] As discussed in footnote 10 and in the next section, many perfect Bayesian equilibria could exist that are based on adversarial beliefs and continuation equilibria.

[17] The concavification of V is the smallest concave function that is pointwise greater than or equal to V.

[18] However, binary-action settings exist in which ex-ante optimality does not imply interim optimality; see Example 2 below.

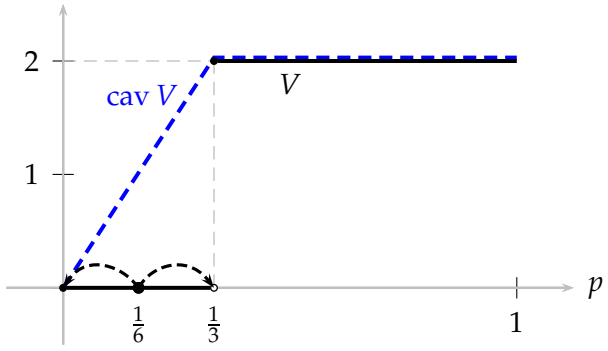

Figure 1: Designer value function (solid lines) and ex-ante utility at the EAO mechanism (dashed lines) in Example 1 (i).

(ii) Three actions. The next example, which is similar to the central bank example in Lipnowski et al. (2022), arises from adding action a^3 to the previous example. Then, the action set becomes $A = \{\underline{a}^0, a^2, a^3\}$. We assume a^3 is the unique optimal action for the agent when $q > \frac{2}{3}$. See the left panel of Figure 2. With a prior $p = \frac{1}{6}$ the EAO mechanism remains unchanged. However, in this three-action setting it is not IO: it is blocked by type $t = 1$, who offers a fully revealing mechanism (the EPO one) yielding $U' = (3, 0)$. In this version of the example, the set of IO allocations is the set of interim utility vectors U for the designer such that $U(1) = 3$ and $U(0) \in [0, \frac{3}{10}]$. The designer ex-ante preferred IO allocation is $(U(1), U(0)) = (3, \frac{3}{10})$ and is obtained from the following direct recommendation mechanism:

This mechanism splits the prior to the posterior $\frac{2}{3}$ with probability $\frac{1}{4}$ and to the posterior 0 with probability $\frac{3}{4}$. The corresponding ex-ante expected utility for the designer is $\frac{3}{4}$, which is strictly lower than the EAO utility cav $V(\frac{1}{6}) = 1$.

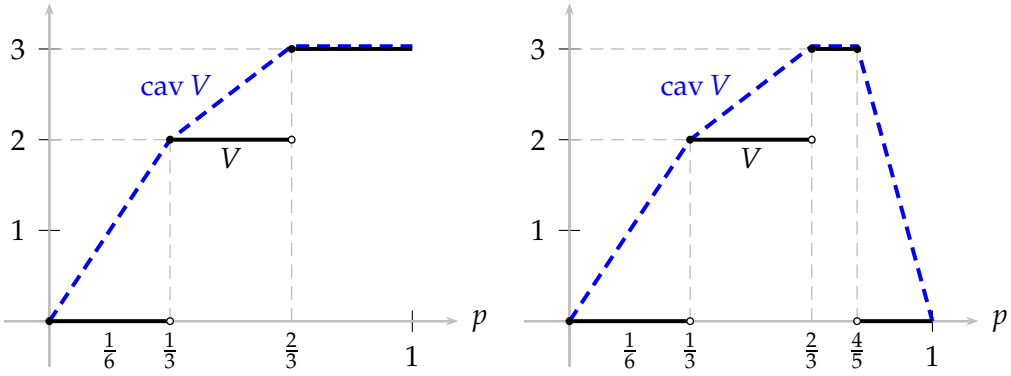

Figure 2: Designer value function (solid lines) and ex-ante utility at the EAO mechanism (dashed lines). Left panel: Example 1 (ii). Right panel: Example 1 (iii).

(iii) Four actions. In the last variation, we add a fourth action, $\bar{a}^0$, so the action set becomes $A = \{\underline{a}^0, a^2, a^3, \bar{a}^0\}$. We assume $\bar{a}^0$ is the unique optimal action for the agent

for $q > \frac{4}{5}$. Contrary to the two cases analyzed above, the ideal action for the designer is only obtained for interior beliefs. See the right panel of Figure 2. For the prior $p = \frac{1}{6}$, the EAO mechanism is the same as before, but it is not IO. It is blocked by *both* types offering a pooling mechanism ν for $q \in [\frac{2}{3}, \frac{4}{5}]$ that recommends for both types action a^3 and yields $U' = (3, 3)$.[19]

Theorem 1 that follows establishes existence of IO mechanisms for every Bayesian incentive problem Γ. We prove this result in Appendix A.1 for the general setting introduced in Section 8. The proof does not impose any additional assumptions on any of the elements of Γ.

Theorem 1 *For any Bayesian incentive problem Γ, at least one interim-optimal mechanism exists.*

The idea of the proof lies in establishing that a neutral optimum (as defined in Myerson, 1983, but without truth-telling conditions for the designer) is IO. Neutral optima exist by the same arguments as in the proof of Theorem 6 in Myerson (1983). To relate interim optimality with neutral optimum, we define interim optimality in terms of "blocked allocations" as follows.

Let $B^{IO}(\Gamma)$ be the set of allocations $U \in \mathbb{R}^T$ such that a belief $q \in \Delta(T)$ and a q-IC allocation U' exist such that $U'(t) > U(t)$ for every $t \in \text{supp}[q]$. By definition, an allocation U is an IO allocation iff it is IC and $U \notin B^{IO}(\Gamma)$. The proof shows $B^{IO}(\Gamma)$ satisfies the axioms of *Domination*, *Openness*, *Extensions* and *Strong solutions*, which establishes that the set of neutral optima is included in the set of IO allocations, and thus, the set of IO allocations is non-empty. These axioms also characterize desirable properties of IO allocations. These axioms are defined in Myerson (1983) and, for completeness, we include their formal definitions in Appendix A.1.

The next proposition shows that if the EPO allocation is EAO, it is the unique IO allocation. In this sense, interim optimality is an in-between notion consistent with ex-ante and ex-post optimality.

Proposition 1 *If the ex-post optimal allocation is ex-ante optimal, then it is the unique interim-optimal allocation.*

Proof. To prove an allocation U^* that is both EPO and EAO is also IO, we use several auxiliary results in Appendix A.1. Because the allocation U^* is EAO, it is undominated (Definition A.7). Hence, if it is EPO, it is ex-post IC, and therefore, it is a strong solution (Definition A.8). We conclude from Proposition A.7 that shows that a strong solution is IO. To show uniqueness of the IO allocation, let U^{IO} be an IO allocation, and assume by way of contradiction that $U^{IO} \neq U^*$. Because U^* is EAO, t exists such that $U^{IO}(t) < U^*(t)$. But U^* is also EPO, so by definition of interim optimality, we have $U^{IO}(t) \geq U^*(t)$ for every t, a contradiction. ∎

In other words, Proposition 1 shows that if an EAO allocation can be obtained by a fully revealing mechanism, this fully revealing mechanism is IO and the corresponding allocation is the unique IO allocation.[20] In particular, the IO allocation is unique

[19] In this example, the EAO mechanism is a core mechanism (see Section 7.1).

[20] Observe that the EPO allocation is always unique but multiple EAO allocations may exist.

when for every t, a correlated equilibrium of the complete-information game at t exists that gives the first-best utility to the designer.

4 Informed-designer game

In this section, we describe the informed-designer game and establish our second main result, Theorem 2, that shows an IO mechanism is a perfect Bayesian equilibrium mechanism of this game. The informed-designer game is the following extensive-form game between the designer and the agents:

1. Nature selects the state of the world, $t \in T$, according to the prior probability distribution $p \in \Delta(T)$;
2. The designer is privately informed about $t \in T$;
3. The designer chooses a non-empty and finite set of signals[21] $X = \prod_{i \in I} X_i$ and an information-disclosure *mechanism*

$$\nu : T \to \Delta(X);$$

4. Agents publicly observe the mechanism ν proposed by the designer;
5. Signals $(x_1, \ldots, x_n)$ are drawn with probability $\nu(x_1, \ldots, x_n \mid t)$. For every i, signal x_i is privately observed by agent i;
6. Every agent i chooses an action $a_i \in A_i$ as a function of his signal $x_i \in X_i$.

The key difference between this game and the usual formulation of information design (as in Bergemann and Morris, 2019) or Bayesian persuasion (Kamenica and Gentzkow, 2011) is that in those settings, the designer is not informed about t (i.e., stage 2 in the description above is absent). That is, he designs an information structure ex ante. This setting corresponds to a mechanism-design problem with verifiable types (an omniscient mediator), and a version of the revelation principle applies (Myerson, 1982, Forges, 1993, Forges and Koessler, 2005, Bergemann and Morris, 2019). By contrast, in the extensive-form game described above, the choice of the mechanism is at the interim stage, so it is an informed-principal problem with verifiable types. Because the revelation principle cannot be applied off the equilibrium path, we allow the designer to choose mechanisms in stage 3 with arbitrary signals as outputs, not just direct recommendation mechanisms.

The informed-designer game is also related to the game studied in the literature on strategic information disclosure as in Milgrom (1981). As we mentioned in the introduction, in this literature, the informed party chooses which piece of evidence to disclose (formally, a message from a type-dependent set of messages), whereas in our setting, the informed party can choose any stochastic information-disclosure mechanism.

[21] Formally, we can define for every i any superset of A_i, and assume the designer chooses a finite subset X_i of that superset.

The extensive-form game we analyze is complex; the designer has private information as in signaling games and, more importantly, the designer's choice set is rich because he chooses disclosure mechanisms (functions from the state to distributions over signals). Our formulation of perfect Bayesian equilibrium relies on a number of auxiliary results due to Myerson (1983) that we include in Appendix A.2 before we prove Theorem 2. That proof establishes that an IO mechanism is an expectational equilibrium as defined in Myerson (1983), which is a refinement of perfect Bayesian equilibrium in the spirit of sequential equilibrium.

Theorem 2 *If μ is an interim-optimal mechanism, then μ is a perfect Bayesian equilibrium mechanism of the informed-designer game.*

The next example presents a binary-state, binary-action setting with a single agent in which the EAO mechanism is a core mechanism as defined in Myerson (1983), but it is not a (Nash) equilibrium of the informed-designer game. Consequently, by Theorem 2, this mechanism is not IO.

Example 2 (State-dependent preferences, two actions) In this example, like in the first, the designer faces one agent. There are two possible states $T = \{t_1, t_2\}$ and two actions for the agent $A = \{a^1, a^2\}$. We show that when the prior is $p(t_1) = p = 3/4$, a profitable deviation from the EAO mechanism exists regardless of the continuation strategy of the agent. The designer's and the agent's utilities are summarized in the following matrix:

	a^1	a^2
t_1	$3,0$	$0,1$
t_2	$0,1$	$1,0$

Let $q(t_1) = q$ denote the belief of the agent that the designer's type is t_1. The unique optimal action for the agent is to choose a^1 if $q < 1/2$ and a^2 if $q > 1/2$. The designer's highest ex-ante expected utility as a function of q is

$$V(q) = \begin{cases} 3q & \text{if } q \leq 1/2 \\ 1-q & \text{if } q > 1/2. \end{cases}$$

When the prior is $p = 3/4$, the EAO mechanism splits uniformly the prior $p = \frac{3}{4}$ to the posteriors $\frac{1}{2}$ and 1. The corresponding direct-recommendation mechanism $\mu : T \to \Delta(A)$ is

$$\mu(a^1 \mid t_1) = 1/3; \mu(a^2 \mid t_1) = 2/3; \mu(a^1 \mid t_2) = 1; \mu(a^2 \mid t_2) = 0.$$

Then, the posterior belief of the agent is $\Pr(t_1 \mid a^2) = 1$, $\Pr(t_1 \mid a^1) = 1/2$ as desired. The EAO allocation is $U = (1,0)$. It is immediate to see this allocation is not a Nash equilibrium allocation of the informed-designer game. The designer can deviate to any pooling mechanism (an experiment that sends the same signal regardless of the state). Suppose that given such a mechanism, the agent chooses a^1 with probability β and a^2 with probability $1-\beta$. If $\beta > \frac{1}{3}$, t_1 strictly benefits, and if $\beta < 1$, t_2 strictly benefits, implying that at least one of the two designer types benefits regardless of the value of β. In this example, neither the EAO allocation $U = (1,0)$ nor the EPO

allocation $(0,0)$ are IO because they are not equilibrium allocations.[22] The allocation $U = (0,1)$, which is simply obtained by a non-revealing experiment, is an IO allocation and a perfect Bayesian equilibrium allocation by Theorem 2.

