Informed Information Design

Frédéric Koessler, Vasiliki Skreta

Informed Information Design*

Frédéric KOESSLER † Vasiliki SKRETA ‡

15 February 2023

Abstract

We study informed persuasion whereby a privately informed designer without ex ante commitment power chooses disclosure mechanisms to influence agents' actions. We characterize the subset of Bayes-correlated equilibria yielding every designer type a payoff higher than what they could get from any disclosure mechanism with credible beliefs. This set of interim optimal mechanisms is non-empty and tractable, and all its elements are perfect Bayesian equilibrium mechanisms of the informed-designer game. Interim optimal mechanisms are characterized via belief-based approaches in pure persuasion settings. We identify single- and multi-agent interactive environments in which ex ante optimal mechanisms are interim optimal.

KEYWORDS: interim information design, Bayesian persuasion, informed principal, Bayes-correlated equilibrium, disclosure games, unraveling, neutral optimum, verifiable types.

JEL CLASSIFICATION: C72; D82.

*We thank the Editor, Emir Kamenica, and four anonymous referees for excellent comments. We also thank Ricardo Alonso, Elchanan Ben-Porath, Francoise Forges, Sergiu Hart, Raphael Levy, Andres Salamanca, Joel Sobel, Tristan Tomala, and Nicolas Vieille for useful feedback as well as seminar participants at Bocconi University, Bonn-Berlin Micro Theory Seminar, Brown, Concordia, CY Paris University, Edinburgh, HEC, Israel Theory seminar, Nancy, Paris Game Theory Seminar, Paris School of Economics, Stanford, Rice, the Workshop in Dynamic Games in Quimper, University of Venice, VSET, Wisconsin, and Yale. Frederic Koessler acknowledges the support of the ANR (StratCom ANR-19-CE26-0010-01). Vasiliki Skreta acknowledges funding by the European Research Council (ERC) consolidator grant 682417 "Frontiers In Design." Francesco Conti and Alkis Georgiadis-Harris provided excellent research assistance.

†HEC Paris and GREGHEC–CNRS, 1 rue de la Libération, 78351, Jouy-en-Josas France. koessler@hec.fr.

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

1 Introduction

Decisions on topics ranging from voting to careers and investments crucially depend 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 designer's purpose is to achieve a certain goal; for example, a seller tries to convince buyers of a product's worth, a politician encourages voters to vote for them, and a pharmaceutical company aims to convince a doctor to prescribe their medicine. In other words, the designer seeks to identify ex ante optimal (EAO) information disclosure mechanisms, which amounts to solving a maximization problem. 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) often have private information that shapes their preferences regarding which procedure to choose. The chosen procedure in turn affects the inferences and the ultimate nature of information that is disclosed.

In this paper, we study *informed* information design. We take the same interim perspective as the influential works on disclosure games[2] but enlarge the designer's choice set. Instead of focusing on deterministic evidence alone, the informed designer chooses any mapping from the state space to distributions over signals. Two key differences exist between the standard information-design setting and our informed persuasion setting. First, the designer's interim incentives differ from their ex ante ones. For example, a high-quality seller prefers to disclose information, but a low-quality seller does not. Similarly, a central bank may choose to fully reveal good news even through ex ante it might have chosen an opaque disclosure rule. 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 features with disclosure games (cf. Milgrom, 1981) and informed-principal problems (cf. Myerson, 1983).

Establishing that a mechanism is part of a perfect Bayesian equilibrium (PBE) of an informed information design game requires showing that for every possible deviation to *any* mechanism, some belief exists alongside some continuation equilibrium play rendering that deviation unprofitable. In general, finding a "correct" combination of belief and continuation equilibrium is difficult, and consequently, so is identifying the mechanisms that are part of a PBE.[3] At the same time, in some settings, any deviation is unprofitable; for example, if there is a state such that the designer always gets the lowest possible payoff when agents assign probability one to that state, any deviation coupled with this belief is unprofitable. In such settings *all* Bayes-correlated equilibria (BCE) are part of a PBE, and this set includes outcomes with the lowest payoffs for the designer. The aforementioned issues with PBE may suggest that we

[1] See, for example, Kamenica and Gentzkow (2011), Bergemann and Morris (2016), Taneva (2019), and Mathevet et al. (2020). Bergemann and Morris (2019), Kamenica (2019), and Forges (2020) provide 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] Example 1 below is a simple binary action, binary state setting in which neither the EAO nor the full disclosure mechanism are part of a PBE.

focus on designer-optimal mechanisms. What is designer optimal, however, depends on the perspective—an EAO mechanism may yield a strictly lower payoff for certain designer types compared with full, partial, or no information disclosure. And in an information design game, designer types can employ such alternative disclosure mechanisms at the interim stage.

In this paper, we provide a tool to tractably identify PBE mechanisms of the informed-designer game, sidestepping the need to consider strategy profiles and associated belief systems. We do so by identifying a class of mechanisms that we call *interim optimal* (IO) mechanisms. IO mechanisms are a tractable subset of BCE mechanisms and are hence defined without reference to the strategies and beliefs of the informed-designer game. In Theorem 1, we show IO mechanisms always exist for general multi-agent information design games. Theorem 2 shows that all IO mechanisms are PBE mechanisms. Furthermore, as their names suggests, IO mechanisms are designer optimal from the interim perspective: IO mechanisms consist of the incentive-compatible mechanisms that yield payoffs for *all* designer types that are higher compared with the best they can obtain by using alternative information disclosure mechanisms with credible beliefs (beliefs that assign positive probability only to types that strictly benefit). As such, IO mechanisms are preferred over full disclosure (unraveling), which may be what the designer prefers when the state is favorable; obfuscation, which may be what the designer prefers when the state is unfavorable; and any other disclosure mechanism coupled with credible beliefs.

An ex ante designer-preferred IO mechanism always exists, and we call it IO*. Proposition 1 and Proposition 4 establish conditions under which IO* and EAO mechanisms coincide. Proposition 4, in particular, implies that in a large class of binary-action settings,[4] the usual ex ante commitment assumption in the information-design literature is without loss: The EAO mechanism is IO*, thus a PBE is robust to interim information disclosure.[5]

IO 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 Lipnowski and Ravid (2020) coin *transparent motives*. When the designer's value function is quasiconvex in beliefs,[6] Proposition 3 shows that each de-

[4]See, for example, Arieli and Babichenko (2019) and Chan et al. (2019), as well as some parametrized examples in Bergemann and Morris (2019), Taneva (2019), and Mathevet et al. (2020).

[5]See Kamenica and Gentzkow (2011) for a discussion of the commitment assumption. Other papers that relax the commitment assumption of the standard information-design paradigm in ways that differ from our approach 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.

[6]Quasiconvexity of the designer's value function naturally arises in many economic environments. Examples include settings in which a salesperson discloses information about the quality of the good with the goal to sell more products (as in Milgrom, 1981 and 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).

signer type getting a payoff weakly higher than that from full disclosure is not only a necessary property, but also sufficient for an incentive-compatible mechanism to be interim optimal.

Proposition 3, in conjunction with Theorem 2, implies full disclosure (i.e., the "unraveling" outcome, which is the unique equilibrium outcome in the leading games on evidence disclosure) is IO and thus a PBE outcome even if the informed party can choose arbitrary mechanisms. We leverage Propositions 2 and 3 to build a constrained information-design program characterizing IO* mechanisms. We illustrate interim optimality in simple examples.

Related literature We contribute to the information design literature (Kamenica and Gentzkow, 2011, Bergemann and Morris, 2016, Taneva, 2019) by identifying a tractable subset of BCE mechanisms that are robust to interim information disclosure and arise at a PBE of the informed-designer game. We also contribute to the informed principal literature stemming from Myerson (1983), by developing a concept that always exists and is a PBE of the informed-designer game. Clearly, when the designer has access to all experiments, the optimum can be achieved by choosing and committing to the experiment before learning additional information, which is not generally true if the set of experiments is restricted. In a pioneering paper, Perez-Richet (2014) studies equilibrium refinements and what happens when an informed designer can choose from constrained-information policies in a single-agent setting with binary actions and states and state-independent payoffs for the designer.[7] Degan and Li (2021) consider a binary-action setting in which the informed sender has a restricted choice set (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. In 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. We provide more detailed comparisons with related works throughout the paper, such as the discussions following the definition of the informed-designer game in Section 3, the last part of Section 4, and Appendix C.

