HAL
open science

Informed Information Design

Frédéric Koessler, Vasiliki Skreta

▶ **To cite this version:**

Frédéric Koessler, Vasiliki Skreta. Informed Information Design. 2021. halshs-03107866v1

HAL Id: halshs-03107866

https://shs.hal.science/halshs-03107866v1

Preprint submitted on 12 Jan 2021 (v1), last revised 22 Dec 2022 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

WORKING PAPER N° 2021 – 03

Information Design by an Informed Designer

Frédéric Koessler
Vasiliki Skreta

JEL Codes: C72; D82
Keywords: Interim information design, Bayesian persuasion, Informed principal, Neutral optimum, Strong-neologism proofness, Core mechanism, Verifiable types.

Funded by a French government subsidy managed by the ANR under the framework of the Investissements d'avenir programme reference ANR-17-EURE-001

Information Design by an Informed Designer*

Frédéric Koessler † Vasiliki Skreta ‡

January 12, 2021

Abstract

A designer is privately informed about the state and chooses an information disclosure mechanism to influence the decisions of multiple agents playing a game. We define an intuitive class of incentive compatible information disclosure mechanisms which we coin *interim optimal mechanisms*. We prove that an interim optimal mechanism exists, and that it is an equilibrium outcome of the interim information design game. An ex-ante optimal mechanism may not be interim optimal, but it is whenever it is ex-post optimal. In addition, in leading settings in which action sets are binary, every ex-ante optimal mechanism is interim optimal. We relate interim optimal mechanisms to other solutions of informed principal problems.

Keywords: interim information design, Bayesian persuasion, informed principal, neutral optimum, strong-neologism proofness, core mechanism, verifiable types.

JEL Classification: C72; D82.

*We thank seminar participants at Brown, Paris School of Economics and UT Austin for useful comments. Frederic Koessler acknowledges the support of the ANR through the program Investissements d'Avenir (ANR-17-EURE-001) and under grant ANR StratCom (ANR-19-CE26-0010-01). Vasiliki Skreta acknowledges funding by the European Research Council (ERC) consolidator grant 682417 "Frontiers In Design." Alkis Georgiadis-Harris provided excellent research assistance.

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

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

1 Introduction and illustrative examples

Decisions ranging from voting, career, investment, to whether or not to get vaccinated depend crucially on the information agents have. In the large and influential literature on Bayesian persuasion and information design, an uninformed designer chooses and commits to a disclosure rule.[1] The purpose of the designer is to achieve a certain goal: A seller tries to persuade buyers that their product is good; a politician voters to vote for them and a pharmaceutical company to convince a doctor to prescribe their medicine. Oftentimes, however, parties selecting the informativeness of a procedure (details on a product brochure; scope and breadth of an investment opportunities study; dimensions on which to test a new vehicle) have private information which shapes their preferences about which procedure to choose, and this, in turn, affects inferences and the ultimate nature of information that can be disclosed in equilibrium. There is a sizable body of research that studies disclosure of evidence by privately informed parties, restricting attention to deterministic evidence.[2] In this paper we take the same interim perspective as the works on disclosure games, but enlarge the choice set that the informed party can choose from: The designer can choose any mapping from the state to a distribution of signals.

We study equilibrium information disclosure mechanisms of an *interim* information design game and investigate how they compare to those arising when the designer can choose and commit to the disclosure mechanism *ex-ante*, before observing state. There are two key differences between the standard information design setting and ours. First, the designer's *interim* incentives differ from his *ex-ante* ones. For example, a high quality seller prefers information to be disclosed, but a low quality one does not. Second, the *choice* of the information disclosure policy can reveal information to the agents. For example, customers can update their expected valuation for a product if the seller designs product testing procedures that have a low probability of uncovering bad characteristics, or if some product features are not tested at all. Interim information design is, thus, not a constrained optimization problem, but a signaling game that shares features

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

[2] See Milgrom (1981), Okuno-Fujiwara, Postlewaite, and Suzumura (1990), Seidmann and Winter (1997), Sher (2011), Hagenbach, Koessler, and Perez-Richet (2014), Hart, Kremer, and Perry (2017) and Ben-Porath, Dekel, and Lipman (2019).

with disclosure games (as in Milgrom, 1981) and informed principal problems *à* la Myerson (1983).

We consider the following setting. There are $n+1$ players: the designer and n agents. Players' payoffs depend on the state of the world $t \in T$ and on the profile of actions chosen. The designer observes the state of the world (which is distributed according to a common prior) and can design any (generalized) information disclosure mechanism: A mapping from T to distributions over signals $\Delta(X)$ where $x \in X$ is a profile of signals and each agent i observes component x_i. The information disclosure mechanism coincides with a Blackwell experiment when there is one agent but it is "generalized" because of two features. First, the output is not necessarily public so each agent can observe a different message. Second, the designer could have actions that are contractually enforceable. After information is disclosed, agents interact in a game, whose continuation equilibrium outcomes depend on the information released by the designer. The set of states of the world and the set of actions for each player can be arbitrary finite sets and we impose no assumption on players' payoff functions.

Our main results are to identify a set of mechanisms, which we call *interim optimal mechanisms*, that always exist (Theorem 1) and that constitute (perfect Bayesian) equilibria of the interim information design game (Theorem 2). Propositions 1 and 2 describe conditions under which an ex-ante optimal mechanism is also interim optimal and thus an equilibrium of the interim information design game. An interim optimal mechanism is an incentive compatible mechanism that is immune to alternative mechanisms when we impose "credibility" constraints to agents' beliefs in the spirit of the notions of core in Myerson (1983) and neologism-proofness in Farrell (1993).

To get a taste of the forces at play when information design is carried at the interim stage as opposed to ex-ante and how these forces shape the identifying properties of interim optimal mechanisms we present two examples.

Example 1 (Transparent motives, three actions) Suppose that the designer (say, the government) faces a single agent (a foreign investor) with three possible actions:

$$A = \{\text{not invest}, \text{invest}, \text{invest and manage}\}.$$

There are two possible states: $T = \{\text{good}, \text{bad}\}$ and the prior is $p(\text{good}) = p = \frac{1}{6}$. The designer's and the agent's payoffs are summarized in the following matrix,

where the first number denotes the designer's payoff and the second the agent's.

	not invest	invest	invest and manage
good	$0, 0$	$2, 2$	$3, 3$
bad	$0, 3$	$2, 2$	$3, 0$

The designer's ranking of the actions is state-independent. The investor's optimal action as a function of his belief $q \in [0, 1]$ that the state is good is:[3]

$$a^*(q) = \begin{cases} \text{not invest} & \text{if } q < 1/3 \\ \text{invest} & \text{if } 1/3 \leq q < 2/3 \\ \text{invest and manage} & \text{if } q \geq 2/3. \end{cases}$$

The resulting designer's indirect utility as a function of q is:

$$V(q) = \begin{cases} 0 & \text{if } q < 1/3 \\ 2 & \text{if } 1/3 \leq q < 2/3 \\ 3 & \text{if } q \geq 2/3. \end{cases}$$