Interim optimality and equilibrium refinements Interim optimality is a notion that only relies on incentive-compatibility notions. The definition of interim optimality does not refer to any informed-designer game or to its set of perfect Bayesian equilibrium assessments.[23] However, Theorem 2 implies interim optimality is a refinement of perfect Bayesian equilibrium.

The set of IO mechanisms is usually much smaller than the set of perfect Bayesian equilibrium mechanisms, because interim optimality gives all the bargaining power to the informed designer. As an illustration, consider the two following examples. In the first example, agents' beliefs play no role, but adversarial equilibrium selection enables support of every IC mechanism as a perfect Bayesian equilibrium. The designer has only one possible type (or multiple payoff-irrelevant types) and there are multiple agents. Assume $\underline{a} \in A$ is a Nash equilibrium in the complete-information game played by the agents, and that $\underline{a}$ is the least-preferred action profile for the designer. In such a situation, the IO mechanism boils down to the designer-optimal correlated equilibrium, exactly as the EAO mechanism: the designer privately recommends players to play according to this correlated equilibrium, which is IC. By contrast, every correlated equilibrium, in particular, the one that induces the worst outcome $\underline{a}$ for the designer with probability 1, constitutes a perfect Bayesian equilibrium.

As a second example, consider Example 1 (iii), where adversarial equilibrium selection plays no role, because the agent has a unique optimal action for (almost all) possible beliefs. As seen previously, in every IO allocation at least one designer's type utility is equal to 3. By contrast, *every* IC allocation, that is, every information-disclosure mechanism, constitutes a perfect Bayesian equilibrium regardless of the prior. In particular, the worst allocation $(0,0)$ for the designer is a perfect Bayesian equilibrium allocation.

Common refinements in signaling games in the spirit of the intuitive criterion (Cho and Kreps, 1987) or D1 (Banks and Sobel, 1987), are, in general, weaker than interim optimality. In particular, in the first example above, off-path beliefs play no role, so $\underline{a}$ is an equilibrium outcome satisfying such refinements. In addition, as discussed after the definition of interim optimality, the notion of blocking in the definition of interim optimality allows the designer to choose a belief under a credibility notion. By contrast, common refinements in signaling games only put restrictions on the set of off-path beliefs for the agents, and, in equilibrium, those beliefs could be chosen, under credibility, to punish the designer. In Example 1 (iii), every IC allocation, in particular, the worst allocation $(0,0)$, survives such refinements regardless of the prior.

[22] This finding is in contrast to the setting in De Clippel and Minelli (2004), where the EPO allocation (and any IC allocation that gives higher payoff to each designer type) is a perfect Bayesian equilibrium allocation. The difference stems from the fact that in De Clippel and Minelli (2004), the agent simply accepts or rejects the mechanism proposed by the informed principal.

[23] The same remark applies to most solution concepts in the informed-principal literature.

5 Belief-based characterization of interim optimality

In Bayesian persuasion, Kamenica and Gentzkow (2011) write the designer's ex-ante payoff as a function of beliefs by incorporating the agent's optimal action which is a function of beliefs. The concavification of the resulting value function, denoted by V, yields an EAO mechanism in terms of an optimal splitting (a distribution of posterior beliefs that average to the prior) without explicitly using the revelation principle and obedience constraints. This elegant approach, based on Aumann and Maschler (1995), has proved powerful and has been broadly applied.

In this section, we provide an analogous belief-based characterization of interim optimality. We assume a single agent is present.[24] We also assume the utility of the designer is state independent: for every state t and action a, the payoff of the designer is equal to $u_0(a)$.[25]

For every $q \in \Delta(T)$, let $A^*(q)$ be the set of optimal actions of the agent when his belief is q:

$$A^*(q) := \arg\max_{a \in A} \sum_{t \in T} q(t) u_1(a, t).$$

For every $q \in \Delta(T)$, let $a^*(q) \in \arg\max_{a \in A^*(q)} u_0(a)$; that is, $a^*(q)$ is a designer-preferred selection among the agent's optimal actions at belief q. For every $q \in \Delta(T)$, let $V(q)$ be the highest utility of the designer when the agent's belief is q:

$$V(q) := u_0(a^*(q)).$$

We start with a preliminary result, Lemma 1, that we employ to prove the two main results of this section.

Lemma 1 *If U is a q-IC allocation, then there exists $\tilde{q} \in \Delta(T)$ with $\operatorname{supp}[\tilde{q}] \subseteq \operatorname{supp}[q]$ such that $U(t) \leq V(\tilde{q})$ for all $t \in \operatorname{supp}[q]$.*

Proof. Let U be a q-IC allocation and let μ be the corresponding mechanism. Let

$$\bar{a} \in \arg \max_{a \in \bigcup_{t \in T} \operatorname{supp}[\mu(\cdot|t)]} u_0(a). \tag{1}$$

Let $\tilde{q} \in \Delta(T)$ be the posterior belief of the agent when he gets recommendation $\bar{a}$ under the mechanism μ: for every t,

$$\tilde{q}(t) = \frac{\mu(\bar{a} \mid t) q(t)}{\sum_{\tilde{t}} \mu(\bar{a} \mid \tilde{t}) q(\tilde{t})},$$

and we have $\operatorname{supp}[\tilde{q}] \subseteq \operatorname{supp}[q]$. Because μ is q-IC, we have $\bar{a} \in A^*(\tilde{q})$. Hence,

$$u_0(\bar{a}) \leq \max_{a \in A^*(\tilde{q})} u_0(a) = V(\tilde{q}).$$

[24] The characterizations we provide apply to settings with multiple agents when the designer is restricted to *public* disclosures. In that case, the same belief-based approach applies by replacing V with the highest utility the designer can get when agents play a Bayes-Nash (instead of Bayes correlated) equilibrium of the symmetric-information game given belief q.

[25] The characterizations below readily extend to settings in which only the ordinal preference of the designer is state independent. State-independent utility for the designer is a common assumption in the literature; see, for example, Dworczak and Martini (2019) and Lipnowski and Ravid (2020).

Then, together with Equation (1), this implies

$$U(t) = \sum_a \mu(a \mid t) u_0(a) \leq u_0(\bar{a}) \leq V(\tilde{q}),$$

for every $t \in \text{supp}[q]$. ∎

The next proposition relies on Lemma 1 to provide a belief-based characterization of interim optimality.

Proposition 2 (Belief-based characterization of interim optimality) *Assume that there is a single agent and that the utility of the designer is state independent. Then, an incentive-compatible allocation $U \in \mathbb{R}^T$ is interim optimal iff*

$$\text{There is no } q \in \Delta(T) \text{ such that } V(q) > U(t) \text{ for every } t \in \text{supp}[q]. \tag{2}$$

Proof. ($\Rightarrow$) Let $U \in \mathbb{R}^T$ be an IC allocation and assume a $q \in \Delta(T)$ exists such that $V(q) > U(t)$ for every $t \in \text{supp}[q]$. Then, the allocation U', with $U'(t) = V(q)$ for every $t \in T$, is q-IC: it corresponds to a non-revealing mechanism in which action $a^*(q)$ is chosen by the agent for every $t \in T$. Hence, $U'(t) > U(t)$ for every $t \in \text{supp}[q]$, which implies U is not IO.

($\Leftarrow$) Let $U \in \mathbb{R}^T$ be an IC allocation and assume it is not IO; that is, $\tilde{q}$ and a $\tilde{q}$-IC allocation U' exist such that $U'(t) > U(t)$ for every $t \in \text{supp}[\tilde{q}]$. The fact that U' is $\tilde{q}$-IC implies by Lemma 1 that $q \in \Delta(T)$ with $\text{supp}[q] \subseteq \text{supp}[\tilde{q}]$ exists such that $V(q) \geq U'(t)$ for all $t \in \text{supp}[\tilde{q}]$. Hence, $V(q) > U(t)$ for every $t \in \text{supp}[q]$. ∎

Proposition 2 implies that to check whether an IC allocation U is IO, it suffices to check a finite number of inequalities that only rely on (2), that is, on the comparison of U to the value function V. To see condition (2) can be rewritten as a finite number of inequalities, for every $S \subseteq T$, let $u^*(S)$ be the highest utility of the designer from an action that is optimal for the agent for some belief with support S. Formally,

$$u^*(S) = \max\{u_0(a) : \exists\, q \in \Delta(T), \text{ such that } \text{supp}[q] = S \text{ and } a = a^*(q)\}.$$

Then, (2) is equivalent to the following:

$$\text{for every } S \subseteq T, \text{ there exists } t \in S \text{ such that } U(t) \geq u^*(S). \tag{IOC}$$

In addition, as in Kamenica and Gentzkow (2011), when the agent breaks ties in favor of the designer (as is the case in an IO mechanism), any IC allocation U can be fully characterized in terms of a splitting of the prior, that is, a probability distribution σ over the set of posteriors $\Delta(T)$ such that the expected value of the posterior is equal to the prior. Because the set of actions is finite, it is without loss of generality to focus on splittings (distributions of posteriors) with finite support (of cardinality at most $|A|$). Any such splitting of p can be represented by $\sigma = (\lambda_k, q_k)_k$ where for every $k = 1, \ldots, |A|$, λ_k is the probability of posterior $q_k \in \Delta(T)$, and $\sum_k \lambda_k q_k = p$. Let $\Sigma(p)$ be the set of all such splittings of p. By Bayes' rule, we get for every $t \in T$

$$U(t) = \sum_k \frac{\lambda_k q_k}{p(t)} u_0(a^*(q_k)).$$

Therefore, we can leverage Proposition 2 to build the following constrained information-design program that characterizes the designer ex-ante preferred IO mechanism, which we call the *IO solution* in what follows:

$$\max_{\sigma \in \Sigma(p)} E_\sigma[V(q)] \text{ subject to (IOC).} \tag{P}$$

This program is a *constrained concavification* problem, that is, an optimal splitting problem under the interim-optimality constraints (IOC). Without the (IOC) constraints, this is simply the program characterizing EAO mechanisms, which yields

$$\max_{\sigma \in \Sigma(p)} E_\sigma[V(q)] = \operatorname{cav} V(p).$$

For illustration, let us consider Example 1 (iii) with four actions. We have

$$u^*(\{1\}) = u^*(\{0\}) = 0 \text{ and } u^*(\{1,0\}) = 3.$$

Hence, the set of IO allocations is the set of allocations induced by some splitting of the prior such that $U(t) = 3$ for some $t \in \{1,0\}$. This condition implies the support of the corresponding splitting σ is either $\{0, \bar{q}\}$, or $\{\bar{q}, 1\}$, or $\{\bar{q}\}$, with $\bar{q} \in [\frac{2}{3}, \frac{3}{4}]$. Therefore, whatever the (interior) prior p, the EPO allocation $(0,0)$ is not IO, and the EAO allocation is IO iff $p \geq \frac{2}{3}$. In Figure 3, we depict in dotted lines the ex-ante utility of the designer at the IO solution.

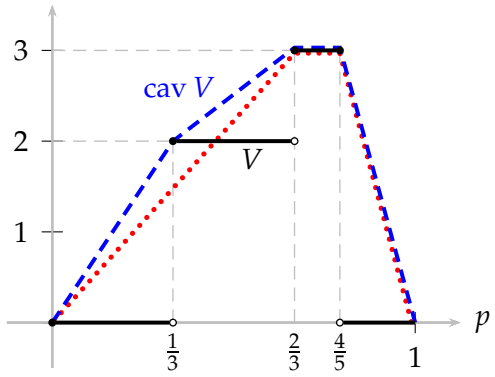

Figure 3: Designer ex-ante utility at the EAO mechanism (dashed lines) and the IO solution (dotted lines) in Example 1 (iii).