The rest of the paper is structured as follows. The next section studies an informed prosecutor example. Section 3 describes the setting, formulates the informed-designer game, and defines EAO and EPO mechanisms. In Section 4, we define interim-optimal mechanisms and prove that they exist and that they are PBE outcomes of the game. Section 5 provides a belief-based characterization of interim optimality. In Section 6, we study interim optimality in multi-agent settings. Section 7 concludes. In Appendix A, we present the general model in which the designer can be imperfectly informed about the state and can commit to enforceable actions. We prove Theorem 1 and Theorem 2 directly for the general setting of Appendix A in Appendix A.1 and Appendix A.2, respectively. Appendix B has the proofs of Section 5. In Appendix C,

[7] Without putting some reasonable restrictions on beliefs, PBE 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. This observation generalizes to some information-design settings with state-independent preferences for the designer, where the EAO mechanism is a PBE mechanism even when it is strictly dominated by the ex post optimal (EPO) mechanism for some designer types (see Zapechelnyuk, 2022).

we compare interim optimality to other leading concepts in the informed-principal literature. Appendix D studies an imperfectly informed prosecutor example.

2 Informed prosecutor

In this section, we provide an introduction to IO mechanisms by reconsidering the leading judge example of Kamenica and Gentzkow (2011), and we illustrate why PBE can have very little predictive power and an EAO mechanism may not be a robust prediction (despite being a PBE) in informed information design games.

Suppose a judge is facing a defendant that is either guilty or innocent, $T = \{t_G, t_I\}$, and the prior on the defendant being guilty is $p(t_G) = p$. The judge's set of actions is a subset of $\{\underline{a}^0, a^2, a^3, \bar{a}^0\}$. The prosecutor is the information designer (player 0) and has state-independent payoffs: $u_0(\underline{a}^0) = u_0(\bar{a}^0) = 0$, $u_0(a^2) = 2$, and $u_0(a^3) = 3$. In contrast to the version in Kamenica and Gentzkow (2011) when the prosecutor chooses the investigation procedure, they know whether the defendant is guilty or not.

We consider three variations that only differ in the number of actions. The first variation corresponds to the judge example in Kamenica and Gentzkow (2011). The judge has two actions: $\underline{a}^0 =$ *acquit* and $a^2 =$ *convict with two decades imprisonment*. The judge's optimal action as a function of their belief $q(t_G) = q$ is $\underline{a}^0$ if $q < 1/3$ and a^2 if $q \geq 1/3$.[8] The next variation arises from adding action a^3 to the previous setting: Now the judge can rule in addition $a^3 =$ *convict with three decades imprisonment*. We assume a^3 is optimal for the judge when $q \geq \frac{2}{3}$. In the last variation, we add a fourth action, $\bar{a}^0 =$ *execution*, and assume $\bar{a}^0$ is the judge's optimal action for $q > \bar{q}$, where $\bar{q} > \frac{2}{3}$. This extreme action yields a payoff of 0 to the prosecutor.

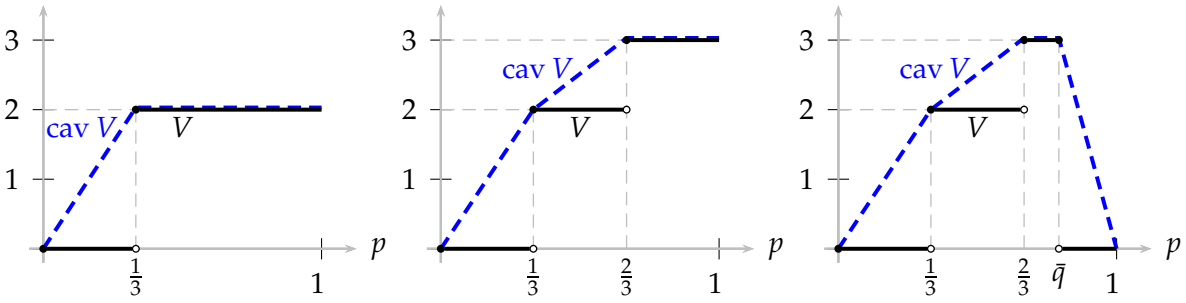

Figure 1: Designer value function (solid lines) and ex ante payoff at the EAO mechanism (dashed lines). Left panel: two actions ($A = \{\underline{a}^0, a^2\}$); middle panel: three actions ($A = \{\underline{a}^0, a^2, a^3\}$); right panel: four actions ($A = \{\underline{a}^0, a^2, a^3, \bar{a}^0\}$).

Figure 1 depicts the designer's expected payoff as a function of p. The EAO distribution of posteriors is obtained from the concavification of the value function V, denoted by cav V, and it is depicted with the dashed line in Figure 1.[9] Note for $p \leq \frac{1}{3}$ the EAO distribution of posteriors is the same for all variations: it splits p (with prior-dependent weights) to two posterior beliefs 0 and $\frac{1}{3}$ inducing the designer-EAO payoff vector $U^{EAO} = (U(t_G), U(t_I)) = (2, \frac{4p}{1-p})$.

[8] For brevity, we only describe the judge's optimal action as a function of the belief and do not provide the payoff details that lead to that function (as usual ties are broken in favor of the prosecutor).

[9] In single-agent settings, the set of BCE mechanisms can be described by Bayes' plausible distributions of posteriors. The concavification of V is the smallest concave function that is point-wise greater than or equal to V.

Can this EAO payoff vector arise at a PBE of the informed prosecutor game and, more broadly, what is the set of PBE payoff vectors for the prosecutor? The answer to the second part of this question is surprisingly simple and also answers the first part. In this example, PBE (interim) payoff vectors (for the designer) coincide with *all* BCE payoff vectors. To see why, observe that action $\underline{a}^0$ leads to the lowest possible payoff for the designer, $(U(t_G), U(t_I)) = (0, 0)$, and it is the agent's unique optimal action for $q = 0$. Therefore, any mechanism yielding a BCE payoff vector is part of a PBE because any off-path deviation coupled with belief $q = 0$ after the observation of the deviation mechanism, and after the observation of any message from this mechanism, leads to action $\underline{a}^0$, rendering the deviation unprofitable.[10] BCE induce very low payoff vectors, for example, $(U(t_G), U(t_I)) = (0, 0)$, which is the payoff arising from no disclosure. Consequently, PBE has unsatisfactory predictive power. In fact, the definition of PBE permits off-path beliefs that make the information revealed from the subsequent experiment irrelevant. Further, it is quite unreasonable for the prosecutor who knows, or is sufficiently convinced,[11] that the defendant is guilty to settle at the interim stage for a payoff lower than the unraveling (full disclosure) payoff.