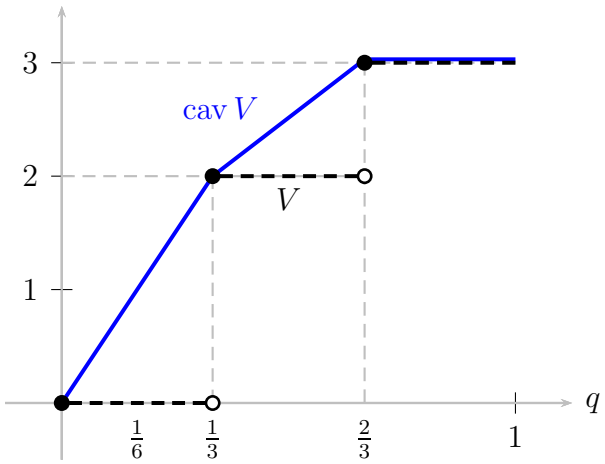
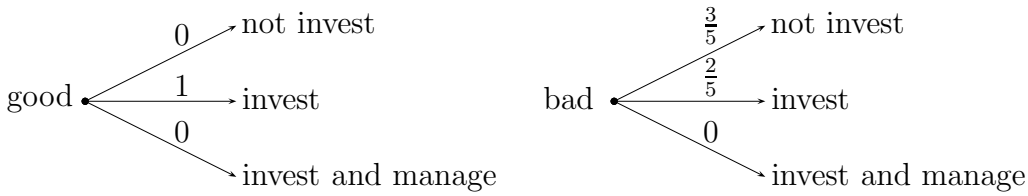


Figure 1: Ex-ante optimal payoff of the designer, $\operatorname{cav} V(q)$, in Example 1.

The ex-ante optimal mechanism for the designer can be obtained directly through the concavification of the designer's indirect utility function V (see Kamenica and Gentzkow, 2011, and Figure 1). In this example, the optimal mechanism is a statistical experiment that splits uniformly the prior $\frac{1}{6}$ into the posteriors 0 and $\frac{1}{3}$. The corresponding direct recommendation mechanism $\mu : T \to \Delta(A)$ is:

[3] In case of indifference, we select the designer's preferred action.



which induces the required beliefs: $\Pr(\text{good} \mid \text{not invest}) = 0$, $\Pr(\text{good} \mid \text{invest}) = \frac{1}{3}$ and results in the ex-ante expected payoff $\operatorname{cav} V(\frac{1}{6}) = \frac{1}{2}V(0) + \frac{1}{2}V(\frac{1}{3}) = 1$. At the ex-ante optimal mechanism, the interim payoff vector for the designer, henceforth *allocation*, is $U = (U(\text{good}), U(\text{bad})) = (2, \frac{4}{5})$: the interim expected payoff of the designer is 2 in the good state and $\frac{2}{5} \times 2 = \frac{4}{5}$ in the bad state. When the state is good, however, the designer can choose a fully revealing experiment thereby inducing action "invest and manage" and get 3 which is strictly higher than the payoff of 2 from the ex-ante optimal mechanism. Hence, in the interim information design game, the interim payoff of the designer should not be less than 3 when the state is good. The allocation resulting from a fully revealing experiment is $(3, 0)$. It is ex-post incentive compatible, so it can be achieved independently of the belief of the agent. More generally, the best the designer types can achieve depends on the beliefs agents hold upon observing a deviation by the designer. Interim optimal mechanisms are defined to be robust to deviations under "reasonable beliefs" for the agents. A mechanism is interim optimal if it is incentive compatible and there is no alternative mechanism which is incentive compatible for some belief that assigns zero probability to states in which the designer does not benefit from the alternative mechanism.

The next example presents a binary state, binary action setting in which the ex-ante optimal experiment is not an even a Nash equilibrium of the interim information design game.

Example 2 (State-dependent motives, two actions) In this example, like in the first, the designer faces one agent. There are two possible states $T = \{t_1, t_2\}$ and two actions for the agent $A = \{a^1, a^2\}$. We show that when the prior is $p(t_1) = p = 3/4$, there is a profitable deviation from the ex-ante optimal experiment whatever the continuation strategy of the agent. 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_2	$0, 1$	$1, 0$

Let q denote the belief of the agent that the designer's type is t_1. The optimal action for the agent is to choose a^1 if $q \leq 1/2$ and a^2 if $q > 1/2$. The designer's indirect utility as a function of q is:

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

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

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

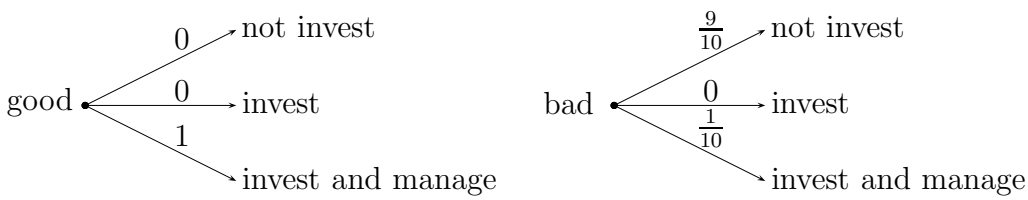
Then, the posterior belief of the agent is $\Pr(t_1 \mid a^2) = 1$, $\Pr(t_1 \mid a^1) = 1/2$ as desired. The ex-ante optimal allocation is $U^{EAO} = (1, 0)$. It is immediate to see that this allocation is not a Nash equilibrium allocation of the interim information design game. The designer can deviate to any pooling experiment (an experiment that sends the same message regardless of the state). Suppose that given such an experiment the agent chooses a^1 with probability β and a^2 with probability $1 - \beta$. If $\beta > \frac{1}{3}$, then t_1 strictly benefits, and if $\beta < 1$, then t_2 strictly benefits, implying that at least one of the two designer types benefits regardless of the value of β.

We formulate the interim information design game as an informed principal problem. The setting is a common value one in Maskin and Tirole (1992)'s terminology because the state of the world can affect all players' payoffs. There are two differences from the usual formulations of informed principal problems. First, in contrast to the setting in Maskin and Tirole (1992), and, for that matter, the majority of work on informed principal problems,[4] the principal cannot "lie" about the state–the input in the experiment is the true state of the world–it is verifiable. In other words, the mediator implementing the mechanism is omniscient in the language of Forges (1993). Second, the experiment's outputs could be "just

[4]To the best of our knowledge there are two exceptions: Types are verifiable in De Clippel and Minelli (2004). Koessler and Skreta (2019) allow for different evidence structures, including verifiable types in a buyer-seller setting with transfers.

signals" and there are not necessarily any contractually enforceable outcomes (as is the norm in the informed principal literature).[5] Our formulation of the interim information design game follows Myerson (1983) with the key difference that the designer's information is verifiable in our setting.

