# Informationally Simple Incentives*

Agathe PERNOUD[†] Simon GLEYZE[‡]

**Abstract**

We consider a mechanism design setting in which agents can acquire costly information on their preferences as well as others'. A mechanism is *informationally simple* if agents have no incentive to learn about others' preferences. This property is of interest for two reasons: First, it is a necessary condition for the existence of dominant strategy equilibria in the extended game. Second, this endogenizes an "independent private value" property of the interim information structure. We show that, generically, a mechanism is informationally simple if and only if it satisfies a separability condition which rules out most economically meaningful mechanisms.

*We thank Emir Kamenica and three anonymous referees for suggestions that greatly improved the paper. We are grateful to Piotr Dworczak, Fuhito Kojima, Shengwu Li, Elliot Lipnowski, Mike Ostrovsky, Eduardo Perez-Richet, Doron Ravid, Al Roth, Ilya Segal, Olivier Tercieux and especially Matt Jackson, Philippe Jehiel and Paul Milgrom for helpful conversations and comments. We also thank seminar participants at PSE, Stanford, Akbarpour–Milgrom discussion group, YES2020, and the CEME Decentralization Conference (2021) for valuable comments and questions. S. Gleyze acknowledges the support of the EUR grant ANR-17-EURE-0001. Edited by Emir Kamenica.

[†]Department of Economics, Stanford University. Email: agathep@stanford.edu

[‡]Uber. Email: gleyze.simon@gmail.com

# 1 INTRODUCTION

Scholars have long understood that institutions can have a strong impact on the formation of preferences. Surprisingly little attention, however, has been paid to how institutions shape the ways in which we constitute our knowledge and acquire information, that is, to how agents' *informational incentives* are affected by institutional rules. This can sometimes be of primary importance because heterogeneous informational incentives can lead to unequal opportunities in voting, labor market outcomes, education choices, investment decisions, etc. Conversely, little is known about how these informational incentives constrain the type of institutions that are actually implementable. In this paper, we make progress toward addressing these questions.

We investigate what kind of mechanisms lead to simple informational incentives, and why simple informational incentives matter for the design of institutions. We consider good allocation problems in which agents' valuations for the good are private, and independently drawn. Agents are uncertain about their preferences, but can acquire information about their preferences as well as others' before entering the mechanism. For instance, students facing a school choice mechanism can not only acquire information on their own preferences for the different schools but also learn about how demanded they are. Similarly, bidders in an auction mechanism can not only learn about their own valuation for the good, but also consult firms to gauge the toughness of the competition. Informational simplicity is defined as acquiring information on one's own preferences only, and not on others'.

One might think that strategy-proofness guarantees informational simplicity since it implies agents have a dominant strategy at the interim stage. Our main result is that this is however not the case: for a large set of information acquisition cost functions, and whenever the mechanism violates a separability condition[1]—which is the case of most economically meaningful

[1] Say a mechanism is separable if agents' reports do not interact with one another in the allocation function: for all players $i$, all messages $m_i, m'_i \in M_i$, and all $m_{-i}, m'_{-i} \in M_{-i}$, the allocation function satisfies $x_i(m_i, m_{-i}) - x_i(m'_i, m_{-i}) = x_i(m_i, m'_{-i}) - x_i(m'_i, m'_{-i})$. All standard auction formats and matching algorithms do not satisfy such separability condition.

mechanisms—, players always have an incentive to learn about others' preferences even though they are not *directly* payoff-relevant. In particular, even strategy-proof mechanisms such as VCG incentivize players to acquire information about others. Moreover, we show that such informational incentives make strategy-proof mechanisms no longer dominance solvable at the ex-ante stage, when players decide what information to acquire. These results hold whenever the cost of information satisfies an Inada condition—which makes it never optimal to become *fully* informed about any state—and a smoothness condition—which guarantees that agents can fine-tune the informativeness of signals without discontinuously changing their cost. Importantly, the result holds even though players' underlying preferences are independent and private. Otherwise, players would have a direct incentive to learn about the preferences of others as it would be informative on their own preferences.

The intuition behind our main result is the following: The set of outcomes that a player can bring about, call this her *opportunity set*, depends on other players' reports to the designer, and hence on the entire vector of fundamentals determining the preferences of the population. Since the value of information on her own preferences depends on her opportunity set, it *indirectly* depends on other players' preferences as well. If gathering a little bit of information about others' preferences is not costlier than learning additional information about her own, then it is generically optimal for the player to devote resources to acquiring information about others' preferences first. That helps her predict others' report, allowing her to acquire more information on herself when it is more valuable.

Finally, we explore the implications of our result for mechanism design. First, it appears that strategic simplicity, as captured by strategy-proofness, is more limited than previously thought. Much of the literature focuses on strategic simplicity at the interim stage, that is once players have acquired their private information. Our results show that strategy-proofness does not guarantee strategic simplicity at the ex-ante stage of information acquisition: Indeed, only informationally simple mechanisms admit equilibria in dominant strategies in the extended game. This is one argument as to why informational simplicity might be valuable in practice: It ensures agents have a dominant strategy when deciding what information to acquire, leading to more robust predictions and fewer strategic mistakes. However, our main result implies that only separable mechanisms satisfy such property.

Second, a direct corollary of our main result is that the standard Independent Private value (IPV) assumption is unlikely to arise endogenously. Of course, this assumption is usually understood as a technically convenient approximation of reality—nothing more. Nevertheless, our result makes precise why this is unlikely to hold in practice, and why departures from IPV in the standard framework lead to discontinuities such as Crémer and McLean (1988)'s full surplus extraction result. Instead, our approach restores a form of continuity: Side bets at the interim stage distort information acquisition at the ex-ante stage—in particular, side bets prevent the efficient amount of information acquisition. Hence *constrained* surplus extraction is feasible, but *full* surplus extraction is not because players internalize the informational incentives generated by side bets.

**Related Literature.** A first strand of the literature investigates information acquisition with fixed mechanisms. Persico (2000) proves a representation theorem for the demand for information in several auction formats. Bergemann et al. (2009) show that with interdependent values the equilibrium level of information acquisition is inefficient under VCG. More recently, Bobkova (2019) investigates the incentives to learn about private versus common value components in auctions.

A second strand of the literature investigates optimal mechanism design with information acquisition. Bergemann and Välimäki (2002) show that in a standard allocation problem with monetary transfers, private values, and information acquisition on own preferences only, VCG is ex-ante efficient. Hatfield et al. (2018) strengthen this result by showing that strategy-proofness is also necessary for ex-post efficient mechanisms to induce ex-ante efficient information acquisition. Interestingly, a corollary of our result is that VCG induces ex-ante *inefficient* information acquisition when agents are allowed to learn about others' preferences in addition to their own, as it endogenously leads to interdependent values at the interim stage. In combination with Jehiel and Moldovanu (2001), this shows the infeasibility of implementing the efficient allocation.

In school choice settings, Immorlica et al. (2018) look for mechanisms that are stable and induce students to acquire information efficiently. Roesler and Szentes (2017) and Ravid et al. (2022) consider monopoly pricing when buyers can flexibly acquire information. In their papers, the seller chooses the mechanism after the buyer chooses her information strategy, whereas in our paper the designer ex-ante commits to a mechanism. Hence in their model the buyer must internalize the seller's strategy, whereas in our paper the designer must internalize agents' future decisions (what we refer to as "informational incentives"). Mensch (2022) considers a screening problem with informational incentives and characterizes the optimal mechanism.