To identify which payoff vectors are robust to unraveling and, more broadly, to beliefs that the designer could credibly induce through disclosure, consider the table that follows in which we describe the designer's payoff vectors in the two-, three-, and four-action variations of the example arising from various disclosure mechanisms: no disclosure (ND), full disclosure (FD), EAO disclosure (inducing two posteriors 0 and $\frac{1}{3}$), and partial disclosure (PD) inducing signals i and g with posteriors 0 and q, where $q > p$ is a posterior inducing the highest payoff for the designer (i.e., $q \geq \frac{1}{3}$ in the two-action example, $q \geq \frac{2}{3}$ in the three-action example, and $q \in [\frac{2}{3}, \bar{q}]$ in the four-action example). All these payoff vectors arise at some PBE of the informed-designer game. We focus on $p \in (0, \frac{1}{3})$ and revisit the example for the remaining priors after we formally define IO in Section 4.

prior $p \in (0, \frac{1}{3})$	$A = \{\underline{a}^0, a^2\}$	$A = \{\underline{a}^0, a^2, a^3\}$	$A = \{\underline{a}^0, a^2, a^3, \bar{a}^0\}$
No disclosure	$0, 0$	$0, 0$	$0, 0$
Full disclosure	$2, 0$	$3, 0$	$0, 0$
Partial disclosure	$2, 2\frac{(1-q)p}{q(1-p)}$	$3, 3\frac{(1-q)p}{q(1-p)}$	$3, 3\frac{(1-q)p}{q(1-p)}$
EAO disclosure	$2, \frac{4p}{1-p}$	$2, \frac{4p}{1-p}$	$2, \frac{4p}{1-p}$

Table 1: Designer's interim payoff vectors $(U(t_G), U(t_I))$

In the two-action variation, there is no ex ante and interim conflict: Both designer types prefer the EAO payoff vector that corresponds to concavification, $U^{EAO} = (2, \frac{4p}{1-p})$, to the one corresponding to full disclosure, $U^{FD} = (2, 0)$, or to any other disclosure mechanism. In addition, because designer type t_G gets their first-best payoff, any alternative mechanism proposal by the designer can only strictly benefit type t_I, so belief credibility should imply that the agent assigns probability zero to t_G and so chooses action $\underline{a}^0$. Therefore, in the two-action variation, the EAO payoff vector (and any BCE

[10] Note that off-path beliefs that do not have full support may be conflicting information subsequently released from off-path mechanisms. Consider, for example, a deviation to a fully revealing mechanism coupled with off-path belief $q = 0$. This mechanism may reveal that the defendant is guilty for sure, yielding inconsistency. In such an instance, Bayes' rule does not apply and the definition of PBE permits any subsequent off-path belief. In particular, the belief we started with, $q = 0$, is PBE-consistent.

[11] In Appendix D we revisit this example and allow for a partially informed prosecutor.

that gives a payoff of 2 to t_G) is robust to interim information disclosure and credible beliefs, and EAO and IO* (the ex ante preferred IO) mechanisms coincide.

In the three-action variation, the payoff vector corresponding to full disclosure is $U^{FD} = (U(t_G), U(t_I)) = (3, 0)$, and the payoff vector corresponding to partial disclosure is $U^{PD} = (U(t_G), U(t_I)) = (3, 3\frac{(1-q)p}{q(1-p)})$. Because we cannot compare the interim payoff vector U^{EAO} with U^{FD} or U^{PD}, there is now ex ante and interim conflict: Type t_G strictly prefers full or partial disclosure to EAO, whereas the opposite holds for t_I. The EAO mechanism is not interim optimal because type t_G strictly prefers full disclosure (t_G gets a payoff of 3 instead of 2), and a belief for the judge that assigns probability one to type t_G is credible because full disclosure strictly benefits type t_G.[12]

In the four-action variation, the ideal action for the prosecutor is only obtained for *interior* beliefs. As in the previous variation, the EAO mechanism is not interim optimal because belief $q \in [\frac{2}{3}, \bar{q}]$ induces action a^3, which strictly benefits both types. To understand why robustness to belief $q \in [\frac{2}{3}, \bar{q}]$ is related to robustness to interim disclosure, it is convenient to represent partial disclosure as a partition of a modified state space $\widetilde{T}$ as follows. Assume for simplicity that the prior is $p = \frac{1}{6}$.

First, consider an hypothetical situation in which the designer is only partially informed about the state. Their type is in $\widetilde{T} = \{t^i, t^g\}$, where t^i knows that the defendant is innocent and t^g believes that the defendant is guilty with probability $q \in [\frac{2}{3}, \bar{q}]$; t^g has probability $\frac{1}{6q}$ and t^i has probability $1 - \frac{1}{6q}$. Now, exactly as in the three-action example, full disclosure of $\widetilde{t} \in \{t^i, t^g\}$ should induce belief in the defendant being guilty equal to $q \in [\frac{2}{3}, \bar{q}]$ for the agent. Hence, for type t^g, action a^3 should be played with probability one, which implies that the EAO mechanism (that never induces an interim payoff higher than 2) is not interim optimal. The EAO mechanism is not robust to unraveling when the designer is partially informed.

Next, we extend the argument by starting from the previous scenario and endowing the designer with an additional, completely informative signal. Specifically, the partially informed designer type t^g becomes type t^g_G by learning that the defendant is guilty (which happens with probability q) or becomes type t^g_I by learning that the defendant is innocent (which happens with probability $1 - q$). The type space is now $\widetilde{T} = \{t^i, t^g_I, t^g_G\}$, with prior $(1 - \frac{1}{6q}, \frac{1-q}{6q}, \frac{1}{6})$. In this setting, which is informationally equivalent to the original one, partial disclosure can be written as a partition that reveals $\{t^i\}$ or $\{t^g_I, t^g_G\}$. Compared with the EAO outcome, t^g_I and t^g_G strictly benefit from partial disclosure with the credible belief that assigns the deviation from the EAO mechanism to t^g_I and t^g_G and assigns belief $\Pr(t^g_G \mid \{t^g_I, t^g_G\}) = q \in [\frac{2}{3}, \bar{q}]$ to the defendant being guilty. More generally, information disclosure can be represented as in works by Green and Stokey (1978), Gentzkow and Kamenica (2017), and Brooks, Frankel, and Kamenica (2022) by a partition of $\widetilde{T} = [0, 1] \times \{t_G, t_I\}$, where the designer knows $(x, t) \in \widetilde{T}$, and the element x of $[0, 1]$ is drawn uniformly. The above partially revealing experiment can be represented by the partition $\{\{g\}, \{i\}\}$, where $i = [0, \frac{6q-1}{5q}] \times \{t_I\}$ and g is the complement. Then, $\Pr(t_G \mid g) = q \in [\frac{2}{3}, \bar{q}]$ and partial revelation strictly benefits the designer for all $\widetilde{t} \in g$.

[12] This unraveling force is at the core of the evidence disclosure works mentioned in the introduction: A high-quality seller who can certify quality does so and charges a high price or sells high quantities.

Hence, in both the three- and four-action variations, for a payoff vector to be interim optimal, it must ensure t_G a payoff of at least 3. In the four-action variation, IO payoff vectors set $U(t_G) = 3$ and $U(t_I) \in [\frac{3(1-\bar{q})}{5\bar{q}}, \frac{3p}{2(1-p)}]$. To obtain the IO set in the three-action variation, we simply set $\bar{q} = 1$. IO* is the same in both variations and is obtained by splitting the prior p to posteriors 0 and $\frac{2}{3}$, yielding a payoff vector $(U(t_G), U(t_I)) = (3, \frac{3p}{2(1-p)})$. Whenever the concavification outcome is robust to "generalized" unraveling forces, as in the case of the two-action variation, it corresponds to IO*; whenever not, the example illustrates that the IO set identifies BCE outcomes that are robust and IO* selects the ex ante designer-preferred one.

3 Model

Environment 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 $I = \{1, \ldots, n\}$ 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.[13] 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 payoff of the designer is $u_0(a, t)$ and the payoff 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

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

Informed-designer game The informed-designer game is the following extensive-form game between the privately informed 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[14] $X = \prod_{i \in I} X_i$ 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 the signal $x_i \in X_i$.

[13] 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 Appendix A.

[14] 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.

Key comparisons The key difference between the informed-designer extensive form 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). This standard 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 pioneered by Myerson (1983). The setting is a common value one in Maskin and Tirole's (1992) terminology because typically the state of the world affects agents' payoffs. Our setup differs from the usual formulations of informed-principal problems in two ways. First, in contrast to Maskin and Tirole (1992) and the majority of works on informed-principal problems,[15] the principal has no truth-telling constraints. Second, the experiment's outputs are signals rather than contractually enforceable outcomes,[16] which makes our game closer to an informed-principal setting with moral hazard.[17]

Chen and Zhang (2020) and Hedlund (2017) study signaling single-agent settings in which the designer is partially informed and chooses an experiment that reveals information about a payoff-relevant state but not directly about their type. More precisely, in those papers, as in our general setting in Appendix A, the state of the world is $(t, \omega) \in T \times \Omega$, and the designer knows only $t \in T$. The designer in Chen and Zhang (2020) and Hedlund (2017) chooses $\mu : \Omega \to \Delta(X)$, whereas the designer in our game chooses a mechanism $\mu : T \times \Omega \to \Delta(X)$. In those papers, the mechanism does not condition directly on the designer's type t, whereas it does in our setting, exactly as it does in Myerson (1983) and in the information design literature (Bergemann and Morris, 2016, 2019). Consequently, our setting and our modeling of mechanisms—the object of choice of the designer–coincide with those in Bergemann and Morris (2016, 2019).