We now proceed to provide simpler necessary and sufficient conditions for interim optimality for settings in which the designer's value function $V(\cdot) = u_0(a^*(\cdot))$ is quasiconvex.[26] When there are two possible types, quasiconvexity of V (written as a function of the probability of one of the two types) means V is weakly increasing, weakly decreasing, or weakly decreasing and then weakly increasing. More generally, when $V(q)$ only depends on the mean of $t \in T \subseteq \mathbb{R}^K$ (as in, e.g., Dworczak and Martini, 2019), that is, it can be written as $V(q) = g(E_q(t))$, V is quasiconvex if g is quasiconvex. Note that in Example 1, V is quasiconvex in the two-action and three-action cases, but not in the four-action case. As discussed in the introduction, quasiconvexity of V naturally arises in many economic environments.

[26]The function $V : \Delta(T) \to \mathbb{R}$ is quasiconvex if its lower contour sets $\{q \in \Delta(T) : V(q) \leq y\}$ are convex sets.

Regardless of the properties of V, a necessary condition for U to be IO is that each type gets at least his EPO allocation. This necessary condition obtains from Proposition 2 by noting $V(\delta_t)$ is the EPO payoff of type t, and we must have $U(t) \geq V(\delta_t)$ for every t. The next proposition shows this condition is also sufficient when V is quasiconvex. As a consequence, in settings with quasiconvex V, the characterization of IO allocations drastically simplifies.

Proposition 3 (Belief-based characterization of interim optimality: quasiconvex V) *Assume that there is a single agent, that the utility of the designer is state independent, and that $V(q)$ is quasiconvex. Then, an incentive-compatible allocation $U \in \mathbb{R}^T$ is interim optimal iff $U(t) \geq V(\delta_t)$ for all $t \in T$. That is, an incentive-compatible allocation U is interim optimal iff each type of the designer gets at least his ex-post optimal allocation. In particular, the ex-post optimal allocation is interim optimal.*

Proof. It suffices to show $U = (V(\delta_t))_{t\in T}$ is IO. Assume it is not. Then, $q \in \Delta(T)$ and a q-IC allocation U' exist such that $U'(t) > V(\delta_t)$ for every $t \in \text{supp}[q]$. Let $y = \max_{\tilde{t}\in\text{supp}[q]} U'(\tilde{t})$ so that $V(\delta_t) < y$ for all $t \in \text{supp}[q]$. By quasiconvexity of V, we get $V(\tilde{q}) < y$ for all $\tilde{q}$ with $\text{supp}[\tilde{q}] \subseteq \text{supp}[q]$. But because U' is q-IC, Lemma 1 implies $U'(t) < y$ for all $t \in \text{supp}[q]$, a contradiction. ∎

In Grossman (1981), and in the persuasion game of Milgrom (1981), the sender's (designer's) payoff is increasing in the mean of the distribution of the state of the world, so his value function V is quasiconvex in beliefs. An immediate, but important, implication of Proposition 3 and Theorem 2 is the following Corollary 1, which connects interim information design with evidence-disclosure games.

Corollary 1 (Interim optimality of full disclosure) *Assume that there is a single agent, that the utility of the designer is state independent, and that $V(q)$ is quasiconvex. Then an ex-post optimal mechanism is interim optimal. Consequently, full disclosure is a perfect Bayesian equilibrium of the informed-designer game.*

The key prediction of the large and influential literature on games of evidence disclosure, stemming from the seminal contributions of Grossman (1981), and of Milgrom (1981), is that full disclosure, the unraveling outcome, is the unique equilibrium outcome. Corollary 1 shows the unraveling outcome is not only a perfect Bayesian outcome when the designer can choose, and therefore deviate to, *any* stochastic evidence disclosure, but is also IO.

Proposition 3 implies that in settings in which V is quasiconvex, the interim optimality constraints (IOC) simplify to the following system of $|T|$ linear constraints:

$$\text{for every } t \in T,\ U(t) \geq u_0(a^*(\delta_t)). \tag{IOC-QC}$$

Hence, the IO solution is obtained by solving the following simplified version of Program (P):[27]

$$\max_{\sigma\in\Sigma(p)} E_\sigma[V(q)] \text{ subject to (IOC-QC).} \tag{P-QC}$$

[27] See Doval and Skreta (2018) for a solution approach.

We illustrate this program in the following two examples. The first example is a simplified version of the lobbyist example in Kamenica and Gentzkow (2011) with a discrete action space that generalizes the three-action version of Example 1 (ii). The EAO allocation is never IO. When the number of actions increases, the EAO allocation tends toward no disclosure, whereas every IO allocation tends toward full disclosure. Hence, the predictions of ex-ante information design dramatically differ from those of interim information design. The second example is the think-tank example in Lipnowski and Ravid (2020). In this example, the EAO mechanism coincides with the IO solution.

Example 3 (Lobbying) Consider the following simplified version of the lobbyist example in Kamenica and Gentzkow (2011). There are two states, $T = \{1, 0\}$, where $t = 1$ corresponds to the good state and $p(1) = p \in (0, 1)$ is the prior probability that the state is good.[28] The action space is

$$A = \left\{0, \frac{1}{K}, \frac{2}{K}, \ldots, \frac{K-1}{K}\right\}, \text{ with } K \geq 3.$$

The designer is a lobbyist and the agent is a politician. The lobbyist wants the politician to choose the highest possible action.[29] The higher the politician's belief that the state is good, the higher the action he chooses. The value function of the designer is

$$V(q) = f(k) \text{ for } q \in \left[\frac{k}{K}, \frac{k+1}{K}\right), k = 0, \ldots, K-1, \text{ and } V(1) = f(K-1),$$

where f is assumed to be strictly increasing and concave.

It is immediate that the EAO mechanism (the optimal splitting obtained by concavification) is as follows: for every k, if $p \in [\frac{k}{K}, \frac{k+1}{K})$, it splits the prior p to the posteriors $\frac{k}{K}$ and $\frac{k+1}{K}$. In line with the predictions of Kamenica and Gentzkow (2011), when K tends toward infinity, the EAO mechanism converges to no disclosure. By contrast, the IO mechanism obtained from the constrained concavification of Program (P-QC) splits any prior $p \leq \frac{K-1}{K}$ to the posteriors 0 and $\frac{K-1}{K}$. For every $K \geq 3$ and $p < \frac{K-1}{K}$, the IO solution is Blackwell more informative than the EAO solution. When K tends toward infinity, this mechanism and every IO mechanism converge to full disclosure. It follows that an informed lobbyist always reveals favorable information at the interim stage and information unravels (see left panel of Figure 4). Note that if we assume f is convex instead of concave, the EAO mechanism coincides with the IO solution of Program (P-QC) and converges to full disclosure (see right panel of Figure 4).

Example 4 (Think tank) Consider the think-tank example in Lipnowski and Ravid (2020). The state space is the same as in the previous example. The agent is the government and the designer is the think tank that wants to implement an agenda. The government can choose one of three actions: implement reform 1, implement reform 2, or keep the status quo. Reform 1 yields a payoff of 1 to the think tank and is optimal for the government when $q \leq \frac{1}{3}$, the status quo yields 0 for the think tank and is

[28] The example can be extended to any state space $T \subseteq \mathbb{R}$ if the value function of the designer only depends on the expected value of the state.

[29] The parameter α in Kamenica and Gentzkow (2011) is set to 0 here.

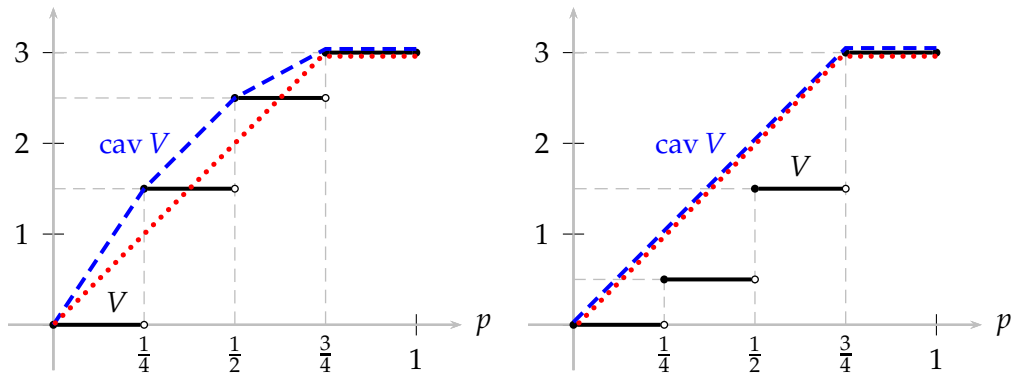

Figure 4: Lobbying: Designer ex-ante utility at the EAO mechanism (dashed lines) and the IO solution (dotted lines) in Example 3 with $K = 4$. Left panel: concave f. Right panel: convex f.

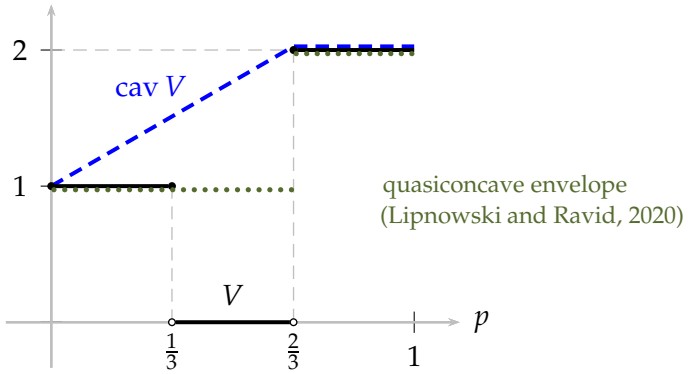

Figure 5: Think tank: Designer ex-ante utility at the EAO mechanism and the IO solution (dashed lines) and the quasiconcvave envelope (dotted lines)

optimal for the government when $q \in [\frac{1}{3}, \frac{2}{3}]$, and reform 2 yields a payoff of 2 to the think tank and is optimal for the government when $q \geq \frac{2}{3}$.

The resulting value function is quasiconvex and the IO solution of Program (P-QC) coincides with the EAO mechanism.[30] Splitting p to 0 and $\frac{2}{3}$ when $p < \frac{2}{3}$, and disclosing no information when $p \geq \frac{2}{3}$, is IO and EAO. Figure 5 illustrates these points and also shows that interim optimality differs from the quasiconcave envelope of V, which is the highest utility of the think tank under cheap talk (see Lipnowski and Ravid, 2020). This observation implies the ability to disclose verifiable information at the interim stage strictly benefits the think tank compared with cheap talk, even though the sender cannot commit ex ante to it.

Interestingly, the reverse observation applies to the lobbying example because, in that example, the unique equilibrium allocation under cheap talk is the non-revealing allocation. Hence, when f is concave, the cheap-talk solution is ex-ante better for the designer than every IO allocation.

[30] If we extend the example by allowing more than two possible reforms for the government, the EAO mechanism may no longer be IO. The IO solution will be similar to the solution with two reforms (it splits the prior to 0 and to the lowest posterior inducing the favorite reform), but the structure of the EAO mechanism will depend on the shape of V as in the previous example (Example 3).

6 Interim optimality in multi-agent settings

In this section, we consider multi-agent Bayesian incentive problems in which each agent has only two actions: $A_i = \{0,1\}$ for every $i \in I$. We first provide general sufficient conditions under which EAO mechanisms are IO. Then, we characterize IO mechanisms and compare them with EAO mechanisms in a class of parametrized environments similar to those studied by Bergemann and Morris (2019), Taneva (2019) and Mathevet et al. (2020).

6.1 Coordinating complementary investments: EAO mechanism is IO

We provide a condition, Assumption 1 below, under which every EAO mechanism is IO. Assumption 1 is always satisfied if there is a single agent, the designer's ideal action is state independent, and there are binary actions, as in the leading judge example in Kamenica and Gentzkow (2011), and in the setting of Perez-Richet (2014). Assumption 1 is also satisfied in many applications with multiple agents in the information-design literature: Alonso and Câmara (2016), Bardhi and Guo (2018), and Chan et al. (2019) consider voting settings, whereas Arieli and Babichenko (2019) consider a setting that encompasses technological adoption. Assumption 1 is also satisfied in the leading applications in Bergemann and Morris (2019) and Taneva (2019).