We say that a mechanism is *interim optimal* if there is no "coalition" of designer types that can benefit from selecting an alternative "blocking" mechanism that results to an incentive-compatible allocation (a vector of interim payoffs for the designer) given a belief for the agents that assigns strictly positive probability *only* to designer types that *strictly* benefit from this deviation. In Example 1, an interim optimal allocation is any incentive compatible allocation for the prior such that $U(\text{good}) = 3$. Indeed, the coalition consisting of the good type alone can block any allocation resulting to a payoff strictly lower than 3 for the good type by simply selecting the full disclosure mechanism. In this example, the set of interim optimal allocations is the set of interim payoff vectors $U \in \mathbb{R}^T$ for the designer such that $U(\text{good}) = 3$ and $U(\text{bad}) \in [0, \frac{3}{10}]$. The highest ex-ante expected designer's payoff achievable at an interim optimal mechanism is obtained with the following direct recommendation mechanism



which splits the prior into the posterior $\frac{2}{3}$ with probability $\frac{1}{4}$ and into 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 ex-ante optimal payoff $\operatorname{cav} V(\frac{1}{6}) = 1$.

In Example 2 the ex-ante optimal allocation $U^{EAO} = (1, 0)$ is not interim optimal because, as we have observed, it is not an equilibrium allocation. Another way to understand why U^{EAO} is not interim optimal is to observe that it is blocked by the following mechanism, for $\varepsilon > 0$ small enough:

$$\nu(a^1 \mid t_1) = 1/2; \nu(a^2 \mid t_1) = 1/2; \nu(a^1 \mid t_2) = 1 - \varepsilon; \nu(a^2 \mid t_2) = \varepsilon.$$

Mechanism ν yields allocation $U^\nu = (1.5, \varepsilon)$, which is strictly higher than the ex-

[5] This makes our game closer to an informed principal moral hazard setting. See, for example, Wagner, Mylovanov, and Tröger (2015) and Mekonnen (2018).

ante optimal one, $U^{EAO} = (1, 0)$, for both t_1 and t_2. The mechanism ν is incentive compatible (satisfies obedience constraints) for every belief $q \in (\frac{2\varepsilon}{1+2\varepsilon}, \frac{2(1-\varepsilon)}{3-2\varepsilon})$. So, in particular, mechanism ν with $\varepsilon = \frac{1}{4}$ and belief $q = 1/2$ blocks the ex-ante optimal mechanism. Any belief, and in particular belief $q = \frac{1}{2}$, is credible because both types of the designer benefit from the deviation. In this example, the allocation $U = (0, 1)$, which is simply obtained by a non-revealing experiment, is an interim optimal allocation and an equilibrium allocation by Theorem 2.

In Section 4 we explore conditions under which the ex-post optimal mechanism (the best full disclosure outcome for the designer) and the ex-ante optimal one are interim optimal. In Proposition 1 we show that if an ex-post optimal mechanism is ex-ante optimal, then it is interim optimal and hence a (perfect Bayesian) equilibrium allocation of the interim information design game.[6] This follows from the facts that an ex-ante optimal mechanism is *undominated*, and an ex-post optimal one is incentive compatible for every belief. Hence, if the ex-post optimal mechanism is ex-ante optimal, then it is a *strong solution* (as defined in Myerson, 1983), which is always interim optimal whenever it exists. Proposition 2 establishes that an ex-ante optimal mechanism is interim optimal and an equilibrium in leading environments in the information design literature in which actions are binary, as in the settings in Alonso and Câmara (2016), Arieli and Babichenko (2019) and Chan, Gupta, Li, and Wang (2019).

Finally, in Section 5, we investigate how interim optimality relates to other leading solutions of informed principal games. Within the context of information design the differences in the sets of core (Myerson, 1983), interim optimal, strong-neologism proof (Mylovanov and Tröger, 2012, 2014; Wagner et al., 2015) and strong unconstrained Pareto optimal mechanisms (Maskin and Tirole, 1990) stem from the beliefs that can accompany alternative mechanism proposals. We illustrate why these notions are less appropriate for the interim information design problem. Strong-neologism proof and strong unconstrained Pareto optimal allocations may fail to exist while core allocations may not be equilibrium allocations.

The rest of the paper is structured as follows. Section 2 describes the setting, defines the mechanism selection game, and formalizes the notion of perfect

[6] Ex-post and ex-ante optimal mechanisms are two key benchmarks that have anchored a large fraction of work on informed principal problems. Among others, Maskin and Tirole (1990) and Mylovanov and Tröger (2014) identify independent-private-values environments with transfers in which ex-ante and ex-post optimal mechanisms coincide and constitute equilibrium mechanisms of the informed principal game.

Bayesian equilibrium. In subsection 2.3 we formally define ex-ante, ex-post optimal, and undominated mechanisms. In Section 3 we define interim optimal mechanisms, prove existence and show that an interim optimal mechanism is an equilibrium outcome of the interim information design game. Sections 4 and 5 proceed as described above.

2 Model

We consider a strategic setting with $n+1$ players. Player 0 is the *information designer* who interacts with n players called *agents*. We index agents by $i = 1, \ldots, n$ and 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. A_0 is the non-empty and finite set of enforceable actions for the designer.[7] Let $A = A_0 \times A_1 \times \cdots A_n$ be the set of action profiles.

The designer is privately informed about the state of the world that affects players' payoffs. Let T be the non-empty and finite set of states. This is the set of types of the designer. The common prior $p \in \Delta(T)$ is assumed to have full support. For every action profile $a \in A$ and type $t \in T$, the utility of the designer is $u_0(a, t)$ and the utility of agent $i \in I$ 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

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

When A_0 is a singleton: $A_0 = \{a_0\}$, Γ corresponds to the basic game as defined in Bergemann and Morris (2019).

2.1 Interim information design game

The interim information design game is the following extensive-form game between the designer and the agents:

1. Nature selects the state of the world, $t \in T$, according to the prior probability distribution $p \in \Delta(T)$;

2. The designer is privately informed about $t \in T$;

[7] In most examples we consider pure information design settings which correspond to cases where A_0 is a singleton: $A_0 = \{a_0\}$.

3. The designer chooses a non-empty and finite set of messages X and an information disclosure *mechanism*

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

where $X = A_0 \times X_1 \times \cdots \times X_n$ and $\nu(a_0, x_1, \ldots, x_n \mid t)$ is the probability of implementing the enforceable action a_0 and sending message x_i privately to each agent i when the actual type of the designer is t;

4. Agents publicly observe the mechanism ν proposed by the designer;

5. Each agent i chooses a function $\gamma_i : X_i \to \Delta(A_i)$ that determines the probability that he chooses action $a_i \in A_i$ as a function of their signal $x_i \in X_i$;

6. The enforceable action a_0 and the profile of messages $(x_1, \ldots, x_n)$ are selected with probability $\nu(a_0, x_1, \ldots, x_n \mid t)$, and for every $i \in I$ action $x_i \in X_i$ is played with probability $\gamma_i(a_i \mid x_i)$.

We are interested in perfect Bayesian equilibria of this game.