Our informed-designer game is also related to the games studied in the literature on strategic information disclosure as in Grossman (1981) and Milgrom (1981). 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 information-disclosure mechanism.

Strategies and PBE definition 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 they choose disclosure mechanisms (functions from states to distributions over signals). For the designer, a strategy specifies for each $t \in T$ an information-disclosure mechanism $\nu : T \to \Delta(X)$. For each possible mechanism ν and for each possible private signal x_i from that mechanism, agent i's strategy specifies a probability distribution over A_i. Appendix A.2 contains formal definitions of

[15] 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.

[16] Our formulation straightforwardly adjusts to accommodate contractible actions, and we do so in Appendix A.

[17] See, for example, Wagner et al. (2015) and Mekonnen (2021).

strategies and continuation games, as well as a number of auxiliary results owing to Myerson (1983).

Below, we rely on the revelation and the inscrutability principles (Myerson, 1983) and provide a simplified formulation of PBE, expressed simply in terms of direct mechanisms instead of strategies in the informed-designer game. This formulation is standard in the informed-principal literature and relies on the following ideas. A strategy profile induces an outcome $\mu : T \to \Delta(A)$, henceforth called a *direct recommendation mechanism*, which specifies a probability distribution over action profiles for each state. Consider any Nash equilibrium or PBE strategy profile inducing $\mu : T \to \Delta(A)$. The strategy of the designer can be replaced by a pooling strategy that specifies the same direct mechanism μ for all designer types. Along the equilibrium path (i.e., when the designer uses μ), each agent's strategy is replaced by an obedient strategy. Specifically, for every $a_i \in A_i$, every agent i plays action a_i when they receive the signal a_i from μ. In that way, the same outcome μ is implemented along the path. Agents get less information than in the original equilibrium because (i) they learn nothing about the designer's type by observing μ (inscrutability), and (ii) they learn the minimal information about the state and others' signals by their recommended action (revelation principle). A direct mechanism is called a PBE mechanism if it is an outcome of a PBE strategy profile and belief system. We proceed to formally define obedient mechanisms and PBE.

Incentive compatible and equilibrium mechanisms A direct mechanism $\mu : T \to \Delta(A)$ is a recommendation system, where $\mu(a_1, \ldots, a_n \mid t)$ is probability that the mechanism privately recommends a_i to each agent i when the actual type of the designer is t. An equilibrium mechanism of the informed-designer game must be obedient, henceforth *incentive-compatible*, given agents' prior p. In a PBE, every off-path continuation outcome should also be incentive compatible for some belief $q \in \Delta(T)$ for the agents. Formally, the mechanism μ is *q-incentive compatible* (q-IC) iff for each agent i obedience is optimal if all the other agents are obedient; that is, for every a_i and a_i' in A_i

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

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 BCE of Γ.

Let

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

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

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

Denote by $\mathcal{U}(\nu, q)$ the set of continuation equilibrium payoff vectors for the designer given the mechanism $\nu : T \to \Delta(X)$ and belief q. By the revelation principle, $\mathcal{U}(\nu, q) \subseteq \boldsymbol{U}(q)$. The following definition of PBE corresponds to the definition of expectational

equilibrium in Myerson (1983), a suitable version of PBE of the informed-designer game in the spirit of sequential equilibrium.

Definition 1 (PBE mechanisms) A mechanism $\mu : T \to \Delta(A)$ is a *perfect Bayesian equilibrium* (PBE) of the informed-designer game if

1. μ is incentive compatible.
2. For every mechanism $\nu : T \to \Delta(X)$, a belief $q \in \Delta(T)$ and a continuation equilibrium payoff vector $(U(t))_{t\in T} \in \mathcal{U}(\nu, q)$ exist such that $U_0(\mu \mid t) \geq U(t)$ for every $t \in T$.

The first condition is a necessary and sufficient condition for each agent's strategy to be a best response to other agents' strategies and the designer's strategy. The difference between a Nash equilibrium and a PBE of the informed-designer game stems from the second condition. In a Nash equilibrium, the payoff vector $(U(t))_{t\in T}$ arising from a deviation to a mechanism ν can arise from any continuation strategy profile for the agents given ν. By contrast, in a PBE the payoff vector $(U(t))_{t\in T}$ should be a continuation equilibrium payoff vector for some belief q.[18] Except for particular cases, such at those in which the designer has state-independent preferences, a PBE is supported by continuation beliefs and payoff vectors that *depend* on the specific mechanism ν the designer deviates to.

Ex ante optimal and ex post optimal mechanisms A mechanism μ is *ex post incentive compatible*[19] 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.[20] We now formally define the concepts of EAO and EPO mechanisms.

Definition 2 (EAO mechanisms) A mechanism μ is *ex ante optimal* (EAO) if μ is incentive compatible, and for every other incentive-compatible mechanism ν, we have

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

[18] Note that if $q(t) = 0$ for some type t, then the off-path belief may conflict with information provided by a deviation mechanism if this mechanism fully identifies type t. In this case, the continuation equilibrium condition in Definition 1 (ii) implicitly assumes that type t is assigned zero probability regardless of the message generated by the mechanism. See also Remark A.1.

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

[20] 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 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.

An EAO mechanism corresponds to the solution of the standard (uniformed) information-design problem (Kamenica and Gentzkow, 2011, Bergemann and Morris, 2019, Taneva, 2019).

Definition 3 (EPO mechanisms) 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 payoff vector is the best correlated equilibrium payoff vector for the designer when t is commonly known, and it is the solution for the designer under complete information.

Towards a new concept: interim optimality In Section 4, we introduce interim optimality, a concept that lies between ex ante and ex post optimality, and argue why it is an appealing concept for informed-designer games. Before doing so we explain why existing concepts have shortcomings that render them less suitable, in general, for informed-designer games. We start with a simple binary state and action example in which neither EAO nor EPO mechanisms are part of a PBE:

Example 1 (State-dependent preferences, binary actions) Suppose that there is only one agent, two states $T = \{1, 0\}$, and two actions for the agent $A = \{a^1, a^2\}$. The designer's and the agent's payoffs are summarized in the following matrix:

	a^1	a^2
$t = 1$	$3, 0$	$0, 1$
$t = 0$	$0, 1$	$1, 0$

Let $q(1) = q$ 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 payoff 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 $p = \frac{3}{4}$, 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

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

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 payoff vector is $U^{EAO} = (1, 0)$. This payoff vector is not a PBE payoff vector of the informed-designer game. In fact, it is not even a Nash equilibrium payoff vector: the designer can deviate to any non-revealing mechanism 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 $\beta < 1$, $t = 0$ strictly benefits, implying that at least one of the two designer types benefits regardless of the value of β. It is clear that the same argument

shows that the EPO payoff vector, $U^{EPO} = (0, 0)$, is not a Nash equilibrium payoff vector.

A natural remedy to the aforementioned issues with EAO and EPO is to simply focus on PBE. Doing so is not compelling for two reasons: tractability and predictive power. Tractability can be an issue because the characterization of PBE of an informed-designer game in general requires identifying the correct belief-continuation play combination for each (not necessarily direct) deviation to mechanism ν that renders ν unprofitable. Finding this correct combination may require characterizing *all* continuation equilibria of ν for *all* possible beliefs! Coming to the second point, the informed prosecutor example underscores that PBE predictions could be weak and unreasonable, because off-path beliefs can be cooked so that informative disclosure mechanisms are totally ignored by the agents. In particular, some PBE mechanisms are not immune to full disclosure, and are therefore inconsistent with simple information unraveling arguments à la Milgrom (1981). In particular, in the three-action variation of the informed prosecutor example, the EAO mechanism is a PBE but is not immune to full disclosure.