Most of the literature investigates information acquisition on one's own preferences only.[2] In this paper, we allow agents to acquire information on others as well, and investigate when it would be optimal for them to do so.

# 2 MOTIVATING EXAMPLE

Two bidders compete in a Second-Price auction to acquire a good. Contrary to the standard approach, bidders are uncertain about their valuation for the good. Bidder 1's valuation is either high $\overline{\omega}_1 \in \Omega_1$ or low $\underline{\omega}_1 \in \Omega_1$, and similarly for bidder 2. We denote the state space by $\Omega = \Omega_1 \times \Omega_2$, and suppose agents' valuations are independently drawn from a uniform distribution, so all four states are equally likely. Let $\overline{\omega}_1 > \overline{\omega}_2 > \underline{\omega}_1 > \underline{\omega}_2$, such that if bidders knew their own valuation for the good and played their dominant strategy, bidder 2 would only win the auction in state $(\underline{\omega}_1, \overline{\omega}_2)$.

[2] A notable exception is Larson and Sandholm (2001) in the computer science literature. They introduce a model in which agents can devote computational resources to discover their own as well as others' valuation. For several auction formats, they show that players compute the valuation of others in equilibrium.

Bidders can engage in costly information acquisition about $\omega = (\omega_1, \omega_2)$ ex ante. Importantly, they can privately acquire information on *any* fundamental, and hence not only learn about their own valuation but also about the other bidder's if they wish to. Information is costly, and bidders trade-off the value and cost of information upon acquiring it. For this example only, we consider the entropic cost function, which has been introduced in the rational inattention literature. Informally, the cost of information is proportional to the expected reduction in uncertainty as measured by the entropy of beliefs:

$$\text{cost of information} = \lambda\Big(\text{prior entropy} - \mathbb{E}[\text{posterior entropy}]\Big)$$

where $\lambda$ is a scaling parameter. This cost function satisfies the key assumptions we impose for our main result: it is smooth, and the marginal cost of becoming fully informed is unbounded (Inada condition). We discuss the necessity of these assumptions later.

We look for an equilibrium in which both agents only learn about their own valuation for the good.[3] For the sake of the example, suppose that each agent's optimal information acquisition strategy leads her to hold one of two posterior beliefs upon entering the auction: one that puts more weight on states in which she has a high valuation $\omega_i = \bar{\omega}_i$, and the other more weight on states where $\omega_i = \underline{\omega}_i$. Let $\overline{m}_i$ be agent $i$'s equilibrium bid at the former belief, and $\underline{m}_i$ her equilibrium bid at the latter. So in equilibrium, either agent $i$ learns that state $\bar{\omega}_i$ is more likely, in which case she submits a high bid $\overline{m}_i$, or that $\underline{\omega}_i$ is more likely, in which case she bids $\underline{m}_i$. Suppose that $\overline{m}_1 > \overline{m}_2 > \underline{m}_1 > \underline{m}_2$, reflecting the ordering of agents' underlying values.[4]

[3]There are also equilibria in which no agent acquires information. For instance, if agent 1 always bids $\overline{\omega}_1$ and agent 2 always bids zero, then neither has an incentive to acquire information and no one wants to deviate. Such equilibrium however yields an outcome that is independent of the state $\omega$—namely, agent 1 always gets the good at zero price. Hence the designer could have achieved the same outcome by using a constant (and hence separable) mechanism that simply allocates the good to agent 1 for free.

[4]It cannot be that $\overline{m}_i > \overline{m}_j > \underline{m}_j > \underline{m}_i$, as that implies agent $j$'s two equilibrium bids always lead to the same outcome: she loses if $i$ bids high and wins if $i$ bids low. Hence agent $j$'s two bids are outcome-equivalent, and $j$ cannot find it optimal to acquire costly information to distinguish between them. The only other possible case is then when $\overline{m}_2 > \overline{m}_1 > \underline{m}_2 > \underline{m}_1$,

Do agents have an incentive to acquire information on the opponent's valuation for the good? We show that the answer is yes, even though the mechanism is strategy-proof at the interim stage, and agents' valuations $\omega_1$, $\omega_2$ are private and independently distributed. To see this, take the perspective of agent $2$. In states where $\omega_1 = \overline{\omega}_1$, her opponent is likely to make a high bid $\overline{m}_1$ ensuring her the object, and learning about her own valuation has no benefit for $2$. On the contrary, in states where $\omega_1 = \underline{\omega}_1$, bidder $2$'s bid impacts the outcome of the auction, and hence information on $\omega_2$ is valuable. Similarly, when $\omega_2 = \underline{\omega}_2$, agent 1 knows that her valuation is higher so does not need acquiring any information on her preferences. When $\omega_2 = \overline{\omega}_2$, she could have a higher or lower valuation than agent 2's most likely bid, and so information on $\omega_1$ is valuable. This shows that the value of information on one's own preferences depends on the realized state for the other. Whether or not agents acquire information on others in equilibrium naturally depends on the cost of information. In particular, agents should not incur a discontinuously high cost upon learning about others, which would offset the associated benefits. Furthermore, agents should not *always* want to become *fully* informed of their own preferences, but should instead equate value and cost of information at the margin.

For the sake of tractability, we characterize what the equilibrium converges to as the scaling parameter $\lambda$ goes to zero.[5] Agents' equilibrium strategies are summarized in Table 1.

and the following reasoning follows verbatim, just exchanging indices across agents.

[5]This simplifies the analysis of the example as it ensures both agents do in fact acquire information in equilibrium. Following Matějka and McKay (2015), we know that equilibrium strategies follow a logit rule under the entropic cost function:

$$\Pr(\overline{m}_1|\omega_1,\omega_2) = \frac{\Pr(\overline{m}_1)}{\Pr(\overline{m}_1) + \Pr(\underline{m}_1)\exp\Big[-\frac{1}{\lambda}\Pr(\overline{m}_2|\omega_1,\omega_2)\,(\omega_1 - \overline{m}_2)\Big]}$$

$$\Pr(\overline{m}_2|\omega_1,\omega_2) = \frac{\Pr(\overline{m}_2)}{\Pr(\overline{m}_2) + \Pr(\underline{m}_2)\exp\Big[-\frac{1}{\lambda}\big(1 - \Pr(\overline{m}_1|\omega_1,\omega_2)\big)\,(\omega_2 - \underline{m}_1)\Big]}.$$

These logit rules make the interdependency between both agents' strategies explicit. For instance, the equilibrium probability agent 2 bids $\overline{m}_2$ in state $(\omega_1,\omega_2)$ is decreasing in the probability her opponent will submit a high bid in that state $\Pr(\overline{m}_1|\omega_1,\omega_2)$.