With some abuse of notation, let $(a_i, \mathbf{1}_{-i})$ denote the action profile where player i plays action a_i and all other players play action 1.

Assumption 1 A subset of types $T^* \subseteq T$ exists such that

(ia) For every $t \in T^*$ and $a \in A$, $u_0(1,\dots,1,t) \geq u_0(a,t)$;

(ib) For every $t \in T \backslash T^*$ and $a \in A$, $u_0(a,t) \geq u_0(0,\dots,0,t)$;

(iia) For every $i \in I$, $t \in T^*$, $u_i(1,\mathbf{1}_{-i},t) - u_i(0,\mathbf{1}_{-i},t) \geq 0$ and for every $a_{-i} \in A_{-i}$

$$u_i(1,\mathbf{1}_{-i},t) - u_i(0,\mathbf{1}_{-i},t) \geq u_i(1,a_{-i},t) - u_i(0,a_{-i},t);$$

(iib) For every $i \in I$, $t \in T \backslash T^*$, $u_i(0,a_{-i},t) > u_i(1,a_{-i},t)$ for every $a_{-i} \in A_{-i}$.

Condition (ia) means for every state in T^*, the best outcome for the designer is that every agent chooses action 1. Condition (ib) means that for every state outside T^*, the worst outcome for the designer is that every agent chooses action 0. In particular, these two assumptions are satisfied when the designer's utility is increasing in the number of agents choosing action 1, as in Arieli and Babichenko (2019). Condition (iia) means for every state in T^*, every agent has the highest incentive to choose action 1 when all the other agents also choose action 1. In particular, this assumption is satisfied when for every state in T^*, the complete-information game $(I,(A_i)_{i\in I},(u_i(\cdot,t))_{i\in I})$ has strategic complements and $a = (1,\dots,1)$ is a Nash equilibrium of that game. Finally, condition (iib) says action 0 is strictly dominant when the state is outside T^* and commonly known. This last part implies $a = (0,\dots,0)$ is the unique Nash equilibrium of the complete-information game $(I,(A_i)_{i\in I},(u_i(\cdot,t))_{i\in I})$ when $t \in T \backslash T^*$.

The set T^* is set of states in which, under complete information, the designer is able to get, at some Nash equilibrium, his first-best. The complement of T^* is the set of states in which the designer always gets his worst outcome under complete information. The next lemma shows that under Assumption 1, at an EAO allocation, the designer gets his first-best utility for every $t \in T^*$.

Lemma 2 *Consider a Bayesian incentive problem with binary actions satisfying Assumption 1.*[31] *If U^* is an ex-ante optimal allocation, then $U^*(t) = u_0(1, \ldots, 1, t)$ for every $t \in T^*$.*

Proof. Let μ be an EAO mechanism, and consider the mechanism μ^* such that $\mu^*(1, \ldots, 1 \mid t) = 1$ for every $t \in T^*$, and $\mu^*(a \mid t) = \mu(a \mid t)$ for every $a \in A$ and $t \in T \backslash T^*$. To prove the lemma, it suffices to show that μ^* is IC and it raises weakly higher utility than μ. From Condition (ia), for every $t \in T$, the designer is not worse off under μ^* than under μ. Hence, it remains to show μ^* is IC. Incentive compatibility for agent i is equivalent to

$$\begin{aligned} &\sum_{t \in T^*} p(t)[u_i(1, \mathbf{1}_{-i}, t) - u_i(0, \mathbf{1}_{-i}, t)] \\ &+ \sum_{t \in T \backslash T^*} p(t) \sum_{a_{-i}} \mu(1, a_{-i} \mid t)[u_i(1, a_{-i}, t) - u_i(0, a_{-i}, t)] \geq 0 \end{aligned}$$

and

$$\sum_{t \in T \backslash T^*} p(t) \sum_{a_{-i}} \mu(0, a_{-i} \mid t)[u_i(0, a_{-i}, t) - u_i(1, a_{-i}, t)] \geq 0.$$

The first inequality follows from Condition (iia) and the fact that μ is IC. The second inequality follows from Condition (iib). ∎

Proposition 4 *Consider a Bayesian incentive problem with binary actions satisfying Assumption 1. Then, an ex-ante optimal mechanism is interim optimal, and therefore a perfect Bayesian equilibrium of the informed-designer game.*

Proof. Let U^* be an EAO allocation. By Lemma 2, $U^*(t) = u_0(1, \ldots, 1, t)$ for every $t \in T^*$. Assume by way of contradiction that U^* is not IO. Then, a q-IC mechanism ν exists such that

$$U_0(\nu \mid t) > U^*(t) \text{ for every } t \in \text{supp}[q].$$

By Condition (ia), $\text{supp}[q] \subseteq T \backslash T^*$. Hence, by Condition (iib) and the fact that ν is q-IC we have $\nu(0, \ldots, 0 \mid t) = 1$ for every $t \in T \backslash T^*$. Finally, Condition (ib) implies $U_0(\nu \mid t) = u_0(0, \ldots, 0, t) \leq U^*(t)$ for every $t \in T \backslash T^*$, a contradiction. ∎

Assumption 1 provides sufficient conditions under which the EAO mechanism is IO. In settings that satisfy Assumption 1, the EAO mechanism is robust at the interim stage and the commitment assumption is without loss of generality. This positive result is strong given the prevalence of settings that satisfy Assumption 1.

[31] Condition (ib) is not required for this lemma.

6.2 Interim optimality in parametrized binary environments

We characterize IO mechanisms in a class of parametrized environments similar to the class of environments studied by Bergemann and Morris (2019), Taneva (2019), and Mathevet et al. (2020). We also illustrate the relevance of Assumption 1 within this class and illustrate how IO mechanisms differ from EAO ones when Assumption (iia) is not satisfied.

Consider two agents, two possible actions for each agent, $A_i = \{0, 1\}$, and two types for the designer, $T = \{0, 1\}$. For example, the agents are firms involved in a game of investment in a project and the designer is privately informed about the profitability of the project. Denote by p the prior probability of type $t = 1$. As in Taneva (2019), the utilities of the agents are given by the following tables, where $c, d > 0$:

$t=0$	$a_2=0$	$a_2=1$
$a_1=0$	c, c	$d, 0$
$a_1=1$	$0, d$	$0, 0$

$t=1$	$a_2=0$	$a_2=1$
$a_1=0$	$0, 0$	$0, d$
$a_1=1$	$d, 0$	c, c

Each agent would like to match the state. If $c > d$, they prefer to match the state jointly (strategic complements), and if $c < d$ they prefer to match the state alone (strategic substitutes). The designer would like both agents to choose action 1: his utility function is

$$u_0((1,1),0) = \overline{V}_0 > 0, \quad u_0((1,1),1) = \overline{V}_1 > 0,$$

and $u_0(a,t) = 0$ if $a \neq (1,1)$. This Bayesian incentive problem satisfies Assumption 1 iff $c \geq d$. The introductory example of Taneva (2019, Section 2), where the designer is a policy-maker who would like to convince two of her peers to vote for a motion, is a special case that obtains for $p = \frac{3}{10}$, $c = 2 \geq d = 1$, and $\overline{V}_0 = \overline{V}_1 = 1$, and hence, it satisfies Assumption 1. It then follows from Proposition 4 that the EAO mechanism characterized in Taneva (2019) is IO.[32]

The following observation is immediate from the definition of interim optimality:

Observation 1 *A mechanism μ is interim optimal iff μ is incentive compatible and $\mu((1,1) \mid t = 1) = 1$.*

We describe below the EAO mechanism and illustrate it may not be IO when Assumption 1 is not satisfied, that is, when $c < d$. Because agents are symmetric, we can focus on symmetric IC mechanisms, summarized by the following parameters $\mu = (\gamma_t, \beta_t, \alpha_t)_{t=0,1}$ with $0 \leq \gamma_t, \beta_t, \alpha_t \leq 1$ and $\gamma_t + 2\beta_t + \alpha_t = 1$:

$t=0$	$a_2=0$	$a_2=1$
$a_1=0$	γ_0	β_0
$a_1=1$	β_0	α_0

$t=1$	$a_2=0$	$a_2=1$
$a_1=0$	γ_1	β_1
$a_1=1$	β_1	α_1

The incentive constraints are:

$$\begin{aligned} (1-p)(\gamma_0 c + \beta_0 d) &\geq p(\gamma_1 d + \beta_1 c), \\ p(\beta_1 d + \alpha_1 c) &\geq (1-p)(\beta_0 c + \alpha_0 d). \end{aligned} \tag{3}$$

[32] Our parametrized class of games although similar, is not a special case of the parametrized class of games considered in Taneva (2019, Section 4), because in that section she assumes $u_0((0,0),0) = u_0((1,1),1)$ and $p = \frac{1}{2}$.

These constraints characterize the set of IC mechanisms μ that correspond to symmetric Bayes correlated equilibria. By definition, the EAO mechanism maximizes the ex-ante expected utility of the designer under these constraints. Hence, by Observation 1, the designer ex-ante preferred IO mechanism solves

$$\max_{\mu} p\alpha_1\overline{V}_1 + (1-p)\alpha_0\overline{V}_0 \text{ subject to (3) and } \alpha_1 = 1.$$

This program simplifies to $\max \alpha_0$ subject to $pc \geq (1-p)(\beta_0 c + \alpha_0 d)$, leading to the following observation:

Observation 2 *The designer ex-ante preferred interim-optimal mechanism is characterized by* $\alpha_0 = \min\{1, \frac{pc}{(1-p)d}\}$, $\alpha_1 = 1$, *and* $\beta_1 = \gamma_1 = \beta_0 = 0$.

From Proposition 4, this solution coincides with the EAO mechanism when $c \geq d$. For instance, in the introductory example of Taneva (2019) mentioned above, we get the ex-ante and IO mechanism $\alpha_0 = \frac{pc}{(1-p)d} = \frac{6}{7}$ and $\alpha_1 = 1$. However, when $c < d$, the EAO mechanism may not be IO. To illustrate, consider the following alternative numerical example with strategic substitutes: $c = 2$, $d = 7$, $\overline{V}_0 = 6$, $\overline{V}_1 = 1$ and $p = 0.3$. It can be checked that the EAO mechanism is

$t=0$	$a_2=0$	$a_2=1$
$a_1=0$	$\frac{11}{14}$	0
$a_1=1$	0	$\frac{3}{14}$

$t=1$	$a_2=0$	$a_2=1$
$a_1=0$	0	$\frac{1}{2}$
$a_1=1$	$\frac{1}{2}$	0

The resulting ex-ante expected utility of the designer is $\frac{9}{10}$, but the mechanism is not IO because $\alpha_1 = 0 \neq 1$. From Observation 2, the designer ex-ante preferred IO mechanism is

$t=0$	$a_2=0$	$a_2=1$
$a_1=0$	$\frac{43}{49}$	0
$a_1=1$	0	$\frac{6}{49}$

$t=1$	$a_2=0$	$a_2=1$
$a_1=0$	0	0
$a_1=1$	0	1

The resulting ex-ante expected utility of the designer is $\frac{57}{70}$, which is strictly lower than at the EAO mechanism.

In this example, the information designer wants agents' actions to be fully (and positively) correlated. Yet, despite this assumption, when the designer has ex-ante commitment power and agents' actions are strategic substitutes, the designer induces negative correlation and the probability that both invest is 0 in state $t = 1$. This negative correlation relaxes the obedience constraints in state $t = 0$, and thus arises for instrumental reasons (see Bergemann and Morris, 2019, for a related discussion). When the designer is informed, the ability to leverage this instrumental role of information is reduced because the designer of type $t = 1$ requires a utility of at least $\overline{V}_1$, which only arises when both agents invest.

7 Interim optimality and other solution concepts

In this section, we discuss the relationship of IO mechanisms with some key concepts of the informed-principal literature.