4 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 robustly select, in the sense that there is no alternative disclosure mechanism the designer could deviate to when facing agents making mechanism-consistent inferences. 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 payoff weakly higher than their EPO payoff. In other words, no designer type t can strictly prefer an alternative mechanism that is IC for $q = \delta_t$ (IC given a belief that assigns probability 1 to t). More generally, because the designer can also partially reveal the state, interim optimality requires that each designer type t does not strictly prefer an alternative mechanism that is IC given a belief q, credible in the sense that q assigns positive probability *only* to designer types who strictly benefit from this alternative mechanism. The set of interim-optimal mechanisms that we define next is robust to such alternative credible mechanism-belief pairs.

Definition 4 (IO mechanisms) 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 $t \in \text{supp}[q]$.

The notion of interim optimality is strong for two reasons. First, the definition of IO implies that 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. Second, the designer is able to choose any belief q that satisfies the credibility requirement.

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. 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 PBE of the informed-designer game and the designer cannot credibly select a better mechanism.[21] Likewise, when an EPO mechanism (the *unraveling* outcome) is IO (see Corollary 1), it satisfies the properties mentioned above. Theorem 2 shows that an IO mechanism also has a solid game-theoretic foundation because it is a PBE of an informed-designer game.

The next proposition shows that if the EPO payoff vector is EAO, it is the unique IO payoff vector. 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 payoff vector is ex ante optimal, then it is the unique interim-optimal payoff vector.*

In other words, Proposition 1 shows that if an EAO payoff vector can be obtained by a fully revealing mechanism, this fully revealing mechanism is IO and the corresponding payoff vector is the unique IO payoff vector.[22] In particular, the IO payoff vector is unique when for every t, a correlated equilibrium of the complete-information game at t exists that gives the first-best payoff to the designer.

Theorem 1 establishes the existence of IO mechanisms for every Bayesian incentive problem Γ. In other words, the statement of Theorem 1 does not impose any additional assumptions on any of the elements of Γ.

Theorem 1 (IO mechanisms exist) *For any Bayesian incentive problem Γ, at least one interim-optimal mechanism exists.*

We prove this result in Appendix A.1. 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 payoff vectors" as follows.

Let $B^{IO}(\Gamma)$ be the set of payoff vectors $U \in \mathbb{R}^T$ such that a belief $q \in \Delta(T)$ and a q-IC payoff vector U' exist such that $U'(t) > U(t)$ for every $t \in \text{supp}[q]$. By definition, a payoff vector U is an IO payoff vector if 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 payoff vectors, and thus, the set of IO payoff vectors is non-empty. These axioms are defined in Myerson (1983) and, for completeness, we include their formal definitions in Appendix A.1.

[21] As already discussed, many PBEs could exist that are based on adversarial beliefs and continuation equilibria.

[22] Observe that the EPO payoff vector is always unique but multiple EAO payoff vectors may exist.

We establish our second main result, Theorem 2, that shows an IO mechanism is a PBE mechanism of the informed-designer game.

Theorem 2 (IO mechanisms are PBE) *If μ is an interim-optimal mechanism, then μ is a perfect Bayesian equilibrium mechanism of the informed-designer game.*

Example 1 is a binary-state, binary-action setting with a single agent in which the EAO mechanism is not a Nash equilibrium of the informed-designer game. Consequently, by Theorem 2, this mechanism is not IO. In this example, neither the EAO payoff vector $U = (1, 0)$ nor the EPO payoff vector $(0, 0)$ are IO because they are not equilibrium payoff vectors.[23] The payoff vector $U = (0, 1)$, which is simply obtained by a non-revealing experiment, is an IO payoff vector and a PBE payoff vector by Theorem 2.

Ex ante preferred IO The set of IO payoff vectors has a nice mathematical structure and, given a prior, we can identify an IO vector that is ex ante preferred, which we call IO*. More precisely, the set of IO payoff vectors of Bayesian incentive problem Γ, $\boldsymbol{U}^{IO}(\Gamma)$, is the intersection of the set of BCE payoff vectors $\boldsymbol{U}(p)$ and the IO-unblocked payoff vectors $\mathbb{R}^T \backslash B^{IO}(\Gamma)$, that is, $\boldsymbol{U}^{IO}(\Gamma) = \boldsymbol{U}(p) \cap (\mathbb{R}^T \backslash B^{IO}(\Gamma))$, which is compact ($\boldsymbol{U}(p)$ is compact and $\mathbb{R}^T \backslash B^{IO}(\Gamma)$ is closed). Then, an IO* payoff vector solves

$$\max_{U \in \boldsymbol{U}^{IO}(\Gamma)} \sum_{t \in T} p(t) U(t).$$

This program has a solution because the objective is linear and the choice set is compact.

Informed prosecutor revisited To illustrate the notions of interim optimality and of IO*, consider again the informed prosecutor example of Section 2, where we have seen that, regardless of the prior, all IC mechanisms are PBE mechanisms. In the two-action variation, first observe that the payoff vector $(2, 0)$ is q-IC for $q = 1$ together with a fully revealing mechanism. This implies that in every IO mechanism, designer type t_G should get their first-best payoff (2). Second, every IC mechanism that gives payoff 2 to type t_G is IO because, starting from such candidate mechanism, belief credibility in Definition 4 implies $q = 0$, and the only q-IC payoff when $q = 0$ is 0. In particular, whatever the prior, the EPO mechanism (full information disclosure) is IO because it induces the payoff vector $(2, 0)$. The EAO mechanism is also IO in this version of the example because the EAO mechanism always gives payoff 2 to type $t = t_G$. More generally, an EAO payoff vector is IO in every single-agent setting with binary actions and state-independent preferences for the designer (see Proposition 4). In this version of the example, the set of IO payoff vectors is the convex hull of the EPO payoff vector, $(2, 0)$, and the EAO payoff vector, $(2, \min\{\frac{4p}{1-p}, 2\})$.

Next, consider the three- and four-action variations. Observe that the payoff vector $(3, 3)$ is q-IC (with a non-revealing mechanism that recommends action a^3 with probability 1) for every belief q in $[\frac{2}{3}, \bar{q}]$. This implies that in every IO mechanism, at least

[23] This finding is in contrast to the setting in De Clippel and Minelli (2004), where the EPO payoff vector (and any IC payoff vector that gives higher payoff to each designer type) is a PBE payoff vector. 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.

one designer type should get their first-best payoff (3). Hence, in both these versions of the example, the EAO mechanism is IO if $p \geq \frac{2}{3}$. It also implies that the fully revealing mechanism is not IO when there are four actions because it induces the payoff vector $(0, 0)$. When the prior is $p = \frac{1}{6}$, in both the three- and four-action variations the set of IO payoff vectors is the set of payoff vectors U such that $U(t_G) = 3$ and $U(t_I) \in [0, \frac{3}{10}]$. The IO* payoff vector is $(U(t_G), U(t_I)) = (3, \frac{3}{10})$ and is obtained from the following direct recommendation mechanism:[24]

Figure 2: Direct recommendation mechanism yielding the IO* payoff vector$(U(t_G), U(t_I)) = (3, \frac{3}{10})$.

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 payoff for the designer is $\frac{3}{4}$, which is strictly lower than the EAO payoff cav $V(\frac{1}{6}) = 1$.

We return to this example in Section 5 after we provide a belief-based characterization of interim optimality.