Table 1: *Equilibrium strategies*

| | $(\overline{\omega}_1, \overline{\omega}_2)$ | $(\overline{\omega}_1, \underline{\omega}_2)$ | $(\underline{\omega}_1, \overline{\omega}_2)$ | $(\underline{\omega}_1, \underline{\omega}_2)$ |
|---|---|---|---|---|
| $\Pr(\overline{m}_1 \vert \omega_1, \omega_2)$ | 1 | $\frac{1}{2}$ | 0 | $\frac{1}{2}$ |
| $\Pr(\overline{m}_2 \vert \omega_1, \omega_2)$ | $\frac{1}{3}$ | 0 | 1 | 0 |

Agent $1$ receives the correct signal, and hence submits the correct bid, whenever the other's valuation is high, as this is when information makes a difference: $\Pr(\overline{m}_1|\overline{\omega}_1, \overline{\omega}_2) \longrightarrow 1$ and $\Pr(\overline{m}_1|\underline{\omega}_1, \overline{\omega}_2) \longrightarrow 0$. Reciprocally, agent $2$ submits the correct bid whenever agent $1$ submits a low bid with non-zero probability, as this is when information is valuable to $2$. An interior probability reflects the fact that the agent is indifferent between the two bids in that state. For instance, in state $(\overline{\omega}_1, \overline{\omega}_2)$, agent 2 is indifferent between bidding high and low as she never wins the auction anyway. Then, optimally, agent 2 does not condition her behavior on this state: $\Pr(\overline{m}_2|\overline{\omega}_1, \overline{\omega}_2) = \Pr(\overline{m}_2) = 0.25 \sum_\omega \Pr(\overline{m}_2|\omega)$ which yields $\Pr(\overline{m}_2|\overline{\omega}_1, \overline{\omega}_2) = 1/3$.[6]

Despite the auction being strategy-proof, each bidder still has an incentive to learn about the opponent's valuation, to assess how much she should learn about her own valuation. Hence strategy-proofness is not enough to guarantee informational simplicity, which illustrates our main finding: a mechanism is informationally simple if and only if it is *de facto* separable, i.e. agents' reports do not interact with one another to determine the allocation. For instance, in the above example, the seller could offer the good to agent 1 at price

[6]For this to be an equilibrium, agent 1 must find it optimal to win the auction at price $\overline{m}_2$ when her signal recommends her to bid $\overline{m}_1$. Similarly, when her signal recommends her to bid low, she must be content with losing against bid $\overline{m}_2$ but winning against $\underline{m}_2$. Overall that requires

$$\mathbb{E}[\omega_1 \mid \overline{m}_1] > \overline{m}_2 > \mathbb{E}[\omega_1 \mid \underline{m}_1] > \underline{m}_2 \iff \frac{3}{4}\overline{\omega}_1 + \frac{1}{4}\underline{\omega}_1 > \overline{m}_2 > \frac{1}{4}\overline{\omega}_1 + \frac{3}{4}\underline{\omega}_1 > \underline{m}_2.$$

Similarly, for agent 2 not to have an incentive to deviate, it must be that

$$\overline{m}_1 > \mathbb{E}[\omega_2 \mid \overline{m}_2] > \underline{m}_1 > \mathbb{E}[\omega_2 \mid \underline{m}_2] \iff \overline{m}_1 > \overline{\omega}_2 > \underline{m}_1 > \frac{1}{4}\overline{\omega}_2 + \frac{3}{4}\underline{\omega}_2.$$

$\overline{\omega}_2$, and if 1 refuses, give it to agent 2 at price $\underline{\omega}_2$. This is a dictatorial mechanism, which hence satisfies our separability condition, as only agent 1 can influence the outcome. It is also informationally simple: agent 1 only wants to learn whether her value is above $\overline{\omega}_2$, and agent 2 does not want to acquire any information at all.

# 3 SETUP

**Environment.** We consider good allocation problems with transferable utilities. A single item needs to be allocated to one of $n$ agents. Let $N = \{1, \ldots, n\}$ be the set of agents. There is a finite set of possible states of the world, or fundamentals, that has a product structure $\Omega = \bigtimes_{i \in N} \Omega_i$, with $|\Omega_i| > 2$.[7] Each player's preferences for the good depend on her own fundamental only: $u_i \in \mathcal{U}_i \subseteq \mathbb{R}^{\Omega_i}$, with $\mathcal{U}_i$ being the open set of possible preferences for $i$. The prior probability distribution $\mu_0 \in \Delta\Omega$ is common knowledge among the players and the designer, and satisfies independence: $\mu_0(\omega) = \prod_{i \in N} \mu_0^i(\omega_i)$ where the superscript $\mu^i$ corresponds to the marginal on dimension $\omega_i$. Should players' preferences depend on the entire vector of fundamentals, or should the fundamentals be correlated, we could not make the distinction between player $i$ acquiring information on her preferences or on others'. Our assumptions ensure that statements such as "player $i$ acquires information on player $j$" have a proper meaning. Moreover, they guarantee that agent $i$ does not have a direct interest in acquiring information on $\omega_{-i}$. Hence in what follows, an agent's payoff will depend on others' fundamentals $\omega_{-i}$ *only indirectly* through the mechanism.

Though we state our results for good allocation problems, they easily extend to more general settings in which there is an abstract set of outcomes and agents have arbitrary preferences over these outcomes. The results also extend to environments without transfers (e.g., matching). We focus on good alloca-

[7] We assume that $\Omega$ is finite for simplicity, but it does not appear to drive our results. We can allow for $|\Omega_i| = 2$ under an additional restriction on the cost of information (see Remark 1).

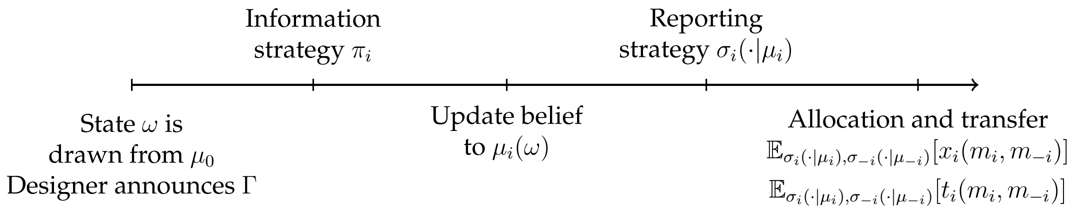


Figure 1: *Timing of the game*

tion problems and quasi-linear utilities for ease of exposition, and to make our impossibility result stronger as transfers give more leeway to the designer.

**Mechanism.** A designer ex-ante commits to a mechanism. A mechanism $\Gamma$ consists of a finite set of messages for each player $M_i$, as well as allocation functions and transfer functions:

$$x_i : M_1 \times \cdots \times M_n \longrightarrow [0, 1]$$
$$t_i : M_1 \times \cdots \times M_n \longrightarrow \mathbb{R}$$

with $\sum_i x_i(m) \leq 1$ for all $m \in M_1 \times \cdots \times M_n$. The ex-post utility of agent $i$ in state $\omega$ under message profile $m$ is quasi-linear in the transfer:

$$x_i(m)u_i(\omega_i) - t_i(m).$$

**Strategies.** At the ex-ante stage, players can acquire costly information about any state.[8] Information acquisition is represented by choosing a distribution over posterior beliefs $\pi_i \in \Delta\Delta\Omega$ that is consistent with the prior $\mathbb{E}_{\pi_i}[\mu_i] = \mu_0$.[9] At the interim stage, players send a message to the designer conditional on

[8]We assume information acquisition is covert for simplicity. None of our results would change if agents could observe the information structure chosen by others before sending their message.