Key comparisons The key difference between this game and the usual formulation of information design (as in Bergemann and Morris, 2019) or Bayesian persuasion (Kamenica and Gentzkow, 2011) is that in those settings the designer is not informed about t (i.e., stage 2 in the description above is absent). That is, he designs an information structure ex-ante. This 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). In our game the choice of the mechanism is at the interim stage, so we study *an informed principal problem with verifiable types*. Since the revelation principle cannot be applied off the equilibrium path, we allow the designer to choose mechanisms in stage 3 with arbitrary signals as outputs, not just direct recommendation mechanisms. When there is a single enforceable action ($|A_0| = 1$), a mechanism is an information structure for the n agents. In addition, if there is only one agent, a mechanism is a Blackwell statistical experiment. Interim information design games are also related to games studied in the literature on strategic information disclosure. As we mentioned in the introduction, in those papers the informed party chooses which piece of evi-

dence to disclose while in our setting the informed party can choose any stochastic information disclosure mechanism.

2.2 Equilibrium 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 much richer as the designer chooses generalized experiments. Our formalization of equilibrium relies a number of auxiliary results due to 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 $\nu_t : T \to \Delta(X)$ when his type is $t \in T$, there is an outcome-equivalent equilibrium is which all designer types offer the same direct mechanism $\mu : T \to \Delta(A)$ (so agents' beliefs at the beginning of Stage 5 are the same as the prior) and agents are obedient *along* the equilibrium path.[8]

A *direct* mechanism is a mapping $\mu : T \to \Delta(A)$, where $\mu(a_0, a_1, \ldots, a_n \mid t)$ is interpreted as the probability that the mediator "running" the mechanism chooses the enforceable action a_0 and privately recommends a_i to each agent i when the actual type of the designer is t.

Incentive compatibility notions. The mechanism μ is *incentive compatible* (IC) iff for each agent i obedience (following the recommendation a_i) is optimal if all the other agents are obedient:[9]

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

More generally, for a common belief $q \in \Delta(T)$ for the agents, the mechanism

[8] Truth-telling constraints are not needed in our setting given that agents have no private information and the designer's information is verifiable.

[9] When $|A_0| = 1$, an incentive compatible mechanism in our model corresponds to a Bayesian solution in Forges (1993, 2006) and to a Bayes-correlated equilibrium in Bergemann and Morris (2016, 2019).

μ is *q-incentive compatible* (q-IC) iff for each agent i we have:

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

Allocations. Let $U_0(\mu \mid t) = \sum_{a\in A} \mu(a \mid t)u_0(a,t)$ denote the interim expected utility of the designer at state t from mechanism μ when agents are obedient. We call *allocation* the corresponding vector of payoffs for each designer type $U = (U_0(\mu \mid t))_{t\in T}$. Let $\boldsymbol{U}(q) \subseteq \mathbb{R}^T$ be the set of q-IC allocations for the designer:

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

Equilibrium. In a perfect Bayesian equilibrium, for every off-path mechanism ν and belief q, agents are required to be sequentially rational in Stage 5, i.e., they play a strategy profile $(\gamma_i)_{i\in I}$ that constitutes a Nash equilibrium given q and ν. For every $x \in X$ and $a \in A$, let

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

be the probability that the action profile a is played when agents play the strategy profile $(\gamma_i)_{i\in I}$ and the outcome of the mechanism is x (which includes the enforceable action x_0).

Let $W_0(\nu, \gamma \mid t)$ be the interim expected payoff of the designer given t, the mechanism ν, and the agents' strategy profile γ:

$$W_0(\nu, \gamma \mid t) = \sum_{x\in X} \sum_{a\in A} \nu(x \mid t)\gamma(a \mid x)u_0(a,t).$$

Let $W_i(\nu, \gamma \mid q)$ be the expected payoff of agent i given belief $q \in \Delta(T)$, the mechanism ν, and the strategy profile γ of the agents:

$$W_i(\nu, \gamma \mid q) = \sum_{t\in T} \sum_{x\in X} \sum_{a\in A} q(t)\nu(x \mid t)\gamma(a \mid x)u_i(a,t).$$

Definition 1 $(\gamma)_{i\in I}$ is a *continuation Nash equilibrium for $\nu : T \to \Delta(X)$ given q* iff for every $i \in I$ and $\gamma_i' : X_i \to \Delta(A_i)$ we have

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

Because agents have symmetric information at the beginning of Stage 5, we require that they have a common belief q at this stage, even off the equilibrium path and this also why we formulated the notions of incentive compatibility for a common belief.

Note that the game induced by ν with prior q has finite sets of pure strategies, so a continuation Nash equilibrium for ν given q always exists. The non-empty and compact set of continuation equilibrium allocations for ν given q is denoted by

$$\mathcal{U}(\nu, q) = \{U \in \mathbb{R}^T : \text{there exists a NE } \gamma \text{ for } \nu \text{ given } q \text{ such that } (W_0(\nu, \gamma \mid t))_{t \in T} = U\}.$$

By the revelation principle, every continuation equilibrium utility allocation for ν given q is q-IC:

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

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

Definition 2 (Equilibrium of the interim information design game) A mechanism $\mu : T \to \Delta(A)$ is an *equilibrium* of the interim information design game iff

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

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

Remark 1 (Definition of equilibrium) Requiring that agents have a common belief at the beginning of Stage 5 is in the spirit of the belief consistency requirement of the sequential equilibrium of Kreps and Wilson (1982) and the strong version of perfect Bayesian equilibrium in Fudenberg and Tirole (1991), and it is standard in the literature. Our results hold under any weaker version of perfect Bayesian equilibrium. We follow Myerson (1983) because there are two important difficulties defining sequential equilibrium or a strong version of perfect Bayesian equilibrium directly in our setting. First, the interim information design 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. While nature moves at the start of the interim information design game to determine the designer's type $t \in T$ 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–so x can have probability zero for some states that are on the support of the agents' beliefs at the beginning of Stage 4.

2.3 Key benchmarks

For the purpose of comparing and deriving properties of equilibrium mechanisms of the interim information design game, we now formally define the concepts of ex-ante optimal, ex-post optimal, undominated mechanisms, and of strong solutions. The latter concept, defined in Myerson (1983), helps us connect ex-ante, ex-post and interim optimal mechanisms in Proposition 1.

Definition 3 (Ex-ante optimal mechanisms) A mechanism μ is *ex-ante optimal* iff μ is incentive compatible and for any other incentive compatible mechanism ν we have

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

When there is a single enforceable action, an ex-ante optimal mechanism corresponds to a solution of the standard information design problem (Kamenica and Gentzkow, 2011; Bergemann and Morris, 2019; Taneva, 2019). An ex-ante optimal mechanism is undominated, in the following sense:

Definition 4 (Dominated and undominated mechanisms) 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.