Interim optimality and other concepts Conceptually and in terms of motivation, interim optimality is related to the axiomatic notion of neutral optimum defined in Myerson (1983). It is also related to other notions in the informed-principal literature that identify specific subsets of incentive-compatible mechanisms as we elaborate in Appendix C. Interim optimality is most closely related 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 payoff vectors is a subset of core payoff vectors (Myerson, 1983); see Proposition C.1. On the other hand, it is a superset of the set of strong neologism-proof payoff vectors (Mylovanov and Tröger, 2012, 2014; Wagner et al., 2015); see Proposition C.2. The difference in the sets of core payoff vectors (Myerson, 1983) and those of interim-optimal payoff vectors stems from the beliefs that can accompany alternative mechanism proposals. The set of strong neologism-proof payoff vectors and that of strong unconstrained Pareto optimal payoff vectors (Maskin and Tirole, 1990) both differ from the set of interim-optimal payoff vectors because both concepts allow types that weakly (rather than strictly) benefit from blocking. Within the context of information design, the set of core payoff vectors may contain vectors that are not perfect Bayesian (and even Nash) equilibrium payoff vectors (see Example 1). At the same time, strong neologism-proof and strong unconstrained Pareto optimal payoff vectors may fail to exist in our setting (see Example 4). Thus, in general, the concepts

[24] In this example, the IO* payoff vector is uniformly preferred to all IO payoff vectors by all designer types. There are examples in which some IO vector other than the IO* payoff vector is preferred by some designer types.

of core, strong neologism proofness, and strong unconstrained Pareto optimality are not suitable for informed-designer games.[25]

Interim optimality and equilibrium refinements Interim optimal mechanisms are a subset of BCE mechanisms, and as such, the definition of interim optimality does not hinge on the informed-designer game or on its PBE strategy profiles and belief systems. However, Theorem 2 implies interim optimality is de facto a refinement of PBE. Indeed, the informed prosecutor example illustrates that the set of IO mechanisms can be much smaller than the set of PBE mechanisms. A natural question pertains to how the set of IO mechanisms compare with mechanisms selected by common refinements in signaling games in the spirit of the intuitive criterion (Cho and Kreps, 1987) or divinity criterion (Banks and Sobel, 1987). Strictly speaking, the aforementioned PBE refinements are not formally defined for informed-designer games and may, a priori, not exist.[26] Even if we could define those refinements, they would not select IO mechanisms, and in fact, in some informed designer settings, they have no power to select equilibria. For instance, in settings analogous to that of four-action variation of the informed prosecutor example, every incentive compatible payoff vector, in particular, the worst payoff vector $(0, 0)$, would survive such refinements regardless of the prior.

These refinements are harder to adapt to multi-agent settings, and even if we did adapt them, they would result in different predictions. To illustrate this point in the simplest possible setting, consider a designer with only one possible type (or multiple payoff-irrelevant types) facing 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 that players 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 one, constitutes a PBE and survives all refinements as, trivially, with one type off-path beliefs play no role.

5 Belief-based characterization of interim optimality in pure persuasion settings

In Bayesian persuasion, Kamenica and Gentzkow (2011) write the designer's ex ante payoff V 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 $\operatorname{cav} V$, 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

[25] The notions of Rothschild-Stiglizt-Wilson payoff vectors of Maskin and Tirole (1992) and those of assured payoff vectors 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.

[26] Even if we adapted their definition they may not be tractable because to identify the set of beliefs consistent with each refinement, it would be necessary to consider *each* deviating (possibly non-direct) mechanism and belief-continuation pair and then compare, for *each* type of the designer, the set of beliefs that could benefit this type compared with the candidate PBE.

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.[27] We also assume the payoff of the designer is state independent: For every state t and action a, the payoff of the designer is equal to $u_0(a)$.[28]

For every $q \in \Delta(T)$, let $A^*(q)$ be the agent's optimal set of actions when their 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 payoff of the designer when the agent's belief is q:

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

The next proposition provides 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 payoff of the designer is state independent. Then, an incentive-compatible payoff vector $U \in \mathbb{R}^T$ is interim optimal iff*

$$\text{There is no } q \in \Delta(T) \text{ such that } V(q) > U(t) \text{ for every } t \in \operatorname{supp}[q]. \tag{1}$$

Proposition 2 implies that to check whether an IC payoff vector U is IO, it suffices to check a finite number of inequalities that only rely on (1), that is, on the comparison of U to the value function V. To see that condition (1) can be rewritten as a finite number of inequalities, for every $S \subseteq T$, let $u^*(S)$ be the highest payoff of the designer from an action that is optimal for the agent for some belief with support S. Formally,

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

Then, (1) 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 payoff vector 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, focusing on splittings (distributions of posteriors) with finite support (of cardinality at most $|A|$) does not sacrifice generality.

[27] 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 payoff the designer can get when agents play a Nash (instead of Bayes correlated) equilibrium of the symmetric-information game given belief q.

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

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 $t \in T$

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

Therefore, we can leverage Proposition 2 to build the following constrained information-design program that characterizes the IO* mechanism:

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

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 the informed prosecutor example with four actions. We have

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

Hence, the set of IO payoff vectors is the set of payoff vectors induced by some splitting of the prior such that $U(t) = 3$ for some $t \in \{t_G, t_I\}$. This condition implies that at least one of the designer's type induces posteriors only in the interval $[\frac{2}{3}, \bar{q}]$. By Bayes' rule, this type can only be t_G if $p < \frac{2}{3}$, and t_I if $p > \bar{q}$. If the other designer type t induces some posterior q outside this interval with positive probability, the designer's type is revealed to the agent: The posterior is $q = 1$ if $t = t_G$ and $q = 0$ if $t = t_I$. Whatever the (interior) prior p, the EPO payoff vector $(0, 0)$ is not IO, and the EAO payoff vector is IO if $p \geq \frac{2}{3}$. In Figure 3, we depict in dotted lines the ex ante payoff of the designer at the IO* mechanism.

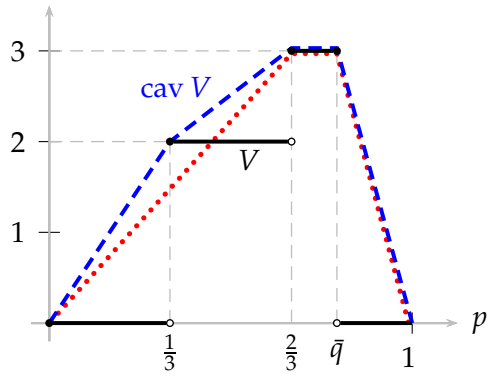

Figure 3: Designer ex ante payoff at the EAO mechanism (dashed lines) and the IO* mechanism (dotted lines) in the four-action informed prosecutor example.

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.[29] 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 the informed prosecutor example, 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.

Regardless of the properties of V, a necessary condition for U to be IO is that each type gets at least their EPO payoff vector. This necessary condition is obtained 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 payoff vectors drastically simplifies.

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

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 their 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 payoff 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 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. In our setting though, the unraveling outcome is not unique and other perfect Bayesian outcomes exist.

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:

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

[29] 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.

Hence, IO* solves the following simplified version of Program (P):[30]

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

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 the informed prosecutor example. In general, the EAO payoff vector is not IO. When the number of actions increases, the EAO payoff vector tends toward no disclosure, whereas every IO payoff vector 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 IO*.

Example 2 (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.[31] 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. The higher the politician's belief that the state is good, the higher the action they choose. 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 clear 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 $p < \frac{K-1}{K}$, IO* 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* mechanism of Program (P-QC) and converges to full disclosure (see right panel of Figure 4).

Example 3 (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 gov-

[30] See Doval and Skreta (2023) for a solution approach.

[31] 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. It can also be interpreted as a generalization of the informed prosecutor example in which the prosecutor has uncertainty about the conviction threshold of the judge.

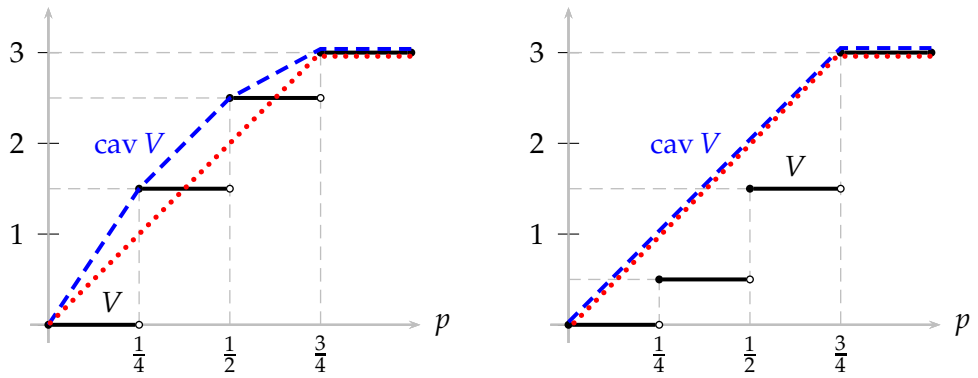

Figure 4: Lobbying: Designer ex ante payoff at the EAO mechanism (dashed lines) and the IO* mechanism (dotted lines) in Example 2 with $K = 4$. Left panel: concave f. Right panel: convex f.