[9]This formulation is equivalent to agents choosing a privately observed signal that can be arbitrarily correlated with the state. Indeed, any signal leads to a specific distribution over posterior beliefs. Reciprocally, any distribution over posteriors $\pi_i$ satisfying this martingale property can be achieved by choosing an appropriate signal (Kamenica and Gentzkow, 2011).

their realized posterior belief $\mu_i$. Let $\sigma_i : \Delta\Omega \longrightarrow \Delta M_i$ denote agent $i$'s reporting strategy. Figure 1 describes the timing of the game. Without loss of generality, we directly work with probability distributions over messages conditional on states: $P_i : \Omega \longrightarrow \Delta M_i$. This object is obtained from $(\pi_i, \sigma_i)$ using Bayes' rule:[10]

$$P_i(m_i \mid \omega) = \sum_{\mu_i \in \text{supp } \pi_i} \sigma_i(m_i \mid \mu_i) \frac{\mu_i(\omega)}{\mu_0(\omega)} \pi_i(\mu_i),$$

where $\text{supp } \pi_i$ denotes the support of $\pi_i$. (Similar notation is used to denote the set of messages in the support of a choice rule $P_i$.) In words, the information strategy and reporting strategy $(\pi_i, \sigma_i)$ lead player $i$ to send message $m_i$ in state $\omega$ with probability $P_i(m_i \mid \omega)$. The more an agent's choice rule in state $\omega$ differs from that in state $\omega'$, the more agent $i$ acquires information to distinguish between states $\omega$ and $\omega'$. Conversely, a choice rule $P_i$ that is independent of $\omega$ can be implemented without acquiring any information about the state.

A Nash Equilibrium is a strategy profile $(P_i^*)_{i \in N}$ such that, for all $i \in N$,

$$P_i^* \in \underset{P_i \in (\Delta M_i)^\Omega}{\arg\max} \sum_{\omega \in \Omega} \mu_0(\omega) \mathbb{E}_{P_i(\cdot|\omega), P_{-i}^*(\cdot|\omega)} \left[ x_i(m_i, m_{-i}) u_i(\omega_i) - t_i(m_i, m_{-i}) \right] - c(P_i)$$

where $c : (\Delta M_i)^\Omega \longrightarrow \mathbb{R}$ is the cost of information acquisition associated with the least informative distribution over posteriors $\pi_i$ that implements $P_i$.[11] This implicitly assumes that more information (in Blackwell's order) is costlier, and

[10]Without loss, an optimal information strategy puts weight on at most $|M_i|$ posteriors, and so we can assume that the support of $\pi_i$ is finite.

[11]Any information and reporting strategies $(\pi_i, \sigma_i)$ induce a unique choice rule $P_i$. Conversely, given a choice rule $P_i$, there exists a least informative distribution over posteriors $\pi_i$ that implements it, in the sense that all other distributions $\pi_i'$ that implement $P_i$ are mean preserving spreads of $\pi_i$. Intuitively, given any choice rule $P_i$, there exists a minimal amount of information the agent has to acquire to be able to correlate her reports to the state as specified by $P_i$. The least informative $\pi_i$ associated with $P_i$ is derived from Bayes rule in the following way. For all $m_i \in \text{supp } P_i$ let

$$\mu_i^{m_i}(\omega) = \frac{P_i(m_i|\omega)\mu_0(\omega)}{\sum_{\omega'} P_i(m_i|\omega')}$$

be $i$'s posterior when she reports $m_i$ to the designer. Then $\text{supp } \pi_i = \{\mu_i^{m_i} | m_i \in \text{supp } P_i\}$ and $\pi_i(\mu_i^{m_i}) = \sum_\omega P_i(m_i|\omega)$.

hence that agents would never buy more information than they need to implement $P_i$. The above formulation makes clear that information on $\omega_{-i}$ can only be valuable to player $i$ if it helps her predict others' equilibrium report $P^*_{-i}(\cdot|\omega)$. Existence of a Nash Equilibrium in pure strategies derives from standard argument, given the assumptions we impose on the cost of information.[12]

**Assumptions on the Cost Function.** As mentioned above, we assume that more information (in Blackwell's order) is costlier. A choice rule $P$ requires more information to be implemented than $P'$ if $P'$ can be derived by adding noise to $P$. Formally, $P'$ is a garbling of $P$ if there exists a positive, column-stochastic matrix $[\Lambda_{m_i,m'_i}]_{m_i,m'_i}$ such that $P'(\cdot|\omega) = \Lambda P(\cdot|\omega)$ for all $\omega$.

**Assumption 1** (Monotonicity)**.** *If $P'$ is a garbling of $P$, then $c(P') < c(P)$.*

Such monotonicity assumption is standard in the literature. Second, we assume that the cost function is smooth—excluding kinks and jumps,—allowing players to fine-tune the informativeness of signals. A stochastic choice rule is interior if its support is the same across all states: $\text{supp}\, P_i(\cdot|\omega) = \text{supp}\, P_i(\cdot|\omega')$ for all $\omega$, $\omega'$. Equivalently, a choice rule is interior if player $i$'s posterior belief always has full support—i.e., player $i$ does not need to rule out some state of the world with certainty in order to implement $P_i$.

**Assumption 2** (Smoothness)**.** *$c$ is twice continuously differentiable and convex over the set of interior stochastic choice rules.*

This ensures that all partial derivatives of $c$ exist and are continuous. In particular, it rules out the possibility that learning about others' preferences

[12] As agents' strategy spaces $(\Delta M_i)^\Omega$ are compact, continuity and convexity of the cost $c$ ensure that best-responses are well-behaved—upper hemicontinuous, non-empty, compact and convex valued—by the Theorem of the Maximum. Kakutani's fixed point theorem then guarantees the existence of an equilibrium. Any mixing between two pure strategies $P$ and $P'$ cannot be strictly optimal as it can always be replicated by a pure strategy at lower cost as soon as $c$ is convex. Note that this does not imply an agent's reporting strategy $\sigma_i$ never involves some mixing in equilibrium: conditional on some posterior $\mu_i$, an agent might send multiple messages with positive probabilities, but this still translates into a pure choice rule $P_i$.

is discontinuously costlier than learning about oneself. More generally, there is no discontinuous change in $c$, or in the partial derivatives of $c$, upon learning a bit about others—e.g., the marginal cost associated with sending some message $m_i$ more often in state $\omega$ is continuous in $P_i$, and does not jump when moving from an informationally simple choice rule $P_i$ to one that is not. Absent this assumption, it is easy to find examples in which agents never acquire information on others, e.g. take a large fixed cost on acquiring information on $\omega_{-i}$ that dominates any benefit from obtaining the good. Since this smoothness assumption is key to our main result, we discuss and relax it in Section 4.1.

Third, we assume that the *marginal* cost of information goes to infinity when a player becomes fully informed about any fundamental (though the total cost can be bounded). This Inada condition guarantees that players' optimal choice rules are interior.

**Assumption 3** (Inada Condition)**.** *For all $\hat{P}_i$ such that $\hat{P}_i(m_i|\omega) = 0$ and $\hat{P}_i(m_i|\omega') > 0$ for some $m_i, \omega, \omega'$,*

$$\lim_{P_i \to \hat{P}_i} \frac{\partial c(P_i)}{\partial P_i(m_i|\omega)} = -\infty.$$

In the statement of Assumption 3, $\hat{P}_i$ is a corner choice rule as agent $i$ needs to know with probability one that state $\omega$ has not realized in order to implement it. The condition then requires that the marginal cost of this strategy is unbounded, which implies that it is never optimal at equilibrium. Absent this assumption, we can construct environments in which it is always optimal for agents to become *fully* informed about their own preferences, as soon as they have some impact over the outcome. Then, in any strategy-proof mechanism, agents would not have an incentive to acquire information on others.