A mechanism μ is *ex-post incentive compatible*[10] iff for every i and 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 incentive compatible mechanism satisfies the agents' obedience constraints when they know the state and it is q-IC for *every* $q \in \Delta(T)$. If there is a single enforceable action ($A_0 = \{a_0\}$), then it maps every t to a correlated equilibrium (Aumann, 1974) of the normal form game $((A_i)_{i \in I}, (u_i(a_0, \cdot, t))_{i \in I})$. An ex-post incentive compatible mechanism always exists in our environment because the set of correlated equilibria is non-empty and the designer's type is verifiable: The designer cannot "lie" to the mediator implementing the mechanism because the mediator is omniscient.[11]

Definition 5 (Ex-post optimal mechanisms) A mechanism μ is *ex-post optimal* iff μ is ex-post incentive compatible and for every other ex-post incentive compatible mechanism ν we have:

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

The ex-post optimal allocation is the best correlated equilibrium allocation for the designer when there is complete information about t. When there are enforceable actions only, the ex-post optimal mechanism corresponds to a best safe mechanism in De Clippel and Minelli (2004). However, the ex-post optimal allocation may be dominated; if it is not, then it is called a strong solution:

Definition 6 (Strong solution (Myerson, 1983)) A mechanism μ is a *strong solution* iff it is ex-post incentive compatible and undominated.

A strong solution is an equilibrium of the informed designer game (see the proof of Theorem 1), and it is a robust prediction of the information design problem because it is incentive compatible for all agent beliefs. However, in many interesting information design problems, a strong solution does not exist. In Example 1, the

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

[11] An ex-post incentive compatible 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), an ex-post incentive compatible mechanism may not exist because a mechanism that is ex-post incentive compatible for the agents may not be incentive compatible for the designer.

ex-post optimal allocation is $(3, 0)$. It is not dominated by the ex-ante optimal allocation $U^{EAO} = (2, \frac{4}{5})$ but it is dominated by the incentive-compatible allocation $(3, U(\text{bad}))$ for every $U(\text{bad}) \in (0, \frac{3}{10}]$, so it is not a strong solution. In Example 2, the ex-post optimal allocation is $(0, 0)$ and as discussed in the introduction, it is not even a Nash equilibrium of our game.[12] It is dominated by the ex-ante optimal allocation $U^{EAO} = (1, 0)$ and by the incentive-compatible allocation $(0, 1)$. In both examples, the ex-post optimal allocation is dominated by an interim optimal allocation as defined in the next section.

3 Interim optimal mechanisms

In this section we define interim optimal mechanisms and establish our main results: Theorem 1 shows that an interim optimal mechanism always exists and Theorem 2 that it is an equilibrium of the interim information design game.

Definition 7 (Interim optimal mechanism) A mechanism $\mu : T \to \Delta(A)$ is *interim optimal* iff μ is incentive-compatible and there is no mechanism ν and belief q such that ν is q-incentive-compatible and $U_0(\nu \mid t) > U_0(\mu \mid t)$ for every $t \in \text{supp}[q]$.

The definition of interim optimality relies on a notion of credibility of beliefs: if the designer selects an alternative mechanism ν, then the agents assign positive probability only to designer types who strictly benefit from the alternative mechanism ν. This credibility requirement is similar but different from other notions of credibility in the informed principal literature (see Section 5). Note that, for every type of the designer, his payoff at an interim optimal mechanism is never lower than at an ex-post optimal mechanism. Indeed, if $U_0(\nu \mid t) > U_0(\mu \mid t)$ for some t and ν is ex-post optimal (and thus ex-post incentive compatible), then ν is q-incentive-compatible for $q = \delta_t$, and therefore μ is not interim optimal. Said differently, at an interim optimal mechanism, each designer type should be better off than under any full disclosure outcome. This basic necessary (but not sufficient) property fails at the ex-ante optimal mechanism of Example 1. More generally, if

[12] This is in contrast to the setting in De Clippel and Minelli (2004) where the ex-post optimal allocation (and any incentive-compatible allocation that dominates it) is an equilibrium allocation. The difference stems from the fact that in that paper the agent simply accepts or rejects the mechanism proposed by the informed principal.

a mechanism is interim optimal, then there is no subset S of designer types such the designer types in that subset strictly benefit from disclosing S to the agents.

Remark 2 (An interim optimal mechanism can be dominated) By definition, an interim optimal mechanism cannot be strictly dominated by another incentive-compatible mechanism. However, it could be weakly dominated, as in Example 1: the interim optimal allocations $(3, U(\text{bad}))$ with $U(\text{bad}) \in [0, \frac{3}{10})$ are weakly dominated by the interim optimal allocation $(3, \frac{3}{10})$. As a refinement of interim optimal allocations, one might select those that are not weakly dominated. However, we see no convincing game theoretic argument for such a selection: In Example 1, if the agent expects the good type of the designer to use a fully informative experiment, which is the best he can do, then it is reasonable that the agent assigns any deviation from full disclosure to the bad type only.

We now establish existence of interim optimal mechanisms for every Bayesian incentive problem $\Gamma = ((A_i, u_i)_{i=0}^I, T, p)$. Note that the proof does not impose any additional assumptions on any of the elements of Γ.

Theorem 1 (Interim optimal mechanisms exist) *An interim optimal mechanism exists for every Bayesian incentive problem* $\Gamma = ((A_i, u_i)_{i=0}^I, T, p)$.

Proof. See the Appendix. ■

The idea of the proof lies in establishing that a neutral optimum (as defined in Myerson, 1983) is interim optimal and neutral optima exist by Theorem 6 in Myerson (1983). To relate interim optimality with neutral optimum we define interim optimality in terms of "blocked allocations" as follows.

Let $B^{IO}(\Gamma)$ be the set of allocations $U \in \mathbb{R}^T$ such that there exists a belief $q \in \Delta(T)$ and a q-IC allocation U' such that $U'(t) > U(t)$ for every $t \in \text{supp}[q]$. By definition, an allocation U is an interim optimal allocation iff it is IC and $U \notin B^{IO}(\Gamma)$. The proof shows that $B^{IO}(\Gamma)$ satisfies the axioms of *Domination*, *Openness*, *Extensions* and *Strong solutions* which establishes that the set of neutral optima is included in the set of interim optimal allocations, and thus the set of interim optimal allocations is non-empty. These axioms also characterize desirable properties of interim optimal allocations. The domination axiom requires that if an allocation U is blocked, then every allocation which is strictly dominated by U is blocked as well. The openness axiom, which is key for existence, requires that if U is blocked, then there exists a neighborhood of U such that every allocation

in that neighborhood is blocked as well. The extension axiom requires that if the designer can commit to additional enforceable actions, then more allocations could be blocked because a larger set of alternative mechanisms are available to the designer. The last axiom requires that if a strong solution (i.e., an undominated ex-post incentive compatible allocation) exists, then it should not be blocked. These axioms are defined in Myerson (1983) and, for completeness, we include their formal definitions in the Appendix.

The next theorem shows that an interim optimal mechanism is an equilibrium mechanism.