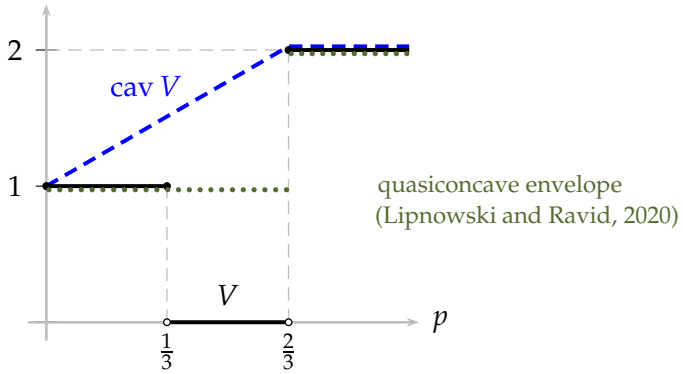

Figure 5: Think tank: Designer ex ante payoff at the EAO mechanism and the IO* mechanism (dashed lines) and the quasiconcave envelope (dotted lines)

ernment 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 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 IO* coincides with the EAO mechanism.[32] 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 payoff 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 might apply to the lobbying example. In that example, the unique equilibrium mechanism under cheap talk is the non-revealing mechanism. Hence, when f is concave, there are intervals of priors such that the cheap-talk solution is ex ante better for the designer than every IO mechanism. For instance, in the left panel of Figure 4, the cheap-talk solution is ex ante better than every IO mechanism for all priors for which the dotted lines are above the plain bold

[32] 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* mechanism 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 2).

lines. When K tends to infinity, the cheap-talk solution is ex ante better than every IO mechanism for almost all priors.

6 Interim optimality in multi-agent binary-action 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 example in Kamenica and Gentzkow (2011). 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 $a \in A$, $u_0(1, \ldots, 1, t) \geq u_0(a, t)$.

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

(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 payoff 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 $a = (1, \ldots, 1)$ is a Nash equilibrium of that game. Finally,

condition (iib) indicates that action 0 is strictly dominant when the state is outside T^* and commonly known. This last part implies $a = (0, \ldots, 0)$ 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, their first-best. The complement of T^* is the set of states in which the designer always gets their worst outcome under complete information.

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.*

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 ex ante commitment assumption is without loss of generality. This positive result is strong given the prevalence of settings that satisfy Assumption 1.[33]

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. Use p to denote the prior probability of type $t = 1$. As in Taneva (2019), the payoffs 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: the designer's payoff 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 their peers to vote for a motion, is a special case obtained for $p = \frac{3}{10}$, $c = 2 \geq d = 1$, and $\overline{V}_0 = \overline{V}_1 = 1$, and hence, it

[33] However, binary-action settings exist in which ex ante optimality does not imply interim optimality (recall Example 1).

satisfies Assumption 1. It then follows from Proposition 4 that the EAO mechanism characterized in Taneva (2019) is IO.[34]

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 the EAO mechanism below and illustrate that it may not be IO 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 the following:

$$\begin{aligned}(1-p)(\gamma_0 c + \beta_0 d) &\geq p(\gamma_1 d + \beta_1 c),\\ p(\beta_1 d + \alpha_1 c) &\geq (1-p)(\beta_0 c + \alpha_0 d).\end{aligned} \tag{2}$$

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 payoff of the designer under these constraints. By Observation 1, the IO* mechanism solves

$$\max_{\mu} p\alpha_1 \overline{V}_1 + (1-p)\alpha_0 \overline{V}_0 \text{ subject to (2) and } \alpha_1 = 1.$$

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 EAO mechanism $\alpha_0 = \frac{pc}{(1-p)d} = \frac{6}{7}$ and $\alpha_1 = 1$, which coincides with IO*. 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

[34] 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 $p = \frac{1}{2}$.

The resulting ex ante expected payoff of the designer is $\frac{9}{10}$, but the mechanism is not IO because $\alpha_1 = 0 \neq 1$. From Observation 2, the 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 payoff 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 payoff of at least $\overline{V}_1$, which only arises when both agents invest.

7 Concluding remarks

In this paper, we identified a class of disclosure mechanisms, which we termed *interim-optimal mechanisms*. These mechanisms are optimal in the sense that the informed designer cannot credibly find an alternative mechanism that strictly improves their 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 PBE 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.

By definition, interim optimal disclosure mechanisms identify mechanisms robust to information unraveling (when news is good); to obfuscation (when news is bad); and to any other disclosure mechanism (paired with credible beliefs). In mechanism and in information design we seek institutions that perform well in expectation (from an ex ante perspective) and that are equilibrium feasible for agents. This latter requirement is captured by the incentive compatibility constraints. When designers are informed, or more broadly, when entities employing the mechanism have private information, they may benefit from switching to another mechanism. Interim optimal mechanisms are robust to such re-optimization. As we have illustrated, our results enable one to check whether an EAO mechanism—the commitment solution in Kamenica and Gentzkow (2011)—is interim optimal and thus robust in this strong sense. Given how crucial credible communication and information disclosure are for settings ranging from monetary policy to public health announcements and to financial, conflict of interest or product characteristic disclosures, the scope and relevance of interim optimal mechanisms are broad and important.

Given the generality of our setting, our 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.

We 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 EAO mechanism is no disclosure. One may then wonder whether an interim-optimal mechanism is always more informative than an EAO one. The answer is no. Recall Example 1 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 EAO mechanism reveals information.

Our setting is general, yet 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 Perez-Richet and Skreta (2022), or the mechanism's output, as in Lipnowski et al. (2022).

Appendix

A Imperfectly informed designer and contracts

We extend the model of Section 3 by allowing the designer to be partially informed about the state of the world 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 their type $t \in T$. The 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 payoff 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 mechanism 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

The interim expected payoff 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 3 and Section 4. 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 if it is EAO or EPO.

For this generalized model, the informed-designer game is the same as in Section 3 except that 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).

We prove Theorem 1 and Theorem 2 for this setting. We prove Proposition 1 last, as its proof relies on concepts defined in the proof of Theorem 1.

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.1 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.2 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 payoff vector of the designer from mechanism μ. Given a Bayesian incentive problem Γ, $B(\Gamma) \subseteq \mathbb{R}^T$ is a set of *blocked payoff vectors*. As mentioned right after Theorem 1 in the main text, we let $B^{IO}(\Gamma)$ be the set of payoff vectors $U \in \mathbb{R}^T$ such that a belief $q \in \Delta(T)$ and a q-IC payoff vector U' exist such that $U'(t) > U(t)$ for every $t \in \text{supp}[q]$. By definition, a payoff vector U is an IO payoff vector iff it is IC and $U \notin B^{IO}(\Gamma)$.

The first axiom requires that if a payoff vector 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 payoff vector 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 payoff vectors 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.3 A mechanism μ is a *neutral optimum* iff μ is incentive compatible and $U_0(\mu) \notin B^*(\Gamma)$.

Lemma A.1 (Myerson, 1983) *For any Bayesian incentive problem Γ, at least one neutral optimum exists.*

The proof Lemma A.1 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 that $B^{IO}(\Gamma)$ 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 payoff vector $U' \in \boldsymbol{U}(q)$ 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 that $B^{IO}(\Gamma)$ satisfies the strong solution axiom (which we state as a proposition below, see Proposition A.1), we start with an auxiliary lemma.

Lemma A.2 *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.2 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.1 *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.2, 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 payoff vector exists because a neutral optimum exists (Lemma A.1). This completes the proof of Theorem 1.