Finally, we impose that if $\omega_{-i}$ is directly and indirectly payoff irrelevant to agent $i$—think of a dictatorial mechanism where $i$ is the dictator—then player $i$ has no incentive to acquire information on $\omega_{-i}$. This condition is only necessary for the converse of our main result, i.e. to show that a separable mechanism is informationally simple. To define this condition formally, let

$V_i(m_i, \omega | P_{-i}, \Gamma) = \mathbb{E}_{P_{-i}(\cdot|\omega)}\Big[x(m_i, m_{-i})u_i(\omega_i) - t_i(m_i, m_{-i})\Big]$ be $i$'s expected payoff from sending message $m_i$ in state $\omega$ given some $P_{-i}$ and mechanism $\Gamma$, and denote by $P_i^*$ an optimal choice rule for player $i$.

**Assumption 4** (Independence of Irrelevant States). *For any mechanism $\Gamma$ and any strategy of others $P_{-i}$, the following must hold: If $V_i(m_i, (\omega_i, \omega_{-i})|P_{-i}, \Gamma)$ is independent of $\omega_{-i}$ for all $m_i$, then so is $P_i^*(\cdot|\omega_i, \omega_{-i})$.*

In words, if an agent's payoff is independent of some dimensions of the state space, then it is not optimal to learn about these payoff-irrelevant dimensions and the induced optimal choice rule does not depend on them. This relates to other assumptions brought forward in the literature.[13] This assumption in particular rules out the possibility that learning about $\omega_i$ is cheaper if one also learns about $\omega_{-i}$.

The most notable example of a cost function satisfying all four conditions is the entropic cost function, which we considered in the motivating example.

**Example 1** (Entropic Cost). *Sims (2003) proposes a cost function based on Shannon's entropy which measures a signal's informativeness as the expected reduction in entropy. The entropic cost associated with a stochastic choice rule $P_i$ writes*

$$c(P_i) = -\sum_{m_i \in \operatorname{supp} P_i} P_i(m_i) \log P_i(m_i) + \sum_{\omega \in \Omega} \mu_0(\omega) \sum_{m_i \in \operatorname{supp} P_i} P_i(m_i|\omega) \log P_i(m_i|\omega),$$

*where $P_i(m_i) = \sum_\omega P_i(m_i|\omega)\mu_0(\omega)$ is the unconditional probability of sending message $m_i$ under $P_i$. Intuitively, the more the agent's choice rule $P_i(\cdot \mid \omega)$ varies across states $\omega$, the higher the cost, as more information about $\omega$ is needed to implement it.*

[13] Independence of irrelevant states is related to but weaker than "invariance under compression" (Caplin et al. (2022)), which is known to hold for the entropic cost function. Invariance under compression requires that splitting a state into two payoff-equivalent states should not change how costly it is to learn about it. In particular, splitting a state $\omega_i$ into $|\Omega_{-i}|$ payoff-equivalent states should not change agent $i$'s optimal strategy, which is our above condition. Our condition only requires the cost function to be invariant under compression of some dimensions of the state space (i.e., only of $\Omega_{-i}$), which is reminiscent of a similar condition in Hébert and La'O (2020). Our notion of independence to irrelevant states is conceptually distinct from prior independence of the cost function. In our setting, agents' prior belief $\mu_0$ is fixed, and so whether the cost of information depends on it or not is immaterial.

We end this section with a comment on static vs. dynamic information acquisition. In our model, agents' choices are static: they simultaneously choose a choice rule $P_i$, or equivalently a signal and a reporting strategy. This is, however, not restricting the manner in which agents can acquire information *per se*. Indeed, any dynamic information acquisition process can be reduced to a single, appropriately chosen, signal. For a large class of cost functions, such reduction is without loss as it leads to a weakly lower overall cost of information.[14] This is for instance the case of the entropic cost function, which can be interpreted as the reduced-form expected cost of an optimal binary search tree over the state space.

# 4 The Generic Complexity of Informational Incentives

In this section we address the following question: Which mechanisms provide players with simple informational incentives, i.e. incentives to only acquire information on their own preferences? We show that players have simple informational incentives if and only if the mechanism is *de facto* separable.

Informational simplicity is captured by the following refinement of Nash equilibrium.

**Definition 1.** *An* ***Informationally Simple Equilibrium*** *(ISE) is a Nash equilibrium $(P_i^*)_{i\in N}$ such that, for all $i$, $P_i$ is independent of $\omega_{-i}$.*[15]

This refinement is of interest for several reasons. First, informational simplicity captures a notion of strategic simplicity which is a priori distinct from strategy-proofness. As it turns out, we will show that informational simplicity is a necessary condition for ex-ante strategy-proofness of the extended game that includes the information acquisition stage. Second, players' interim information structure satisfies the Independent Private value (IPV) assumption

[14]See Zhong and Bloedel (2020) for a characterization of cost functions satisfying this property.

[15]Equivalently, the signal that agent $i$ acquires is independent of $\omega_{-i}$.

if and only if the equilibrium is informationally simple. Hence our analysis shades light on whether we should expect such information structure to arise endogenously.

Given others' strategy $P_{-i}^*$, player $i$ chooses a stochastic choice rule $P_i$ so as to solve the following program:

$$\max_{P_i:\Omega\to\Delta M_i} \quad \sum_{\omega\in\Omega} \mu_0(\omega)\mathbb{E}_{P_i(\cdot|\omega),P_{-i}^*(\cdot|\omega)}\Big[x_i(m_i,m_{-i})u_i(\omega_i) - t_i(m_i,m_{-i})\Big] - c(P_i).$$

Conditional on acquiring some information, the first-order conditions with respect to $P_i(m_i|\omega)\mu_0(\omega)$ yield necessary restrictions on player $i$'s best response: for all $m_i$ in the support of $P_i^*$ and for all $\omega \in \Omega$,

$$(\star) \qquad \mathbb{E}_{P_{-i}^*(\cdot|\omega)}\Big[x_i(m_i,m_{-i})u_i(\omega_i) - t_i(m_i,m_{-i})\Big] + \frac{\gamma_i(\omega)}{\mu_0(\omega)} = \frac{\partial c(P_i^*)}{\partial P_i^*(m_i|\omega)\mu_0(\omega)},$$

where $\gamma_i(\omega)$ is the Lagrange multiplier associated with the constraint that the choice rule $P_i(\cdot|\omega)$ must sum to one. The left-hand side captures the marginal gain from sending message $m_i$ in state $\omega$ (had player $i$ been fully informed of the state), rather than any other message $m_i'$. The Lagrange multiplier $\gamma_i(\omega)$ is indeed the shadow price of the constraint that the choice rule sums to one, and hence captures the fact that sending $m_i$ more often implies sending $m_i' \neq m_i$ less often. The right-hand side represents the marginal cost associated with sending message $m_i$ more often in state $\omega$.