7.1 Core

We say the mechanism μ is *IC given* R, where $R \subseteq T$, iff it is q-IC for $q(\cdot) = p(\cdot \mid R)$, that is, for each agent i we have:

$$\sum_{a_{-i} \in A_{-i}} \sum_{t \in R} p(t)\mu(a \mid t)[u_i(a,t) - u_i((a_i', a_{-i}), t)] \geq 0, \text{ for every } a_i \text{ and } a_i' \text{ in } A_i. \quad (4)$$

Let

$$S(\nu, \mu) := \{t \in T : U_0(\nu \mid t) > U_0(\mu \mid t)\},$$

be the set of designer types who strictly prefer the mechanism ν over μ. A core mechanism is defined by Myerson (1983) as follows:

Definition 4 A mechanism $\mu : T \to \Delta(A)$ is a *core mechanism* iff μ is incentive compatible and no mechanism ν exists such that $S(\nu, \mu) \neq \emptyset$ and such that ν is incentive compatible given S for every $S \supseteq S(\nu, \mu)$.

To establish that IO allocations are core allocations, we rely on an alternative, simpler definition of core mechanisms in Lemma 3 below. To show this equivalence, we use the fact that an ex-post IC mechanism always exists, because in information-design settings, the mediator is omniscient; thus, no truth-telling conditions exist for the designer.

Lemma 3 *A mechanism $\mu : T \to \Delta(A)$ is a core mechanism iff μ is incentive compatible and no mechanism ν exists such that $S(\nu, \mu) \neq \emptyset$ and such that ν is incentive compatible given $S(\nu, \mu)$.*

Proof. The "if" part is immediate by definition. To establish the "only if" part, we show that if μ is IC and a mechanism ν exists such that $S(\nu, \mu) \neq \emptyset$ and such that ν is IC given $S(\nu, \mu)$, then μ is not a core mechanism; that is, a mechanism $\tilde{\nu}$ exists such that $S(\tilde{\nu}, \mu) \neq \emptyset$ and such that $\tilde{\nu}$ is IC given S for every $S \supseteq S(\tilde{\nu}, \mu)$. Consider the following mechanism:

$$\tilde{\nu}(t) = \begin{cases} \nu(t) & \text{if } t \in S(\nu, \mu) \\ \nu'(t) & \text{if } t \notin S(\nu, \mu), \end{cases}$$

where ν' is any ex-post IC mechanism. It is immediate to show $\tilde{\nu}$ is IC given S for every $S \supseteq S(\tilde{\nu}, \mu)$. ∎

Like an IO mechanism, a core mechanism has a natural interpretation in terms of deviations of coalitions of designer types. An IC mechanism μ is not a core mechanism iff a coalition of types $S \subseteq T$ and mechanism ν that is IC given S exist, such that all types in S strictly benefit from ν compared with μ. Note the belief of the agents after the deviation can either be interpreted as coming from a strategic inference that $t \in S$ or as a direct inference from a verifiable disclosure of the set S from the deviating coalition. An IO mechanism is similar to a core mechanism but allows for more blocking mechanisms. The definition of IO mechanism does not require the blocking mechanism ν to be IC given $S(\nu, \mu)$; the blocking mechanism could be IC for *some* belief q whose support is included in $S(\nu, \mu)$ (i.e., $\text{supp}[q] \subseteq S(\nu, \mu)$). This definition allows for more flexibility: agents can modify arbitrarily the relative likelihoods of the different types

in $S(\nu, \mu)$, whereas in the definition of the core mechanism, beliefs are "passive" because they keep the relative likelihoods of the different types in $S(\nu, \mu)$ constant. In other words, interim optimality entails a larger set of blocking mechanisms that is the driving force of the following result:

Proposition 5 *If μ is an interim-optimal mechanism, then μ is a core mechanism.*

Proof. Follows directly from the alternative definition of core in Lemma 3 and the definition of IO mechanisms (Definition 3). ∎

The reverse of this proposition is not true. In Example 1 (iii), for the prior $p = \frac{1}{6}$, every IC allocation is a core allocation. In Example 2, the core allocation $(1, 0)$ (which is EAO for the assumed prior) is not IO. This last example also shows that a core allocation is not necessarily an equilibrium allocation because, as seen previously, $(1, 0)$ is not an equilibrium allocation.

We finish this section by noting the set of core allocations changes when some of the designer's types are decomposed into multiple payoff-equivalent types. Consider the four-action version of Example 1 (iii) with the prior $p = \frac{1}{3}$. The EAO allocation $U = (2, 2)$ is a core allocation: it is not blocked by $T = \{1, 0\}$, because $U = (2, 2)$ is EAO, so it is not strictly dominated by an allocation that is IC for the prior. It is also not blocked by a single type t, because this type would get 0 (at belief $q = 0$ or $q = 1$). Now consider the Bayesian incentive problem $T' = \{1, t_0, t_0'\}$ with $p' = (\frac{1}{3}, \frac{1}{6}, \frac{1}{2})$ and assume all payoffs at $t = t_0$ and $t = t_0'$ are the same as the payoffs at $t = 0$. Types t_0 and t_0' are two types that are payoff-equivalent to $t = 0$, and the total probability of t_0 and t_0' is the same as the probability of $t = 0$. The EAO allocation is $(2, 2, 2)$, which is equivalent to the previous one. But now this allocation is no longer a core allocation: it is blocked by the coalition $S = \{1, t_0\}$ and allocation $U' = (3, 3, 0)$ because now the belief is $p'(t = 1 \mid S) = \frac{2}{3}$, and therefore, the allocation $U' = (3, 3, 0)$ is IC given S, and both types $t = 1$ and $t = t_0$ are strictly better off compared with the EAO allocation. The set of IO mechanisms is immune to such payoff-equivalent decompositions of types, which could be interpreted as private randomization devices or sunspots.

7.2 SUPO and SNP mechanisms

Maskin and Tirole (1990) introduced the notion of a strong unconstrained Pareto optimal mechanism, which exists and is an equilibrium of some informed-principal problems with private values and transfers.

Definition 5 (Maskin and Tirole, 1990) A mechanism $\mu : T \to \Delta(A)$ is *strong unconstrained Pareto optimal (SUPO)* iff it is incentive compatible and no belief $q \in \Delta(T)$ together with a q-incentive-compatible mechanism ν exist such that $U_0(\nu \mid t) \geq U_0(\mu \mid t)$ for every $t \in T$, with a strict inequality for some $t \in T$, and a strict inequality for all $t \in T$ if $\text{supp}[q] \neq T$.

As already observed by Mylovanov and Tröger (2012), SUPO mechanisms usually fail to exist if there are no transfers. For instance, Example 1 (i) has no SUPO mechanism for $p < \frac{1}{3}$, and Example 1 (ii) has is no SUPO mechanism for $p < \frac{2}{3}$.

Mylovanov and Tröger (2012) introduced a similar concept, called a strong neologism-proof mechanism, which exists in more general *private* value adverse-selection environments and is also a perfect Bayesian equilibrium mechanism of the informed-principal game in such environments. Let

$$U_0^{FB}(t) = \max\{u_0(a,t) : a \in A\},$$

be the first-best utility for type t of the designer, that is, the highest possible utility of the designer when his type is t.

Definition 6 (Mylovanov and Tröger, 2012) A mechanism $\mu : T \to \Delta(A)$ is *strong neologism-proof (SNP)* iff it is incentive compatible and there is no belief $q \in \Delta(T)$ such that $q(t) = 0$ if $U_0(\mu \mid t) = U_0^{FB}(t)$, together with a q-incentive-compatible mechanism ν such that $U_0(\nu \mid t) \geq U_0(\mu \mid t)$ for every $t \in \text{supp}[q]$, with a strict inequality for some $t \in \text{supp}[q]$.

In the next example, even SNP mechanisms do not exist. Failure of existence is related to the fact that the set of blocked allocations in the definitions of SUPO and SNP is not necessarily an open set. By contrast, the set of blocked allocations in the definition of an IO allocation is an open set.[33]

Example 5 (SUPO and SNP allocations may not exist) Consider the following example with a single agent, $T = \{t_1, t_2\}$ and $A = A_1 = \{a^1, a^2, a^3\}$:

	a^1	a^2	a^3
t_1	$0,0$	$1,1$	$2,-1$
t_2	$0,1$	$1,0$	$0,1$

The first-best allocation is $U_0^{FB} = (2,1)$. If $p(t_1) = p < \frac{1}{2}$ every IC allocation is dominated by the allocation $(1,1)$, which is q-IC for $q(t_1) \geq \frac{1}{2}$, so no SUPO or SNP allocation exist. However, it is immediate that the set of IO allocations is the set of IC allocations in which the utility of type t_1 is equal to 1. In particular, the EPO allocation $(1,0)$ is IO whatever the prior.

Proposition 6 *If μ is a strong neologism-proof mechanism, then μ is an interim-optimal mechanism and therefore a perfect Bayesian equilibrium of the informed-designer game.*

Proof. Let μ be an IC mechanism that is not an IO mechanism; that is, $q \in \Delta(T)$ and a q-IC mechanism ν exist such that $\text{supp}[q] \subseteq S(\nu, \mu)$. By definition, for every $t \in S(\nu, \mu)$, we have $U_0(\mu \mid t) < U_0(\nu \mid t) \leq U_0^{FB}(t)$. Because $\text{supp}[q] \subseteq S(\nu, \mu)$, we get $q(t) = 0$ if $U_0(\mu \mid t) = U_0^{FB}(t)$ and $U_0(\mu \mid t) < U_0(\nu \mid t)$ for every $t \in \text{supp}[q]$. Hence, μ is not an SNP mechanism. We conclude by Theorem 2. ∎

To summarize, we have the following relationships in our Bayesian incentive environment: neutral optima and SNP mechanisms are IO, IO mechanisms are perfect Bayesian equilibrium mechanisms, and IO mechanisms are core mechanisms. We have also observed that the set of SUPO and SNP mechanisms may be empty, that some core mechanisms may not be equilibrium mechanisms, and that some perfect

[33] See Appendix A.1.

Bayesian equilibrium mechanisms are not IO. Whether IO mechanisms are neutral optima in general or under specific assumptions is an open and difficult question that is left for future research.

8 Imperfectly informed designer and enforceable actions

We extend the model of Section 2 by allowing the designer to be partially informed about the payoff-relevant state of nature and by adding a set of enforceable actions (such as fines and bonuses) for the designer. The designer has a non-empty and finite set of enforceable actions A_0 and we denote by $A = \prod_{i=0}^{n} A_i$ the set of action profiles. The designer is privately informed about his type $t \in T$. The payoff-relevant state is now $(t, \omega) \in T \times \Omega$, where Ω is non-empty and finite. No player observes $\omega \in \Omega$. The marginal probability distribution of T is $p \in \Delta(T)$ and has full support. The conditional probability distribution of Ω is given by $\pi : T \to \Delta(\Omega)$, and $\pi(\omega \mid t)$ denotes the probability of ω given t. The utility of each player $i = 0, 1, \dots n$ is $u_i(a, t, \omega)$. A Bayesian incentive problem is now given by $\Gamma = ((A_i)_{i=0}^{n}, (u_i)_{i=0}^{n}, T, \Omega, p, \pi)$. We get a Bayesian incentive problem as defined in the main text as a particular case in which $|A_0| = |\Omega| = 1$.

A direct mechanism is a mapping $\mu : T \times \Omega \to \Delta(A)$, where $\mu(a_0, a_1, \dots, a_n \mid t, \omega)$ is the probability that the mediator implements the enforceable action a_0 and privately recommends a_i to each agent i when the state is (t, ω). The notion of incentive compatibility directly extends to this more general setting. Let $q \in \Delta(T)$ denote the agents' beliefs about the designer's type. The mechanism μ is q-IC iff for every i in I, and a_i and a_i' in A_i,

$$\sum_{a_{-i} \in A_{-i}} \sum_{t \in T} \sum_{\omega \in \Omega} q(t)\pi(\omega \mid t)\mu(a \mid t, \omega) \left[u_i(a, t, \omega) - u_i((a_i', a_{-i}), t, \omega) \right] \geq 0,$$

and it is IC if it is p-IC. [34]

The interim expected utility of the designer's type t from mechanism μ is $U_0(\mu \mid t) = \sum_{a \in A} \sum_{\omega \in \Omega} \pi(\omega \mid t)\mu(a \mid t, \omega)u_0(a, t, \omega)$. The definitions of ex-post, interim, and ex-ante optimality are exactly the same as their counterparts in Section 2 and Section 3. When the designer has no private information ($|T| = 1$), the interim-design problem is equivalent to the standard ex-ante design problem, and a mechanism is IO iff it is EAO or EPO. We prove Theorem 1 and Theorem 2 in the appendix for this more general setting.