Theorem 2 (Interim optimal are equilibrium mechanisms) *If μ is an interim optimal mechanism, then μ is an equilibrium mechanism of the interim information design game.*

Proof. Let μ be an interim optimal mechanism. By definition, it is incentive compatible. 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, $(a_0, x_1, \ldots, x_n) \in X$ is drawn with probability $\nu(a_0, x_1, \ldots, x_n \mid t)$. In the third stage, each player 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 $u_i(a_0, a_1, \ldots, a_n, t)$.

Since the fictitious game $G(\nu, \mu)$ is a finite extensive form game, it has an equilibrium in behavioral strategies. 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

$$\sum_{t \in T} q(t)(W_0(\nu, \gamma \mid t) - U_0(\mu \mid t)),$$

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

$$W_i(\nu, \gamma \mid q).$$

By construction, $(\gamma)_{i \in I}$ is a Nash equilibrium for $\nu : T \to \Delta(X)$ given q according to Definition 1, so by the revelation principle $U = (U(t))_{t \in T} = (W_0(\nu, \gamma \mid t))_{t \in T}$ is a q-IC allocation, i.e., $U \in \boldsymbol{U}(q)$. Let

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

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

4 When is ex-ante information design interim optimal?

In this section we state sufficient conditions under which the ex-ante optimal mechanism is interim optimal, and therefore an equilibrium mechanism of interim information design game. Under those conditions, the ability to ex-ante commit to a mechanism brings no extra value to the designer.[13]

4.1 When full disclosure is ex-ante optimal

The first proposition shows that if the solution of the ex-ante information design problem is full disclosure (i.e., ex-post optimal, Definition 5), then it is a strong solution, and therefore it is interim optimal and an equilibrium mechanism of the interim information design game.

Proposition 1 *If an ex-ante optimal mechanism is ex-post incentive compatible, then it is interim optimal and an equilibrium of the interim information design game.*

Proof. The ex-ante optimal mechanism is undominated. Hence, if it is ex-post incentive compatible, then it is a strong solution. A strong solution is interim

[13]Other papers that relax the commitment assumption of the standard information design paradigm in different ways than us include Lipnowski, Ravid, and Shishkin (2019), Lipnowski and Ravid (2020) and references therein. In Lipnowski et al. (2019) 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 committing to a disclosure rule) by an informed party that has state-independent preferences over action profiles.

optimal (see the proof of Theorem 1) and an equilibrium of the information design game by Theorem 2. ■

Of course, ex-post incentive compatibility is not a necessary condition for the ex-ante optimal mechanism to be interim optimal. The next section describes an important class of Bayesian incentive problems in the information design literature in which the ex-ante optimal mechanism is usually not ex-post incentive compatible but is always interim optimal.

4.2 When actions are binary and motives "transparent"

In this section we consider a pure information design setting, i.e., the set of enforceable actions of the designer is a singleton. Each agent has only two actions: $A_i = \{0, 1\}$ for every $i \in I$. We make the following assumption for every $i \in I$:

Assumption 1 There exists a subset of types $T^* \subseteq T$ 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 $t \in T^*$, $u_i(1, \ldots, 1, t) - u_i(0, 1, \ldots, 1, t) \geq 0$ and $u_i(1, \ldots, 1, t) - u_i(0, 1, \ldots, 1, t) \geq u_i(1, a_{-i}, t) - u_i(0, a_{-i}, t)$ for every $a_{-i} \in A_{-i}$;

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

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

The set T^* is set of states in which under complete information the designer is able to get, at some Nash equilibrium, his first best. The complement of T^* is the set of states in which the designer always gets his worst outcome under complete information.

Assumption 1 is always satisfied if there is a single agent and the designer's utility is state-independent. This includes the leading "judge" example in Kamenica and Gentzkow (2011), and the setting of Perez-Richet (2014). Assumption 1 is also satisfied in many applications with multiple agents in the information design literature: Alonso and Câmara (2016) and Chan et al. (2019), consider voting settings, whereas Arieli and Babichenko (2019) a setting that encompasses technological adoption. Assumption 1 is also satisfied in the coordination games (the investment examples) in Bergemann and Morris (2019) and Taneva (2019).

The next lemma shows that under Assumption 1, the designer gets his first best for every $t \in T^*$.

Lemma 1 *Consider a Bayesian incentive problem with binary actions satisfying Assumption 1.*[14] *If U^* is an ex-ante optimal allocation, then $U^*(t) = u_0(1, \ldots, 1, t)$ for every $t \in T^*$.*

Proof. Let μ be an ex-ante optimal 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 ex-ante optimal. 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 incentive compatible. Incentive compatibility for agent i is equivalent to

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

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 incentive-compatible. The second inequality follows from Condition (iib). ■

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

Proposition 2 (Binary actions) *Consider a Bayesian incentive problem with binary actions satisfying Assumption 1. Then, an ex-ante optimal mechanism is interim optimal, and therefore an equilibrium of the interim information design game.*

Proof. Let U^* be an ex-ante optimal allocation. By Lemma 1, $U^*(t) = u_0(1, \ldots, 1, t)$ for every $t \in T^*$. Assume by way of contradiction that U^* is not interim optimal. Then, there exists a q-IC mechanism ν 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), $\nu(0, \ldots, 0 \mid t) = 1$ 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. ■

5 Interim optimality and other solution concepts

In this section we discuss the relationship between interim optimal allocations and some key concepts in the informed principal literature: core allocations (Myerson, 1983), strong unconstrained Pareto optimal allocations (Maskin and Tirole, 1990) and strong neologism-proof (Mylovanov and Tröger, 2012, 2014). In Proposition 3 we show that interim optimal allocations are core allocations, so core allocations always exist. However, core allocations may fail to be equilibrium allocations of the interim information game, and are therefore not necessarily interim optimal. In example 3 we illustrate that strong unconstrained Pareto optimal and strong neologism-proof allocations may fail to exist in our setting. When it exists, a strong neologism-proof allocation is interim optimal and is therefore an equilibrium allocation.

5.1 Core

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

$$\sum_{a_{-i} \in A_{-i}} \sum_{t \in R} p(t)\mu(a \mid t)[u_i(a,t) - u_i((a_i', a_{-i}), t)] \geq 0, \text{ for every } a_i \text{ and } a_i' \text{ in } A_i. \quad (1)$$

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 has been defined by Myerson (1983) as follows:

Definition 8 (Core mechanism) A mechanism $\mu : T \to \Delta(A)$ is a *core mechanism* iff μ is IC and there is no mechanism ν such that $S(\nu, \mu) \neq \emptyset$ and such that ν is IC given S for every $S \supseteq S(\nu, \mu)$.

To establish that interim optimal allocations are core allocations we rely on an alternative, simpler definition of core mechanisms in Lemma 2 below. To show this equivalence we use the fact that an ex-post incentive compatible mechanism always exists when the state is verifiable and, therefore, there are no truth-telling conditions for the designer.