A.2 Proof of Theorem 2

To prove Theorem 2, 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 their 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 a PBE, for every off-path mechanism ν and belief q, agents are required to be sequentially rational in Stage 6 of the informed-designer game. Each agent i chooses a function $\gamma_i : X_i \to \Delta(A_i)$ that determines the probability that i chooses action $a_i \in A_i$ as a function of the signal $x_i \in X_i$. Sequential rationality for the agents requires the strategy profile $(\gamma_i)_{i \in I}$ to constitute a continuation 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 payoff 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 payoff 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.4 $(\gamma_i)_{i\in I}$ is a *continuation 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 3, 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 equilibrium for ν given q always exists. The non-empty and compact set of continuation equilibrium payoff vectors 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 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 payoff vector for ν given q is q-IC:

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

These observations lead to the definition of PBE in Definition 1 in the main text, which is what Myerson (1983) calls an expectational equilibrium. In particular, the set of equilibrium payoff vectors is a subset of the set of IC payoff vectors $\boldsymbol{U}(p)$.

Remark A.1 (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 PBE 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 PBE 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.

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 payoff 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 payoff 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 payoff 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 payoff of player $i \in I$ is

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

By construction, $(\gamma_i)_{i \in I}$ is an equilibrium for $\nu : T \times \Omega \to \Delta(X)$ given q according to Definition A.4, 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 payoff vector; 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 payoff vector 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 a PBE. This completes the proof of Theorem 2.

A.3 Proof of Proposition 1

To prove a payoff vector U^* that is both EPO and EAO is also IO, we use several auxiliary results in Appendix A.1. Because the payoff vector U^* is EAO, it is undominated (Definition A.1). Hence, if it is EPO, it is ex post IC, and therefore, it is a strong solution (Definition A.2). We conclude from Proposition A.1 that U^* is IO. To show uniqueness of the IO payoff vector, let U^{IO} be an IO payoff vector, 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. ∎

B Proofs of Section 5

To prove the results of Section 5 we start with a preliminary lemma.

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

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

$$\bar{a} \in \arg \max_{a \in \bigcup_{t \in T} \text{supp}[\mu(\cdot|t)]} u_0(a). \tag{3}$$

Let $\tilde{q} \in \Delta(T)$ be the posterior belief of the agent when they get 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 $\text{supp}[\tilde{q}] \subseteq \text{supp}[q]$. Because μ is q-IC, we have $\bar{a} \in A^*(\tilde{q})$. Hence,

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

Then, together with Equation (3), this implies

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

for every $t \in \text{supp}[q]$. ∎

B.1 Proof of Proposition 2

($\Rightarrow$) Let $U \in \mathbb{R}^T$ be an IC payoff vector and assume a $q \in \Delta(T)$ exists such that $V(q) > U(t)$ for every $t \in \text{supp}[q]$. Then, the payoff vector 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 payoff vector and assume it is not IO; that is, $\tilde{q}$ and a $\tilde{q}$-IC payoff vector 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 B.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 $t \in \text{supp}[q]$. ∎

B.2 Proof of Proposition 3

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 payoff vector 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 B.1 implies $U'(t) < y$ for all $t \in \text{supp}[q]$, a contradiction. ∎

B.3 Proof of Proposition 4

To prove Proposition 4 we use the following lemma, which shows that under Assumption 1, at an EAO payoff vector, the designer gets their first-best payoff for every $t \in T^*$.

Lemma B.2 *Consider a Bayesian incentive problem with binary actions satisfying Assumption 1.*[35] *If U^* is an EAO payoff vector, 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 payoff 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 that 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). ∎

We now prove Proposition 4:

Let U^* be an EAO payoff vector. By Lemma B.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. ∎

C 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. We do so for the baseline model of a perfectly informed designer presented in the main text.

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

C.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 the following:

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

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 C.1 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 payoff vectors are core payoff vectors, we rely on an alternative, simpler definition of core mechanisms in Lemma C.1 below. To show this equivalence, we use the fact that an ex post IC mechanism always exists, because in information-design settings no truth-telling conditions exist for the designer.

Lemma C.1 *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 obvious 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 straightforward to show $\tilde{\nu}$ is IC given S for every $S \supseteq S(\tilde{\nu}, \mu)$. ∎

A core mechanism has a natural interpretation in terms of deviations of "coalitions" of designer types. Such deviations could also be interpreted as deviations of a partially informed designer. An IC mechanism μ is not a core mechanism iff a set 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. 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 arbitrarily modify 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 constitute the driving force of the following result:

Proposition C.1 *If μ is an interim-optimal mechanism, then μ is a core mechanism.*

Proof. Follows directly from the alternative definition of core in Lemma C.1 and the definition of IO mechanisms (Definition 4). ∎

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

C.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 C.2 (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, the informed prosecutor example has no SUPO mechanism for $p < \frac{1}{3}$.

Mylovanov and Tröger (2012) also introduced a similar concept, called a strong neologism-proof mechanism, which exists in more general *private* value adverse-selection environments and is also a PBE 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 payoff for type t of the designer, that is, the highest possible payoff of the designer when their type is t.

Definition C.3 (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 payoff vectors in the definitions of SUPO and SNP

is not necessarily an open set. By contrast, the set of blocked payoff vectors in the definition of an IO payoff vector is an open set.[36]

Example 4 (SUPO and SNP payoff vectors may not exist) Consider the following example with a single agent, $T = \{1, 0\}$ and $A = A_1 = \{a^1, a^2, a^3\}$:

	a^1	a^2	a^3
$t = 1$	$0, 0$	$1, 1$	$2, -1$
$t = 0$	$0, 1$	$1, 0$	$0, 1$

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

Proposition C.2 *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 PBE mechanisms, and IO mechanisms are core mechanisms. We have also observed that the sets of SUPO and SNP mechanisms may be empty, that some core mechanisms may not be equilibrium mechanisms, and that some PBE 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.

D Imperfectly informed designer: illustration

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 the informed prosecutor example with three actions $A = \{\underline{a}^0, a^2, a^3\}$. Let $\Omega = \{1, 0\}$ and assume the payoff 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, and $\omega = 1$ corresponds to the state in which the defendant is guilty. That is, if $\bar{\pi}$ is the belief of the agent about $\omega = 1$, their 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 $\frac{1}{6}$. The designer's payoff 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

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

[36] See Appendix A.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 their belief about state $\omega = 1$. When the agent has belief q about the designer's type $t = \bar{t}$, their belief about $\omega = 1$ is $q\bar{t}$ and their prior belief about $\omega = 1$ is $p\bar{t} = \frac{1}{6}$.

Because payoffs do not directly depend on t, the ex ante expected payoff 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 payoffs 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 their EPO payoff $U_0^{EPO}(\bar{t}) = \operatorname{cav} V(\bar{t})$. Hence, to characterize when an EAO mechanism is IO it suffices to focus on the EAO mechanism μ that maximizes the payoff 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.

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.

BROOKS, B., A. FRANKEL, AND E. KAMENICA (2022): "Information hierarchies," *Econometrica*, 90, 2187–2214.

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*, 72, 869–908.

DOVAL, L. AND V. SKRETA (2023): "Constrained information design," *Mathematics of Operation Research*, forthcoming.

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

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

——— (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.

GENTZKOW, M. AND E. KAMENICA (2017): "Bayesian persuasion with multiple senders and rich signal spaces," *Games and Economic Behavior*, 104, 411–429.

GREEN, J. AND N. STOKEY (1978): *Two representations of information structures and their comparisons*, 271, Institute for Mathematical Studies in the Social Sciences.

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*, 88, 1631–1660.

LIPNOWSKI, E., D. RAVID, AND D. SHISHKIN (2022): "Persuasion via weak institutions," *Journal of Political Economy*, 130, 2705–2730.

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

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

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

MEKONNEN, T. (2021): "Informed principal, moral hazard, and limited liability," *Economic Theory Bulletin*, 9, 119–142.

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*, 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.

PEREZ-RICHET, E. AND V. SKRETA (2022): "Test design under falsification," *Econometrica*, 90, 1109–1142.

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*.