To better understand this trade-off, suppose that the cost of any information is very high: then players send only one message that maximizes their average payoff across states. Conversely, if the marginal cost is sufficiently low, then agents send exactly the payoff maximizing message in each state as in a game with perfect information. Hence, for intermediate costs, players achieve a trade-off between the gains from sending the optimal message and the cost associated with discovering what is the optimal message.

We can rearrange and substitute out the Lagrange multiplier to obtain an interpretation of the FOC in terms of value of information. Take the FOC with respect to $P_i(m_i' \mid \omega)\mu_0(\omega)$ and substract from the previous FOC. This gives the marginal gain from reporting $m_i$ relative to $m_i'$ in state $\omega$, net of marginal costs:

$$\mathbb{E}_{P^*_{-i}(\cdot|\omega)}\Big[x_i(m_i, m_{-i})u_i(\omega_i) - t_i(m_i, m_{-i}) - [x_i(m_i', m_{-i})u_i(\omega_i) - t_i(m_i', m_{-i})]\Big] \\ = \frac{\partial c(P_i^*)}{\partial P_i^*(m_i|\omega)\mu_0(\omega)} - \frac{\partial c(P_i^*)}{\partial P_i^*(m_i'|\omega)\mu_0(\omega)}.$$

In general the value of information for $i$ seems to depend on other agents' reports, as $m_{-i}$ impacts the chosen outcome. In an informationally simple equilibrium, however, player $i$'s strategy must be independent of $\omega_{-i}$. This requires the value of information for player $i$ to be independent of other players' realized state.

We now show that this independence cannot generically be satisfied unless the mechanism is *de facto* separable. A statement holds **generically** if it is false only for a set of utilities $\mathcal{U}^0 \subseteq \bigtimes_i \mathcal{U}_i$ whose closure has Lebesgue measure zero.[16] A mechanism is **separable** if agents' reports do not interact with one another in the allocation function: for all $i$, all $m_i, m_i' \in M_i$, and all $m_{-i}, m_{-i}' \in M_{-i}$,

$$x_i(m_i, m_{-i}) - x_i(m_i', m_{-i}) = x_i(m_i, m_{-i}') - x_i(m_i', m_{-i}').$$

Hence others' report $m_{-i}$ can only impact the level in $i$'s outcome, but cannot interact with how $i$'s report affects her outcome. Note that if $i$ cannot influence the outcome altogether, then the condition is trivially satisfied as both sides of the equations are always zero. More generally, it can be that several agents influence the outcome with positive probability, but never *jointly*—e.g., the mechanism might randomly (and independently of reports $m$) pick a dictator $i^*$ and condition the outcome on her report $m_{i^*}$ only. Finally, say a mechanism $\Gamma$ is ***de facto* separable** under equilibrium $P^*$ if there exists a separable mechanism $\hat{\Gamma}$ and an equilibrium $\hat{P}^*$ of $\hat{\Gamma}$ that is outcome equivalent.[17]

[16]Mathematical genericity does not necessarily imply genericity "in practice" and very much depends on the chosen universe of preferences. We show that Informational Simplicity is only feasible for preferences that are non-generic, within the open set of allowed preferences $\bigtimes_i \mathcal{U}_i$. But the set of preferences that are relevant in practice could be itself small and non-generic.

[17]That is, an equilibrium $\hat{P}^*$ of $\hat{\Gamma}$ leads to the same state-dependent allocation:

In most settings of interest, the goal of running a mechanism is precisely to aggregate agents' information, and to choose an allocation based on all pieces of information jointly. Yet we show that this is generically incompatible with informational simplicity.

**Theorem 1.** *Fix any mechanism $\Gamma$. Generically, if $P^*$ is an Informationally Simple equilibrium of $\Gamma$, then $\Gamma$ is de facto separable under $P^*$.*

*Conversely, if both the outcome functions and transfer functions of $\Gamma$ are separable,*[18] *then all equilibria of $\Gamma$ are Informationally Simple.*

Theorem 1 states that informational simplicity is generically impossible to achieve under most mechanisms of interest. Whenever the mechanism is not de facto separable, it is generically impossible to design transfers that incentivize agents to only learn about themselves. The economic intuition is that players have uncertainty about their "opportunity set," i.e. which outcomes they can bring about. Their opportunity set depends on others' preferences, which makes it valuable to condition how much they learn about their own preferences on the realization of others'. That allows them to acquire more information on their own preferences when the stakes are higher—i.e., when they face a larger opportunity set.[19] Doing so is free at the margin, as the smoothness of the cost function implies it has no discontinuous jump or kink when agents start acquiring information on others. In the proof, we show that this interdependence between the value of information for $i$ and others' preferences is so rich that it cannot be offset by appropriately designed transfers.

$\sum_{\hat{m}\in\hat{M}}\prod_i \hat{P}_i^*(\hat{m}_i|\omega)\hat{x}_i(\hat{m}) = \sum_{m\in M}\prod_i P_i^*(m_i|\omega)x_i(m)$ for all $i$, $\omega$.

[18] That is $t_i(m_i, m_{-i}) - t_i(m_i', m_{-i}) = t_i(m_i, m_{-i}') - t_i(m_i', m_{-i}')$ for all $i$, $m_i$, $m_i' \in M_i$, $m_{-i}$, $m_{-i}' \in M_{-i}$, and similarly for $(x_i)_i$. Note that our baseline definition of separability only imposes restrictions on the outcome functions $(x_i)_i$. This may however not be enough to guarantee Informational Simplicity, as interdependencies in transfers $(t_i)_i$ might incentivize agents to learn about each other. For instance, if transfers generate a coordination game across agents, then agents might coordinate on conditioning their play on one particular agent's state $\omega_i$, which then implies all agents $j \neq i$ learn about another person.

[19] Note that agents do not care about others' preferences $\omega_{-i}$ *per se*, but only because it helps them predict others' report to the designer. Hence if agents were allowed to acquire information on what others know—i.e. their posterior beliefs $\mu_{-i}$—as in Denti (2018), then they would do so instead of learning about their underlying preferences $\omega_{-i}$.

Note that under an interim strategy-proof mechanism, agents want to learn about others only because it helps them assess how much they should learn about themselves. Hence it is important for Theorem 1 that agents do not know their own preferences, and that becoming fully informed on their own preferences is never optimal by the Inada condition. It is also essential that agents be able to condition how much they learn about themselves on what they learn about others. This is most intuitive if learning is sequential—e.g., if agents first buy a signal about $\omega_{-i}$ and then, conditional on its realization, buy a signal on $\omega_i$. This is captured by our framework as $c$ can be interpreted as the reduced-form cost of an optimal dynamic process of information acquisition.[20]

The necessity part of the theorem is not a statement about the primitives: if $P^*$ is an Informationally Simple equilibrium of $\Gamma$ then the mechanism *need not be* separable, but under $P^*$ the mechanism acts *as if* it were separable.[21] For instance, when the value of information is very small for all but one player, it is very possible that even if all agents can impact the outcome in the mechanism, only one decides to acquire information in equilibrium. The designer could have then replicated the induced outcome by running a dictatorial mechanism, in which only that agent's private information would have been elicited. That being said, the theorem is sufficient for mechanism design purposes: the point is that informational simplicity is impossible to achieve unless the designer's objective does not require eliciting multiple agents' information and using it jointly to decide on the outcome. Whether or not the mechanism is truly separable or only separable *de facto* is immaterial.