Example 6 (Imperfectly informed designer) To illustrate how the precision of information of the designer affects IO mechanisms beyond the extreme cases in which the designer is uninformed or perfectly informed about the state, consider Example 1 (ii) with three actions $A = \{\underline{a}^0, a^2, a^3\}$. Let $\Omega = \{1, 0\}$ and assume the utility function of the agent only depends on $a \in A$ and $\omega \in \Omega$, where the payoff-relevant state is now ω instead of t. That is, if $\bar{\pi}$ is the belief of the agent about $\omega = 1$, his optimal action is $\underline{a}^0$ if $\bar{\pi} < \frac{1}{3}$, a^2 if $\bar{\pi} \in [\frac{1}{3}, \frac{2}{3})$ and a^3 if $\bar{\pi} \geq \frac{2}{3}$. The marginal probability of $\omega = 1$ is $\frac{1}{6}$.

[34] Note Chen and Zhang (2020) and Hedlund (2017) study signaling settings in which the designer chooses $\mu : \Omega \to \Delta(X)$ instead of a mechanism $\mu : T \times \Omega \to \Delta(X)$. Another difference is that in those papers, the designer's type t is unverifiable, whereas it is verifiable in our setting. See Remark A.1.

The designer's utility only depends on the agent's action. The type of the principal is now a signal $t \in \{0, \bar{t}\}$ about the payoff-relevant state, with

$$\pi(\omega = 1 \mid t = 0) = 0 \text{ and } \pi(\omega = 1 \mid t = \bar{t}) = \bar{t} \in [\frac{1}{6}, 1].$$

Hence, the prior marginal probability of type $t = \bar{t}$ is $p = \frac{1}{6\bar{t}}$, and type t of the principal simply corresponds to his belief about state $\omega = 1$. When the agent has belief q about the designer's type $t = \bar{t}$, his belief about $\omega = 1$ is $q\bar{t}$ and his prior belief about $\omega = 1$ is $p\bar{t} = \frac{1}{6}$.

Because utilities do not directly depend on t, the ex-ante expected utility of the designer at the EAO mechanism does not depend on the precision of the designer's information, $\bar{t}$, and is the same as in the original example. However, the interim utilities of the designer at an EAO mechanism depend on $\bar{t}$. They also depend on the EAO mechanism that is used, except when $\bar{t} = 1$, in which case the EAO mechanism is unique. Every EAO mechanism $\mu : T \times \Omega \to \Delta(A)$ satisfies the following:

$$\Pr(a = a^2 \mid \omega = 1) = \mu(a^2 \mid \bar{t}, 1) = 1,$$

and

$$\begin{aligned}\Pr(a = a^2 \mid \omega = 0) &= \Pr(t = \bar{t} \mid \omega = 0)\mu(a^2 \mid \bar{t}, 0) + \Pr(t = 0 \mid \omega = 0)\mu(a^2 \mid 0, 0), \\ &= \frac{1 - \bar{t}}{5\bar{t}}\mu(a^2 \mid \bar{t}, 0) + \frac{6\bar{t} - 1}{5\bar{t}}\mu(a^2 \mid 0, 0) = \frac{2}{5}.\end{aligned}$$

Such a mechanism is IO iff the high-type designer gets at least his EPO utility allocation $U_0^{EPO}(\bar{t}) = \operatorname{cav} V(\bar{t})$. Hence, to characterize when an EAO mechanism is IO it suffices to focus on the EAO mechanism μ that maximizes the utility of type $\bar{t}$. Such an EAO mechanism is

$$\mu(a^2 \mid \bar{t}, 1) = 1 \text{ and } \mu(a^3 \mid t, \omega) = 0 \text{ for every } t \text{ and } \omega,$$

$$\mu(a^2 \mid \bar{t}, 0) = \begin{cases} 1 & \text{if } \frac{1-\bar{t}}{5\bar{t}} \leq \frac{2}{5}, \text{ i.e., } \bar{t} \geq \frac{1}{3}, \\ \frac{2\bar{t}}{1-\bar{t}} & \text{if } \bar{t} \leq \frac{1}{3}, \end{cases}$$

$$\mu(a^2 \mid 0, 0) = \begin{cases} \frac{3\bar{t}-1}{6\bar{t}-1} & \text{if } \bar{t} \geq \frac{1}{3}, \\ 0 & \text{if } \bar{t} \leq \frac{1}{3}. \end{cases}$$

If $\bar{t} > \frac{1}{3}$, we get $U_0(\mu \mid \bar{t}) = 2 < U_0^{EPO}(\bar{t}) = \operatorname{cav} V(\bar{t})$, so no EAO mechanism is IO. If $\bar{t} \leq \frac{1}{3}$, we get $U_0(\mu \mid \bar{t}) = 6\bar{t} = U_0^{EPO}(\bar{t}) = \operatorname{cav} V(\bar{t})$, so the EAO mechanism μ is IO. To conclude, in this example, we have shown that if the precision of the designer's information is low ($\bar{t} \in [\frac{1}{6}, \frac{1}{3}]$), an EAO mechanism exists that is IO. Otherwise, if the precision of the designer's information is high ($\bar{t} > \frac{1}{3}$), no EAO mechanism exists that is IO.

9 Concluding remarks

In this paper, we identified a class of disclosure mechanisms, which we coined *interim-optimal mechanisms*. These mechanisms are optimal in the sense that the informed designer cannot credibly find an alternative mechanism that strictly improves his interim payoff. We established that the notion of interim optimality is well founded because an interim-optimal mechanism always exists, and every interim-optimal mechanism is implementable as a perfect Bayesian equilibrium of the informed-designer game. Interim-optimal mechanisms can be tractably characterized in common settings using Kamenica and Gentzkow's (2011) belief-based approach and other state-of-the-art tools.

Interim optimality provides a benchmark to evaluate whether an ex-ante optimal mechanism–the commitment solution in Kamenica and Gentzkow (2011)–is robust in the sense that the designer will select it in settings in which he has private information when choosing the disclosure mechanism. At the same time, when an ex-ante optimal mechanism turns out not to be interim optimal, we provide tools to identify which disclosure mechanisms an informed designer with all the bargaining power is willing to use. Given the generality of our setting, these results open the door to an array of problems in information design in which the assumption of ex-ante commitment to a mechanism is not compelling and it is more natural to assume the designer possesses some private information when selecting the informativeness of a procedure, as in the settings we mentioned in the introduction.

The situations we explored also shed new light on the information-unraveling prediction, which is focal in the large and influential literature on disclosure games stemming from the classical works of Milgrom (1981) and Grossman (1981). We showed that the unraveling outcome is interim optimal and thus a robust prediction even when the informed designer can choose any disclosure mechanism. By contrast, in the settings of Milgrom (1981) and Grossman (1981), if the designer's value function is concave in beliefs, the ex-ante optimal mechanism is no disclosure. One may then wonder whether an interim-optimal mechanism is always more informative than an ex-ante optimal one. The answer is no. Recall Example 2 in which the interim-optimal mechanism is to reveal no information at all. There, at the interim, the designer wants to keep the agent in the dark, whereas the ex-ante optimal mechanism reveals information.

Whereas our setting is general, our results apply even more broadly: straightforward extensions include allowing for private information on the side of the agents and for non-contractible actions for the designer. Other interesting, but not immediate, extensions include the relaxation of the verifiability assumption on the part of the designer as well as the assumption that the informed designer cannot tamper with the mechanism's input, as in Richet-Perez and Skreta (2022), or the mechanism's output, as in Lipnowski et al. (2022). The latter scenarios could be particularly interesting for applied settings.

A Appendix

In this appendix, we consider the Bayesian incentive problem of Section 8:

$$\Gamma = ((A_i, u_i)_{i=0}^n, T, \Omega, p, \pi).$$

This generalized model allows the designer to be imperfectly informed about the state (the state is $(t, \omega) \in T \times \Omega$ but the designer is only informed about t) and to choose enforceable actions in any non-empty and finite set A_0. A direct mechanism is a mapping $\mu : T \times \Omega \to \Delta(A)$, and the interim expected utility of the designer's type t from mechanism μ is $U_0(\mu \mid t)$. The model of the main text is obtained as a special case of this generalized model by assuming $|\Omega| = |A_0| = 1$. We prove Theorem 1 and Theorem 2 directly for this general setting.

Remark A.1 As in the baseline setting of Section 2, we follow the information-design literature and take the inputs in the disclosure mechanism to be the realized state (t, ω) (so the mediator is omniscient, and no truth-telling constraints exist for the designer). An alternative setting would be one in which the mechanism depends on ω and the *reported* type of the designer.

To prove Theorem 1, we rely on the notion of strong solution defined in Myerson (1983). Strong solution is used in one of the axioms of Myerson (1983) in Appendix A.1 and to connect EAO, EPO and IO mechanisms in Proposition 1. The definition of a strong solution relies on the concept of undominated mechanisms, which we define next.

Definition A.7 A mechanism μ is *dominated by* ν iff $U_0(\mu \mid t) \leq U_0(\nu \mid t)$ for every $t \in T$, with a strict inequality for at least one t. A mechanism μ is *strictly dominated* by ν iff $U_0(\mu \mid t) < U_0(\nu \mid t)$ for every $t \in T$. A mechanism μ is *undominated* iff μ is incentive compatible and μ is not dominated by any other incentive-compatible mechanism.

An EAO mechanism is undominated. An EPO mechanism, however, may be dominated; if it is not, it is a strong solution.

Definition A.8 A mechanism μ is a *strong solution* iff it is ex-post incentive compatible and undominated.

A.1 Proof of Theorem 1

We first present the axiomatic definition of neutral optimum in Myerson (1983), adapted to our setting with verifiable information. Let $U_0(\mu) := (U_0(\mu \mid t))_{t \in T} \in \mathbb{R}^T$ be the utility allocation vector of the designer from mechanism μ. Given a Bayesian incentive problem Γ, $B(\Gamma) \subseteq \mathbb{R}^T$ is a set of *blocked allocations*. As mentioned right after Theorem 1 in the main text, we let $B^{IO}(\Gamma)$ be the set of allocations $U \in \mathbb{R}^T$ such that a belief $q \in \Delta(T)$ and a q-IC allocation U' exist such that $U'(t) > U(t)$ for every $t \in \text{supp}[q]$. By definition, an allocation U is an IO allocation iff it is IC and $U \notin B^{IO}(\Gamma)$.

The first axiom requires that if an allocation U is blocked and U' is strictly dominated by U, U' is blocked as well:

Axiom 1 (Domination) *For every $U, U' \in \mathbb{R}^T$, if $U \in B(\Gamma)$ and $U'(t) < U(t)$ for every t, then $U' \in B(\Gamma)$.*

The next axiom requires that if U is blocked, a neighborhood of U exists such that every allocation in that neighborhood is blocked too.

Axiom 2 (Openness) *$B(\Gamma)$ is an open set of $\mathbb{R}^T$.*

A Bayesian incentive problem $\overline{\Gamma} = ((\overline{A}_0, (A_i)_{i=1}^n), (\overline{u}_0, (\overline{u}_i)_{i=1}^n), T, \Omega, p, \pi)$ is an *extension* of the Bayesian incentive problem $\Gamma = ((A_0, (A_i)_{i=1}^n), (u_0, (u_i)_{i=1}^n), T, \Omega, p, \pi)$ if $A_0 \subseteq \overline{A}_0$ and

$$\overline{u}_i(a, t, \omega) = u_i(a, t, \omega), \text{ for every } i = 0, 1, \ldots, n, t \in T, \omega \in \Omega \text{ and } a \in A.$$

That is, an extension $\overline{\Gamma}$ of Γ is a Bayesian incentive problem in which, compared with Γ, the designer can commit to additional enforceable actions. The idea of the next axiom is that in $\overline{\Gamma}$, more allocations could therefore be blocked.