Lemma 2 (Equivalent definition of core mechanism) *A mechanism $\mu : T \to \Delta(A)$ is a core mechanism iff μ is IC and there is no mechanism ν such that $S(\nu, \mu) \neq \emptyset$ and such that ν is IC given $S(\nu, \mu)$.*

Proof. The "if" part is direct by definition. To show the "only if" part we show that if μ is IC and there is mechanism ν such that $S(\nu, \mu) \neq \emptyset$ and such that ν is IC given $S(\nu, \mu)$, then μ is not a core mechanism, i.e., there exists a mechanism $\tilde{\nu}$ 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

$$\tilde{\nu}(t) = \begin{cases} \nu(t) & \text{if } t \in S(\nu, \mu) \\ \nu^{EPIC}(t) & \text{if } t \notin S(\nu, \mu), \end{cases}$$

where ν^{EPIC} is any ex-post incentive compatible mechanism. It is immediate to show that $\tilde{\nu}$ 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; an IC mechanism μ is not a core mechanism iff there exists a coalition of types $S \subseteq T$ and mechanism ν which is IC given S, such that all types in S strictly benefit from ν compared to μ. Notice that the belief of the agents after the deviation can either be interpreted as coming from a strategic inference that $t \in S$, or as direct inference from a verifiable disclosure of the set S from the deviating coalition. An interim optimal mechanism is similar to a core mechanism but allows for more blocking mechanisms. The definition of interim optimal mechanism does not require the blocking mechanism ν to be incentive compatible given $S(\nu, \mu)$; the blocking mechanism could more generally be incentive compatible for some belief q whose support is included in $S(\nu, \mu)$ (i.e., $\text{supp}[q] \subseteq S(\nu, \mu)$). This allows for more flexibility for off path beliefs: Agents can modify arbitrarily the relative likelihoods of the different types in $S(\nu, \mu)$, whereas in the definition of the core mechanism off path beliefs 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 which is the driving force of the following result:

Proposition 3 (An interim optimal mechanism is a core mechanism) *If μ is an interim optimal mechanism, then μ is a core mechanism.*

Proof. Follows directly from the alternative definition of core in Lemma 2 and the definition of interim optimal mechanisms (Definition 7). ■

The reverse of this proposition is not true. In Example 2, the core allocation $(1, 0)$ (which is ex-ante optimal for the assumed prior) is not interim optimal. This example also shows that a core allocation is not necessarily an equilibrium allocation because, as seen previously, $(1, 0)$ is not an equilibrium allocation.

5.2 Strong unconstrained Pareto optimality and strong neologism-proofness

Maskin and Tirole (1990) introduced the notion of the strong unconstrained Pareto optimal (SUPO) mechanism, which exists and is an equilibrium of some informed principal problems with private values and transfers, as well as in some interdependent value environments with verifiable types (see, e.g., Koessler and Skreta, 2019). For completeness and ease of comparison we include the definition:

Definition 9 (Maskin and Tirole, 1990) A mechanism $\mu : T \to \Delta(A)$ is *strong unconstrained Pareto optimal (SUPO)* iff it is incentive compatible and there is no belief $q \in \Delta(T)$ together with a q-IC mechanism ν 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 $\text{supp}[q] \neq T$.

A solution concept similar to strong unconstrained Pareto optimality is strong neologism-proofness, introduced by Mylovanov and Tröger (2012) who established that a strong neologism-proof mechanism exists in general *private* value adverse selection environments and is an equilibrium mechanism of the informed principal game in such environments. It is also applicable to some moral hazard settings (see Wagner et al., 2015). Let

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

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

Definition 10 (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-IC 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, SUPO and SNP mechanisms do not exist. Failure of existence is related to the fact that the set of blocking allocations in the definition of SUPO and SNP is not necessarily an open set. On the contrary, the set of blocking allocations in the definition of an interim optimal allocation is an open set.

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

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

The first best allocation is $U_0^{FB} = (2, 1)$. If $p < \frac{1}{2}$ every incentive-compatible allocation is dominated by the allocation $(1, 1)$, which is q-IC for $q \geq \frac{1}{2}$, so there is

no SUPO and no SNP allocation. However, every incentive-compatible allocation in which the utility of type t^1 is equal to 1 is interim optimal.

Proposition 4 (SNP mechanism is interim optimal) *If μ is a strong neologism-proof mechanism then μ is an interim optimal mechanism, and therefore an equilibrium of the interim information design game.*

Proof. Let μ be an IC mechanism that is not an interim optimal mechanism, i.e., there exists $q \in \Delta(T)$ and a q-IC mechanism ν 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 a strong neologism-proof mechanism. We conclude by Theorem 2. ■

The next proposition shows that if μ is a strong solution, then it is strong neologism-proof.

Proposition 5 (A strong solution is SNP) *If μ is a strong solution, then μ is a strong neologism-proof mechanism.*

Proof. See the Appendix. ■

To summarize, we have the following relationships:

$$\text{strong solution} \Rightarrow \begin{cases} \text{strong neologism proof} \\ \text{neutral optimum} \end{cases} \Rightarrow \text{interim optimal} \Rightarrow \text{core}$$

We have also established that, in general, the reverse implications are not true. Thus, important classes of allocations that exist and are equilibrium allocations for other informed principal settings fail to exist in information design settings.

6 Other related literature

There is a small set of papers that study information design at the interim stage. A pioneering paper is Perez-Richet (2014) who studies equilibrium refinements and constrained information policies in a single-agent setting with binary actions and states, and state-independent utilities for the designer. Hedlund (2017) studies a binary state setting in which the designer is partially informed, and shows that

equilibrium outcomes that satisfy D1 are either fully disclosing (the experiment fully reveals the state) or fully separating (the choice of experiment reveals the state). In Chen and Zhang (2020) the principal's type is binary and it indexes the distribution of the buyer's values. The seller offers experiments that provide information about the buyer's value. They show that private information hurts the seller. The aforementioned papers examine single-agent settings. Eliaz and Serrano (2014) analyze a setting where an informed planner discloses information about the state to two interacting agents who play multi-action versions of prisoner's dilemma. The planner knows the state and sends each agent a private message which consists of any subset of states that contains the true one.

Clearly, when the designer has access to all experiments, the optimum can be achieved by choosing and committing to the experiment before knowing the state. This is not generally true if the set of experiments is restricted. Degan and Li (2015) examine a binary state, binary action setting in which the informed sender chooses the signal's precision. Alonso and Câmara (2018) focus on whether or not 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 considered a general 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.

Appendix

A Blocked allocations and additional proofs

The following axioms were originally defined in Myerson (1983). For completeness, we include their formal definitions in what follows. Let $U_0(\mu) := (U_0(\mu \mid t))_{t \in T} \in \mathbb{R}^T$ be the utility allocation vector of the designer from mechanism μ. Given a Bayesian incentive problem Γ,

$$B(\Gamma) \subseteq \mathbb{R}^T$$

is a set of *blocked allocations*.

The first axiom requires that if an allocation U is blocked and U' is strictly dominated by U, then 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, then there exists a neighborhood of U such that every allocation in that neighborhood is blocked too.