**Remark 1.** The assumption of independence of irrelevant states (Assumption 4) only comes into play to prove that under a separable mechanism, all equilibria are informationally simple (sufficiency). Indeed, under a separable mechanism, others' preferences $\omega_{-i}$ are (directly and indirectly) payoff-irrelevant to agent $i$. Hence her choice rule is independent of $\omega_{-i}$ only if As-

[20]See the discussion at the end of Section 3.

[21]Any separable mechanism—such as a dictatorial mechanism—is *de facto* separable, but every non-separable mechanisms—such as VCG—can also be *de facto* separable.

sumption 4 holds. This is for instance the case under the entropic cost (Example 1). Even though we do not need Assumption 4 to prove the necessity part of Theorem 1, imposing it simplifies the proof greatly. In the Appendix, we include both the general proof that does not require Assumption 4, and the shorter one that does. In the latter, we can dispense of the assumption that $|\Omega_i| > 2$.

**A Knife-Edge Example.** We now go through a knife-edge case for which our main result does not hold. Indeed, since the latter is a genericity result, there exists a degenerate set of utility functions under which a mechanism can be both informationally simple and non-separable.

There are three goods $\{A, B, C\}$ to be allocated to two agents $\{i, j\}$.[22] For simplicity, there are no transfers. The mechanism used is a simultaneous version of the serial dictatorship: Agents report their preferences, agent $i$ gets her favorite good, then agent $j$ her favorite good among the remaining ones.

Agent $i$'s most preferred good is either $A$ or $B$: $u_i = (u_{iA}, u_{iB}, u_{iC}) \in \{(4, 2, 1), (2, 4, 1)\}$. Agent $j$ always values good $A$ and $B$ equivalently: $u_j \in \{(2, 2, 0), (0, 0, 2)\}$. An agent's allocation $x_i = (x_{iA}, x_{iB}, x_{iC})$ specifies with which probability she receives each good. Agent $i$ and $j$ each have two possible messages they can send to the designer: $M_i = \{m_A, m_B\}$ and $M_j = \{m_{AB}, m_C\}$. Intuitively, think of a message $m_l$ as indicating to the designer that the agent wants good $l$. The allocation function is given in Table 2.

Table 2: *Allocation Function*

| $x_i(m_i, m_j)$ | $m_{AB}$ | $m_C$ |
|---|---|---|
| $m_A$ | $(1, 0, 0)$ | $(1, 0, 0)$ |
| $m_B$ | $(0, 1, 0)$ | $(0, 1, 0)$ |

| $x_j(m_i, m_j)$ | $m_{AB}$ | $m_C$ |
|---|---|---|
| $m_A$ | $(0, 1, 0)$ | $(0, 0, 1)$ |
| $m_B$ | $(1, 0, 0)$ | $(0, 0, 1)$ |

[22] Knife-edge examples with only one good are more complex and require non-zero transfers, and so we give one with several goods for ease of exposition.

As in the motivating example, consider the entropic cost function. We know that optimal choice rules then follow a logit rule (Matějka and McKay (2015)). Since agent $i$'s allocation does not depend on agent $j$'s report, she only learns about her own preferences in equilibrium: the equilibrium probability that $i$ reports $m_A$ in state $\omega$ equals

$$P_i^*(m_A|\omega) = \frac{P_i^*(m_A)\exp[\frac{1}{\lambda}u_{iA}(\omega_i)]}{P_i^*(m_A)\exp[\frac{1}{\lambda}u_{iA}(\omega_i)] + P_i^*(m_B)\exp[\frac{1}{\lambda}u_{iB}(\omega_i)]},$$

with $P_i^*(m_l) = \sum_\omega P_i^*(m_l|\omega)$. Note that it is independent of $\omega_j$: agent $i$ does not acquire any information about $j$'s preferences. More surprisingly, agent $j$ also only learns about her own preferences, despite the fact that her allocation is impacted by $i$'s report: the equilibrium probability that $j$ reports $m_{AB}$ in state $\omega$ equals

$$P_j^*(m_{AB}|\omega) = \frac{P_j^*(m_{AB})\exp[\frac{1}{\lambda}u_{jA}(\omega_j)]}{P_j^*(m_{AB})\exp[\frac{1}{\lambda}u_{jA}(\omega_j)] + P_j^*(m_C)\exp[\frac{1}{\lambda}u_{jC}(\omega_j)]},$$

which is independent of $\omega_i$.[23] That is because her utility function has some symmetry that makes the value of information on $\omega_j$ independent of $m_i$. Indeed, if agent $i$ picks good $A$, agent $j$ is left choosing between $B$ and $C$, and the utility value of making the correct choice is 2. Similarly, if agent $i$ picks good $B$, agent $j$ is left choosing between $A$ and $C$, and the utility value of making the correct choice is, again, 2. The equilibrium is informationally simple, even though two agents acquire information and jointly impact the outcome in equilibrium.

[23] To derive this logit rule, we use the same formula as for the motivating example: agent $j$ reports $m_{AB}$ in state $\omega$ with probability

$$\frac{P_j^*(m_{AB})\exp[\frac{1}{\lambda}((1-P_i^*(m_A|\omega))u_{jA}(\omega_j) + P_i^*(m_A|\omega)u_{jB}(\omega_j))]}{P_j^*(m_{AB})\exp[\frac{1}{\lambda}((1-P_i^*(m_A|\omega))u_{jA}(\omega_j) + P_i^*(m_A|\omega)u_{jB}(\omega_j))] + P_j^*(m_C)\exp[\frac{1}{\lambda}u_{jC}(\omega_j)]},$$

which simplifies to the above expression since $u_{jB}(\omega_j) = u_{jA}(\omega_j)$ in all states $\omega_j$.

### 4.1 Discussion on Fixed Costs

The proof of Theorem 1 leverages the assumption that the cost of information is smooth. This seems to be a relevant approximation in some settings. For instance, consider a school choice problem in which students must send a rank-order list of schools to a central authority, and can beforehand acquire information on the different schools. They can learn about their own preferences over schools—e.g., by looking at the set of courses offered and whether they look interesting to them—but also about others'—e.g., by asking about the popularity of the school, and admission cutoffs. Arguably, acquiring some information about a school's popularity is not very costly.

A natural concern is that in other settings, the smoothness assumption may be missing relevant factors, and in particular overlooks the possibility that learning about others may be discontinuously harder than learning about oneself. This would mechanically make Informational Simplicity easier to achieve, and we investigate the robustness of Theorem 1 to such discontinuity.

Consider the same smooth cost of information as in our main setup, but suppose that as soon as an agent decides to learn a bit about others, it has to pay an additional fixed cost $\kappa$. Fix an arbitrary mechanism $\Gamma$ and let $\mathcal{U}_{IS}(\kappa) \subseteq \bigtimes_i \mathcal{U}_i$ be the set of utility functions for which (i) there exists an informationally simple equilibrium $P^*$ of $\Gamma$, and (ii) $\Gamma$ is not de facto separable. Let $\rho(\kappa)$ be the Lebesgue measure of $\mathcal{U}_{IS}(\kappa)$.