Axiom 3 (Extensions) *If $\overline{\Gamma}$ is an extension of Γ, then $B(\Gamma) \subseteq B(\overline{\Gamma})$.*

The last axiom requires that a strong solution should never be blocked.

Axiom 4 (Strong solutions) *If μ is a strong solution of Γ, then $U_0(\mu) \notin B(\Gamma)$.*

Let $\mathbf{H}$ be the set of all functions $B(\cdot)$ satisfying the four axioms, and for every Γ, let

$$B^*(\Gamma) = \bigcup_{B \in \mathbf{H}} B(\Gamma).$$

Note that B^* satisfies the four axioms. The set of neutral optima is the smallest possible set of unblocked IC mechanisms:

Definition A.9 A mechanism μ is a *neutral optimum* iff μ is incentive compatible and $U_0(\mu) \notin B^*(\Gamma)$.

Lemma A.4 (Myerson, 1983) *For any Bayesian incentive problem Γ, at least one neutral optimum exists.*

The proof Lemma A.4 is the same as the proof of Theorem 6 in Myerson (1983). The necessary and sufficient conditions that characterize the neutral optima in Theorem 7 in Myerson (1983) are simpler in our setting because incentive compatibility conditions are simply the agents' obedience constraints, whereas in Myerson (1983) truth-telling constraints are also in place. Formally, in Theorem 7 in Myerson (1983), the shadow price for the constraint that type t of the designer should not be tempted to claim to be type s is always zero.

The next step is to show $B^{IO}(\Gamma)$ satisfies the axioms of *Domination*, *Openness*, *Extensions*, and *Strong solutions*.

Domination. Let $U \in B^{IO}(\Gamma)$; that is, $q \in \Delta(T)$ and a q-IC allocation $U' \in \boldsymbol{U}(q)$ exist such that $U'(t) > U(t)$ for every $t \in \text{supp}[q]$. If $\tilde{U}(t) < U(t)$ for every $t \in T$, then

$U'(t) > U(t) > \tilde{U}(t)$ for every $t \in \text{supp}[q]$. Hence, $\tilde{U}$ is blocked by U'; that is, $\tilde{U} \in B^{IO}(\Gamma)$.

Openness. For every $t \in T$, let $\varepsilon(t) \in \mathbb{R}$, $\varepsilon(t) \neq 0$, and $\tilde{U}(t) = U(t) + \varepsilon(t)$. For every $t \in \text{supp}[q]$, we have $U'(t) > U(t)$, so for $\varepsilon(t)$ close enough to zero, we get $U'(t) > \tilde{U}(t)$. Hence, $\tilde{U}$ is blocked by U'; that is, $\tilde{U} \in B^{IO}(\Gamma)$.

Extensions. If U' is q-IC in Γ, it is also q-IC in an extension $\overline{\Gamma}$ of Γ. Hence, if U is blocked by U' in Γ, it is also blocked by U' in $\overline{\Gamma}$. Therefore, $B^{IO}(\Gamma) \subseteq B^{IO}(\overline{\Gamma})$.

Strong solutions. To show $B^{IO}(\Gamma)$ satisfies the strong solution axiom (which we state as a proposition below, see Proposition A.7), we start with an auxiliary lemma.

Lemma A.5 *If ν is q-IC and ν' is q'-IC, then for every $\alpha \in [0,1]$, the mechanism ν^*, defined by*

$$\nu^*(a \mid t, \omega) = \frac{\alpha q(t)}{q^*(t)} \nu(a \mid t, \omega) + \frac{(1-\alpha) q'(t)}{q^*(t)} \nu'(a \mid t, \omega),$$

for every $a \in A$, $t \in \text{supp}[q^*]$ and $\omega \in \Omega$, with $q^*(t) = \alpha q(t) + (1-\alpha) q'(t)$ for every $t \in T$, is q^*-IC.

Proof. The mechanism ν^* is q^*-IC iff for every a_i and a_i' in A_i

$$\sum_{a_{-i} \in A_{-i}} \sum_{t \in T} \sum_{\omega \in \Omega} q^*(t) \pi(\omega \mid t) \nu^*(a \mid t, \omega) [u_i(a, t, \omega) - u_i((a_i', a_{-i}), t, \omega)] \geq 0,$$

that is,

$$\sum_{a_{-i} \in A_{-i}} \sum_{t \in T} \sum_{\omega \in \Omega} \left(\alpha q(t) \pi(\omega \mid t) \nu(a \mid t, \omega) + (1-\alpha) q'(t) \pi(\omega \mid t) \nu'(a \mid t, \omega) \right) \times [u_i(a, t, \omega) - u_i((a_i', a_{-i}), t, \omega)] \geq 0,$$

or, equivalently,

$$\begin{aligned} & \alpha \sum_{a_{-i} \in A_{-i}} \sum_{t \in T} \sum_{\omega \in \Omega} q(t) \pi(\omega \mid t) \nu(a \mid t, \omega) [u_i(a, t, \omega) - u_i((a_i', a_{-i}), t, \omega)] \\ + \quad & (1-\alpha) \sum_{a_{-i} \in A_{-i}} \sum_{t \in T} \sum_{\omega \in \Omega} q'(t) \pi(\omega \mid t) \nu'(a \mid t, \omega) [u_i(a, t, \omega) - u_i((a_i', a_{-i}), t, \omega)] \geq 0. \end{aligned}$$

The first term is positive for every a_i and a_i' in A_i because ν is q-IC and the second term is positive for every a_i and a_i' in A_i because ν' is q'-IC. Hence, ν^* is q^*-IC. ∎

The intuition for Lemma A.5 is as follows. If ν is q-IC and ν' is q'-IC, and $q^* = \alpha q + (1-\alpha) q'$, then when the prior belief is q^*, the designer can first use an information-disclosure policy that splits the prior belief q^* to the posterior belief q with probability α and to the posterior belief q' with probability $1-\alpha$. By Bayes' rule, the probability that the posterior is q conditional on t is $\frac{\alpha q(t)}{q^*(t)}$, and the probability that the posterior is q' conditional on t is $\frac{(1-\alpha) q'(t)}{q^*(t)}$. Then, the designer uses the q-IC mechanism ν when the posterior is q, and the q'-IC mechanism ν' when the posterior is q'.

Proposition A.7 *If μ is a strong solution, then μ is interim optimal.*

Proof. Assume by way of contradiction that μ is a strong solution but not IO. Then, $q \in \Delta(T)$ and a q-IC mechanism ν exist such that $U_0(\nu \mid t) > U_0(\mu \mid t)$ for every $t \in \text{supp}[q]$.

For every $t \in T$, let

$$q'(t) = \frac{p(t) - \alpha q(t)}{1 - \alpha},$$

where $\alpha \in (0,1)$ is small enough that $p(t) > \alpha q(t)$; that is, $q'(t) > 0$ for all $t \in T$. This is possible because p is assumed to have full support. Note $\sum_{t \in T} q'(t) = \sum_{t \in T} \frac{p(t) - \alpha q(t)}{1-\alpha} = 1$, so $q' \in \Delta(T)$ is a full-support belief: $\text{supp}[q'] = T$.

Define the following mechanism ν^*:

$$\nu^*(a \mid t, \omega) := \frac{\alpha q(t)}{p(t)} \nu(a \mid t, \omega) + \frac{(1-\alpha) q'(t)}{p(t)} \mu(a \mid t, \omega),$$

for every $t \in T$, $\omega \in \Omega$, and $a \in A$. Note $p(t) = \alpha q(t) + (1-\alpha) q'(t)$ for every $t \in T$, ν is q-IC and μ is q'-IC because μ is ex-post IC. Hence, from Lemma A.5, the mechanism ν^* is IC for the prior p. In addition, for every $t \in T$, we have by construction

$$U_0(\nu^* \mid t) = \frac{\alpha q(t)}{p(t)} U_0(\nu \mid t) + \frac{(1-\alpha) q'(t)}{p(t)} U_0(\mu \mid t).$$

We get $U_0(\nu^* \mid t) \geq U_0(\mu \mid t)$ for every $t \in T$, with a strict inequality for every $t \in \text{supp}[q]$. It follows that ν^* is IC and dominates μ, a contradiction to the assumption that μ is a strong solution. ∎

We conclude $B^{IO}(\Gamma) \subseteq B^*(\Gamma)$, and therefore, a neutral optimum is IO. Hence, an IO allocation exists because a neutral optimum exists (Lemma A.4). This completes the proof of Theorem 1.

A.2 Proof of Theorem 2

For the generalized model of Section 8, the informed-designer game is the same as in Section 4 except that in the first stage, nature selects the state of the world (t, ω) with probability $p(t)\pi(\omega \mid t)$; in the third stage, the designer chooses a mechanism $\nu : T \times \Omega \to \Delta(X)$, where $X = A_0 \times X_1 \times \cdots \times X_n$; in the fifth stage, $\nu(a_0, x_1, \ldots, x_n \mid t, \omega)$ is the probability of implementing the enforceable action a_0 and sending message x_i privately to each agent i when the state is (t, ω).

To prove Theorem 2, we show that an IO mechanism is an expectational equilibrium as defined in Myerson (1983), which is a refinement of perfect Bayesian equilibrium in the spirit of sequential equilibrium (Kreps and Wilson, 1982). We first adapt some auxiliary definitions and results developed in Myerson (1983).

Revelation and inscrutability principles Following Myerson (1983), we can rely on the revelation and the inscrutability principles, which allow us to conclude that for every equilibrium in which the designer uses a generalized mechanism $\nu_t : T \times \Omega \to \Delta(X)$ when his type is $t \in T$, an outcome-equivalent equilibrium exists in which all designer types offer the same direct mechanism $\mu : T \times \Omega \to \Delta(A)$ (so agents' beliefs

at the beginning of Stage 6 are the same as the prior) and agents are obedient *along* the equilibrium path.

Continuation equilibrium In an expectational equilibrium, for every off-path mechanism ν and belief q, agents are required to be sequentially rational in Stage 6. Each agent i chooses a function $\gamma_i : X_i \to \Delta(A_i)$ that determines the probability that he chooses action $a_i \in A_i$ as a function of his signal $x_i \in X_i$. Sequential rationality for the agents requires the strategy profile $(\gamma_i)_{i\in I}$ to constitute a (continuation) Nash equilibrium given q and ν. For every $x \in X$ and $a \in A$, let

$$\gamma(a \mid x) = \begin{cases} \prod_{i\in I} \gamma_i(a_i \mid x_i) & \text{if } x_0 = a_0 \\ 0 & \text{otherwise,} \end{cases}$$

be the probability that the action profile a is played when agents play the strategy profile $(\gamma_i)_{i\in I}$ and the outcome of the mechanism is x.

Let $W_0(\nu, (\gamma_i)_{i\in I} \mid t)$ be the interim expected utility of the designer given t, the mechanism ν, and the agents' strategy profile $(\gamma_i)_{i\in I}$:

$$W_0(\nu, (\gamma_i)_{i\in I} \mid t) = \sum_{\omega\in\Omega} \sum_{x\in X} \sum_{a\in A} \pi(\omega \mid t)\nu(x \mid t, \omega)\gamma(a \mid x)u_0(a, t, \omega).$$

Let $W_i(\nu, (\gamma_i)_{i\in I} \mid q)$ be the expected utility of agent i given belief $q \in \Delta(T)$, the mechanism ν, and the strategy profile $(\gamma_i)_{i\in I}$ of the agents:

$$W_i(\nu, (\gamma_i)_{i\in I} \mid q) = \sum_{\omega\in\Omega} \sum_{t\in T} \sum_{x\in X} \sum_{a\in A} q(t)\pi(\omega \mid t)\nu(x \mid t, \omega)\gamma(a \mid x)u_i(a, t, \omega).$$

Definition A.10 $(\gamma_i)_{i\in I}$ is a *continuation Nash equilibrium for* $\nu : T \times \Omega \to \Delta(X)$ *given* q iff for every $i \in I$ and $\gamma_i' : X_i \to \Delta(A_i)$, we have

$$W_i(\nu, (\gamma_i)_{i\in I} \mid q) \geq W_i(\nu, (\gamma_i', \gamma_{-i}) \mid q).$$

Because agents have symmetric information at the beginning of Stage 6 of the extensive form defined in Section 4, we require that they have a common belief q at this stage, even off the equilibrium path. This reason is also why we formulated the notion of incentive compatibility for a common belief.