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

We say that a Bayesian incentive problem $\bar{\Gamma} = ((\bar{A}_0, (A_i)_{i \in N}), T, \bar{u}_0, (\bar{u}_i)_{i \in I}, p)$ is an *extension* of the Bayesian incentive problem $\Gamma = ((A_0, (A_i)_{i \in N}), T, u_0, (u_i)_{i \in I}, p)$ if $A_0 \subseteq \bar{A}_0$ and

$$\bar{u}_i(a, t) = u_i(a, t), \text{ for every } i = 0, 1, \ldots, n, \; t \in T \text{ and } a \in A_0 \times A_1 \times \cdots \times A_n.$$

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

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

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.

Definition 11 A mechanism μ is a *neutral optimum* (NO) iff μ is IC and $U_0(\mu) \notin B^*(\Gamma)$.

If $B^{IO}(\Gamma)$ satisfies all four axioms, then $B^{IO}(\Gamma) \subseteq B^*(\Gamma)$, and therefore a neutral optimum is interim optimal. Hence, an interim optimal allocation exists because a neutral optimum exists (Theorem 6, Myerson, 1983).

We start with a auxiliary lemma which we use below to show that $B^{IO}(\Gamma)$ satisfies the strong solution axiom. We also use it in the proof of Proposition 5.

Lemma 3 (Convex combination of IC mechanisms) *If ν is q-IC and ν' is q'-IC, then for every $\alpha \in [0,1]$, the mechanism ν^*, defined by*

$$\nu^*(a \mid t) = \frac{\alpha q(t)}{q^*(t)}\nu(a \mid t) + \frac{(1-\alpha)q'(t)}{q^*(t)}\nu'(a \mid t), \text{ for every } a \in A \text{ and } t \in \operatorname{supp}[q^*],$$

with $q^*(t) = \alpha q(t) + (1-\alpha)q'(t)$ for every $t \in T$, is q^*-IC.

The intuition of this result 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'. The formal proof is as follows.

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

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

i.e.,

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

or, equivalently,

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

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. ■

Proof of Theorem 1. We prove that a neutral optimum is interim optimal which amounts to verifying that $B^{IO}(\Gamma)$ satisfies the axioms of *Domination*, *Openness*, *Extensions* and *Strong solutions*.

Domination. Let $U \in B^{IO}(\Gamma)$, i.e., there exists $q \in \Delta(T)$ and $U' \in \boldsymbol{U}(q)$ 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', i.e., $\tilde{U} \in B^{IO}(\Gamma)$.

Openness. For every $t \in T$, let $\varepsilon(t) \in \mathbb{R}^*$ 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', i.e., $\tilde{U} \in B^{IO}(\Gamma)$.

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

Strong solutions. Assume by way of contradiction that μ is a strong solution but not interim optimal. Then, there exists $q \in \Delta(T)$ and a q-IC mechanism ν such that

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

Consider the following splitting of p: $p(t) = \alpha q(t) + (1-\alpha)q'(t)$, $\alpha \in (0,1)$, and the mechanism ν^* that implements ν with probability α and μ with probability $(1-\alpha)$. ν^* is p-IC because ν is q-IC and μ is q'-IC (because μ is ex-post IC). In addition, $U_0(\nu^* \mid t) > U_0(\nu \mid t)$ for every $t \in \text{supp}[q]$ because $U_0(\nu^* \mid t)$ is a convex combination of $U_0(\nu \mid t)$ and $U_0(\mu \mid t)$. Moreover ν^* is incentive compatible by Lemma 3. Hence, ν^* dominates μ, a contradiction to the assumption that μ is a strong solution. ■

Proof of Proposition 5. We show that if an IC mechanism μ is not SNP, then it is not a strong solution. Assume by way of contradiction that μ is a strong

solution but not SNP. Then, there exists $q \in \Delta(T)$ and a q-IC mechanism ν such that $q(t) = 0$ if $U_0(\mu \mid t) = U_0^{FB}(t)$ and $U_0(\nu \mid t) \geq U_0(\mu \mid t)$ for every $t \in \text{supp}[q]$, with a strict inequality for some $t \in \text{supp}[q]$.

For every $t \in T$, let

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

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

Define the following mechanism ν^*:

$$\nu^*(a \mid t) := \frac{\alpha q(t)}{p(t)} \nu(a \mid t) + \frac{(1 - \alpha) q'(t)}{p(t)} \mu(a \mid t), \text{ for every } t \in T \text{ and } a \in A.$$

Note that $p(t) = \alpha q(t) + (1 - \alpha) q'(t)$ for every $t \in T$, ν is q-IC and μ is q'-IC because it is ex-post incentive-compatible. Hence, from Lemma 3 the mechanism ν^* is incentive compatible 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 $U_0(\nu^* \mid t) \geq U_0(\nu \mid t)$, with a strict inequality for some $t \in \text{supp}[q]$. We conclude that ν^* dominates μ, a contradiction to the assumption that μ is a strong solution. ■

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.

Ben-Porath, E., E. Dekel, and B. L. Lipman (2019): "Mechanisms with evidence: Commitment and robustness," *Econometrica*, 87, 529–566.

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

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

Chan, J., S. Gupta, F. Li, and Y. Wang (2019): "Pivotal persuasion," *Journal of Economic Theory*, 180, 178 – 202.

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

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 (2015): "Persuasive signalling," *Available at SSRN 1595511*.

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

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

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

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

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

Forges, F. and F. Koessler (2005): "Communication Equilibria with Partially Verifiable Types," *Journal of Mathematical Economics*, 41, 793–811.

Fudenberg, D. and J. Tirole (1991): *Game Theory*, MIT Press.

Hagenbach, J., F. Koessler, and E. Perez-Richet (2014): "Certifiable Pre-Play Communication: Full Disclosure," *Econometrica*, 82, 1093–1131.

Hart, S., I. Kremer, and M. Perry (2017): "Evidence games: Truth and commitment," *American Economic Review*, 107, 690–713.

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

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

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

Koessler, F. and V. Skreta (2019): "Selling with evidence," *Theoretical Economics*, 14, 345–371.

Kreps, D. M. and R. Wilson (1982): "Sequential Equilibria," *Econometrica*, 50, 863–894.

Lipnowski, E. and D. Ravid (2020): "Cheap talk with transparent motives," *Econometrica*, forthcoming.

Lipnowski, E., D. Ravid, and D. Shishkin (2019): "Persuasion via weak institutions," *Available at SSRN 3168103*.

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

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

Mathevet, L., J. Perego, and I. Taneva (2020): "On information design in games," *Journal of Political Economy*, 128, 1370–1404.

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

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

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

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

Mylovanov, T. and T. Tröger (2012): "Informed principal problems in generalized private values environments," *Theoretical Economics*, 7, 465–488.

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

Okuno-Fujiwara, A., M. Postlewaite, and K. Suzumura (1990): "Strategic Information Revelation," *Review of Economic Studies*, 57, 25–47.

Perez-Richet, E. (2014): "Interim Bayesian Persuasion: First Steps," *The American Economic Review*, 104, 469–474.

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.