Note the game induced by ν with prior q has finite sets of pure strategies, so a continuation Nash equilibrium for ν given q always exists. The non-empty and compact set of continuation equilibrium allocations for ν given q is denoted by $\mathcal{U}(\nu, q)$: it is the set of all $U \in \mathbb{R}^T$ such that a continuation Nash equilibrium $(\gamma_i)_{i\in I}$ for ν given q exists such that $(W_0(\nu, (\gamma_i)_{i\in I} \mid t))_{t\in T} = U$. By the revelation principle, every continuation equilibrium utility allocation for ν given q is q-IC:

$$\mathcal{U}(\nu, q) \subseteq \boldsymbol{U}(q).$$

These observations lead to the following definition of equilibrium, which is what Myerson (1983) calls an expectational equilibrium:

Definition A.11 A mechanism $\mu : T \to \Delta(A)$ is an *expectational equilibrium* of the informed-designer game iff

1. μ is incentive compatible;
2. for every generalized mechanism ν, a belief $q \in \Delta(T)$ and a continuation Nash equilibrium allocation $(U(t))_{t\in T} \in \mathcal{U}(\nu, q)$ exist such that $U_0(\mu \mid t) \geq U(t)$ for every $t \in T$.

In particular, the set of expectational equilibrium allocations is a subset of the set of IC allocations $\boldsymbol{U}(p)$.

Remark A.2 (Definition of equilibrium) Requiring that agents have a common belief at the beginning of Stage 6 is in the spirit of the belief-consistency requirement of the sequential equilibrium of Kreps and Wilson (1982) and the strong version of perfect Bayesian equilibrium in Fudenberg and Tirole (1991), and it is standard in the literature. We follow Myerson (1983) because two important difficulties emerge in defining sequential equilibrium or a strong version of perfect Bayesian equilibrium directly in our setting. First, the informed-designer game is not a finite game, because the set of possible mechanisms is not finite and not even countable. Second, the definition of sequential equilibrium requires that nature moves at the start of the game with a full-support probability distribution. Whereas nature moves at the start of the informed-designer game to determine the state (t, ω) with a full-support probability distribution, nature also moves later in the game to determine the mechanism's output x, and at that point the mechanism may not have full support.

Proposition A.8 *If μ is an interim-optimal mechanism, then μ is an expectational equilibrium of the informed-designer game.*

Proof. Let μ be an IO mechanism. By definition, it is IC. Fix a deviation of the designer to ν and consider the following fictitious $(n+1)$-player extensive-form game $G(\nu, \mu)$. In the first stage, player 0 chooses $t \in T$. In the second stage, nature draws $\omega \in \Omega$ with probability $\pi(\omega \mid t)$. In the third stage, $(a_0, x_1, \ldots, x_n) \in X$ is drawn with probability $\nu(a_0, x_1, \ldots, x_n \mid t, \omega)$. In the fourth stage, each player $i \in I$ is privately informed about x_i and chooses an action a_i. The utility of player 0 is $u_0(a_0, a_1, \ldots, a_n, t) - U_0(\mu \mid t)$, and for each $i \in I$, the utility of player i is $u_i(a_0, a_1, \ldots, a_n, t)$.

The fictitious game $G(\nu, \mu)$ has an equilibrium in behavioral strategies because it is a finite extensive-form game. Take such an equilibrium profile of behavioral strategies: $q \in \Delta(T)$ for player 0, and $\gamma_i : X_i \to \Delta(A_i)$ for each player $i \in I$. The corresponding expected utility for player 0 is

$$\sum_{t\in T} q(t)(W_0(\nu, (\gamma_i)_{i\in I} \mid t) - U_0(\mu \mid t)),$$

and the expected utility of player $i \in I$ is

$$W_i(\nu, (\gamma_i)_{i\in I} \mid q).$$

By construction, $(\gamma_i)_{i\in I}$ is a Nash equilibrium for $\nu : T \times \Omega \to \Delta(X)$ given q according to Definition A.10, so by the revelation principle $U = (U(t))_{t\in T} = (W_0(\nu, (\gamma_i)_{i\in I} \mid$

$t))_{t\in T}$ is a q-IC allocation; that is, $U \in \boldsymbol{U}(q)$. Let

$$S = \{t \in T : U(t) > U_0(\mu \mid t)\}.$$

If S is non-empty, the equilibrium strategy q of player 0 should assign strictly positive probability to actions in S only, namely, $\text{supp}[q] \subseteq S$. That is, we have $U(t) > U_0(\mu \mid t)$ for every $t \in \text{supp}[q]$. Hence, μ is not an IO mechanism, a contradiction. Therefore, S is empty, which means the belief q and continuation equilibrium allocation U given ν and q constructed in the fictitious game above satisfy $U_0(\mu \mid t) \geq U(t)$ for every t. Hence, for every designer's type, the deviation from μ to ν is not profitable for the designer. Because this construction can be done for every ν, μ is an expectational equilibrium. ■

We conclude the proof of Theorem 2 by Proposition A.8 that implies an IO mechanism is a perfect Bayesian equilibrium mechanism because an expectational equilibrium is a perfect Bayesian equilibrium mechanism.

References

ALONSO, R. AND O. CÂMARA (2016): "Persuading Voters," *American Economic Review*, 106, 3590–3605.

——— (2018): "On the value of persuasion by experts," *Journal of Economic Theory*, 174, 103–123.

ARIELI, I. AND Y. BABICHENKO (2019): "Private bayesian persuasion," *Journal of Economic Theory*, 182, 185–217.

AUMANN, R. J. (1974): "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics*, 1, 67–96.

AUMANN, R. J. AND M. B. MASCHLER (1995): *Repeated Games of Incomplete Information*, Cambridge, Massachusetts: MIT Press.

BALKENBORG, D. AND M. MAKRIS (2015): "An undominated mechanism for a class of informed principal problems with common values," *Journal of Economic Theory*, 157, 918–958.

BANKS, J. AND J. SOBEL (1987): "Equilibrium Selection in Signaling Games," *Econometrica*, 55, 647–662.

BARDHI, A. AND Y. GUO (2018): "Modes of persuasion toward unanimous consent," *Theoretical Economics*, 13, 1111–1149.

BEN-PORATH, E., E. DEKEL, AND B. L. LIPMAN (2019): "Mechanisms with evidence: Commitment and robustness," *Econometrica*, 87, 529–566.

BERGEMANN, D. AND S. MORRIS (2016): "Bayes correlated equilibrium and the comparison of information structures in games," *Theoretical Economics*, 11, 487–522.

——— (2019): "Information design: A unified perspective," *Journal of Economic Literature*, 57, 44–95.

CHAN, J., S. GUPTA, F. LI, AND Y. WANG (2019): "Pivotal persuasion," *Journal of Economic Theory*, 180, 178 – 202.

CHEN, Y. AND J. ZHANG (2020): "Signalling by Bayesian Persuasion and Pricing Strategy," *The Economic Journal*, 130, 976–1007.

CHO, I. K. AND D. KREPS (1987): "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102, 179–221.

DE CLIPPEL, G. AND E. MINELLI (2004): "Two-person bargaining with verifiable information," *Journal of Mathematical Economics*, 40, 799–813.

DEGAN, A. AND M. LI (2021): "Persuasion with costly precision," *Economic Theory*, forthcoming.

DOVAL, L. AND V. SKRETA (2018): "Constrained information design: Toolkit," *arXiv preprint arXiv:1811.03588*.

DWORCZAK, P. AND G. MARTINI (2019): "The simple economics of optimal persuasion," *Journal of Political Economy*, 127, 1993–2048.

ELIAZ, K. AND R. SERRANO (2014): "Sending information to interactive receivers playing a generalized prisoners dilemma," *International Journal of Game Theory*, 43, 245–267.

FARRELL, J. (1993): "Meaning and Credibility in Cheap-Talk Games," *Games and Economic Behavior*, 5, 514–531.

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

——— (2006): "Correlated equilibrium in games with incomplete information revisited," *Theory and decision*, 61, 329–344.

——— (2020): "Games with incomplete information: from repetition to cheap talk and persuasion," *Annals of Economics and Statistics*, 3–30.

FORGES, F. AND F. KOESSLER (2005): "Communication Equilibria with Partially Verifiable Types," *Journal of Mathematical Economics*, 41, 793–811.

FUDENBERG, D. AND J. TIROLE (1991): *Game Theory*, MIT Press.

GROSSMAN, S. J. (1981): "The Informational Role of Warranties and Private Disclosure about Product Quality," *Journal of Law and Economics*, 24, 461–483.

HAGENBACH, J., F. KOESSLER, AND E. PEREZ-RICHET (2014): "Certifiable Pre-Play Communication: Full Disclosure," *Econometrica*, 82, 1093–1131.

HART, S., I. KREMER, AND M. PERRY (2017): "Evidence games: Truth and commitment," *American Economic Review*, 107, 690–713.

HEDLUND, J. (2017): "Bayesian persuasion by a privately informed sender," *Journal of Economic Theory*, 167, 229–268.

KAMENICA, E. (2019): "Bayesian persuasion and information design," *Annual Review of Economics*, 11, 249–272.

KAMENICA, E. AND M. GENTZKOW (2011): "Bayesian Persuasion," *American Economic Review*, 101, 2590–2615.

KOESSLER, F. AND V. SKRETA (2019): "Selling with evidence," *Theoretical Economics*, 14, 345–371.

KREPS, D. M. AND R. WILSON (1982): "Sequential Equilibria," *Econometrica*, 50, 863–894.

LIPNOWSKI, E. AND D. RAVID (2020): "Cheap talk with transparent motives," *Econometrica*, forthcoming.

LIPNOWSKI, E., D. RAVID, AND D. SHISHKIN (2022): "Persuasion via weak institutions," *Journal of Political Economy*, forthcoming.

MASKIN, E. AND J. TIROLE (1990): "The principal-agent relationship with an informed principal: The case of private values," *Econometrica: Journal of the Econometric Society*, 379–409.

——— (1992): "The principal-agent relationship with an informed principal, II: Common values," *Econometrica: Journal of the Econometric Society*, 1–42.

MATHEVET, L., J. PEREGO, AND I. TANEVA (2020): "On information design in games," *Journal of Political Economy*, 128, 1370–1404.

MEKONNEN, T. (2018): "Informed Principal, Moral Hazard, and Limited Liability," *Moral Hazard, and Limited Liability (July 8, 2018)*.

MILGROM, P. (1981): "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12, 380–391.

MYERSON, R. (1983): "Mechanism design by an informed principal," *Econometrica: Journal of the Econometric Society*, 1767–1797.

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

MYLOVANOV, T. AND T. TRÖGER (2012): "Informed principal problems in generalized private values environments," *Theoretical Economics*, 7, 465–488.

——— (2014): "Mechanism Design by an Informed Principal: Private Values with Transferable Utility," *The Review of Economic Studies*, 81, 1668–1707.

OKUNO-FUJIWARA, A., M. POSTLEWAITE, AND K. SUZUMURA (1990): "Strategic Information Revelation," *Review of Economic Studies*, 57, 25–47.

PEREZ-RICHET, E. (2014): "Interim bayesian persuasion: First steps," *American Economic Review*, 104, 469–74.

RICHET-PEREZ, E. AND V. SKRETA (2022): "Test design under falsification," *Econometrica*, forthcoming.

SEIDMANN, D. J. AND E. WINTER (1997): "Strategic Information Transmission with Verifiable Messages," *Econometrica*, 65, 163–169.

SHER, I. (2011): "Credibility and determinism in a game of persuasion," *Games and Economic Behavior*, 71, 409.

TANEVA, I. (2019): "Information design," *American Economic Journal: Microeconomics*, 11, 151–85.

WAGNER, C., T. MYLOVANOV, AND T. TRÖGER (2015): "Informed-principal problem with moral hazard, risk neutrality, and no limited liability," *Journal of Economic Theory*, 159, 280–289.

ZAPECHELNYUK, A. (2022): "On the equivalence of optimal persuasion by uninformed and informed principals," *mimeo*.