**Proposition 1.** *For any mechanism $\Gamma$, $\rho(\kappa)$ is increasing and continuous in $\kappa$, with $\rho(0) = 0$.*

Theorem 1 corresponds to the corner case in which $\kappa = 0$: It is generically impossible to design a mechanism that admits an informationally simple equilibrium and that is non-separable under that equilibrium. As $\kappa$ increases, the set of preferences for which informational simplicity can be achieved by non-separable mechanisms grows. Interestingly, it grows *continuously*, so our benchmark with $\kappa = 0$ is not a knife-edge case: Adding a small cost to learning about others does make informational simplicity easier to achieve, but only in very few settings. It is only as $\kappa$ tends to infinity that informational simplicity becomes generically feasible.

# 5 IMPLICATIONS FOR MECHANISM DESIGN

## 5.1 The Limits of Strategy-Proofness

Strategic simplicity is valued in mechanism design for robustness and for leveling the playing field across players. A lot of attention has been given to interim strategy-proof mechanisms, i.e., mechanisms under which agents have a dominant strategy at the interim stage, taking as given their private information. Formally, interim strategy-proofness requires that, for all $i$ and $\mu_i$, there exists $m_i \in M_i$ such that

$$\begin{aligned} x_i(m_i, m_{-i})\mathbb{E}_{\mu_i}[u_i(\omega_i)] - t_i(m_i, m_{-i}) \\ \geq x_i(m_i', m_{-i})\mathbb{E}_{\mu_i}[u_i(\omega_i)] - t_i(m_i', m_{-i}) \quad \forall m_i' \in M_i, m_{-i} \in M_{-i}. \end{aligned}$$

Little is known, however, about the strategic complexity of the acquisition of agents' private information at the ex-ante stage. This is important as many inequalities may arise due to suboptimal information acquisition and strategic mistakes.

Say a mechanism is ex-ante strategy-proof if agents have a dominant strategy in the overall game that includes the information acquisition stage. Formally, for all agent $i$, there exists a choice rule $P_i$ such that

$$\begin{aligned} \mathbb{E}_{P_i, P_{-i}}[x_i(m)u_i(\omega_i) - t_i(m)] - c(P_i) \\ \geq \mathbb{E}_{P_i', P_{-i}}[x_i(m)u_i(\omega_i) - t_i(m)] - c(P_i') \quad \forall P_i' \in (\Delta M_i)^{\Omega}, P_{-i} \in (\Delta M_{-i})^{\Omega}. \end{aligned}$$

We show that informational simplicity is a necessary condition for ex-ante dominance solvability.

**Proposition 2.** *Fix any mechanism $\Gamma$. If $P^*$ is an equilibrium in dominant strategy of $\Gamma$, then $P^*$ is informationally simple.*

The standard notion of strategy-proofness ensures that agents have a dominant strategy once they have acquired information, but the stronger requirement of Informational Simplicity is needed to guarantee agents also have a dominant strategy when choosing what information to acquire. The intuition behind Proposition 2 is that information about others is valuable because it helps predict their reports at the interim stage. Hence if an agent learns about another, her equilibrium information strategy has to depend on the strategy of the other player: the mechanism is not ex-ante dominance solvable.

Together, Proposition 2 and Theorem 1 yield that, generically, there exists an equilibrium in dominant strategy of $\Gamma$ only if $\Gamma$ is de facto separable. Therefore, in the extended game that includes information acquisition, agents virtually never have a dominant strategy under non-separable mechanisms.

## 5.2 Independent Private Values

A direct corollary of Theorem 1 is that the standard Independent Private value assumption is unlikely to arise endogenously.

**Corollary 1.** *Fix any mechanism $\Gamma$. Generically, the equilibrium posterior beliefs $(\mu_i, \mu_{-i})$ are (unconditionally) independent across players only if $\Gamma$ is de facto separable.*

Therefore the interim information structure is endogenously correlated, which creates interdependent values across players. Put differently, the IPV assumption does not arise endogenously whenever the mechanism is non-separable and the technology of information acquisition satisfies our conditions.

Why has research in mechanism design been limited to the IPV case despite the practical importance of information correlation? A theoretical argument due to Crémer and McLean (1988) suggests that as soon as there is some correlation in agents' ex-ante private information,[24] the designer can extract all surplus by constructing appropriate side bets. This result highlights that

[24] Formally, whenever the matrix of the conditional probabilities of other players' signals conditional on one's own signal has full rank.

the *independence* of private information across players is necessary for them to earn an information rent as in Myerson (1981). The limits of that result to risk aversion, limited liability, collusion among the agents, etc. have been explored extensively. However what has been explored less is how such results rely on the exogenous nature of private information: If agents anticipate the designer will exploit the correlation structure in their information, why would they acquire such information in the first place? We show that full surplus extraction (in Nash equilibrium) is generically impossible to achieve when taking into account informational incentives. Therefore, our main result together with the impossibility of full surplus extraction suggest that there is room for studying mechanism design with correlated information.

First, we need to properly define what full surplus extraction means in a setting where private information is endogenous. We say full-surplus extraction is feasible if there exists a mechanism that can extract the maximal surplus that can be generated in the economy. Namely, given an environment and a technology of information acquisition, there exists a maximal total surplus that can be generated, which balances total gains from the allocation and total information costs. Full surplus extraction requires that we reach an equilibrium that generates this surplus, and then extract it entirely using transfers.

As in Crémer and McLean (1988), there is one good to be allocated. (Our result easily extends to multiple goods.) Let the **ex-post efficient allocation** at belief profile $\mu = (\mu_i)_{i \in N}$ be the allocation that maximizes total expected welfare:

$$x^*(\mu) \in \arg\max_{x \in \Delta N} \sum_{i \in N} \sum_{\omega \in \Omega} \mu_i(\omega) u_i(\omega_i) x_i.$$

The **maximum total surplus** that can be generated in the economy equals:

$$\text{Max. Total Surplus } = \max_{P \in \times_i \mathcal{P}_i} \sum_{i \in N} \sum_{\omega \in \Omega} \mu_0(\omega) \mathbb{E}_{P_i, P_{-i}} \left[ x_i^*(\mu) u_i(\omega_i) \right] - c(P_i),$$

where $\mathcal{P}_i = \{P_i \in (\Delta\Omega)^\Omega \mid \mu_i(\omega) = \frac{P_i(\mu_i|\omega)\mu_0(\omega)}{\sum_{\omega'} P_i(\mu_i|\omega')\mu_0(\omega')} \forall \mu_i \in \operatorname{supp} P_i\}$ is the set of distribution over beliefs that are consistent with Bayes' rule. Let $P^\dagger$ be a strategy profile that maximizes total surplus. Note that if $P^\dagger$ is informationally simple—that is, it is socially efficient to have agents acquiring information on themselves only—then using side bets to extract all surplus is trivially precluded. However, we know from Theorem 1 that this is generically not the case whenever the ex-post efficient allocation is non-separable on the support of $P^\dagger$.[25] Hence, whenever efficiency requires that multiple agents learn about their valuations for the good, $P^\dagger$ is generically not informationally simple: it is more efficient for an agent to condition her learning about herself on others' valuations so as to save on information costs whenever possible. In most settings of interest, the information structure that maximizes total surplus then exhibits interdependent beliefs across agents, and allows in principle for the possibility of side bets à la Crémer McLean. We however show that extracting all surplus is generically infeasible, as the anticipation of such side bets distorts agents' incentive to acquire information ex ante.

A (direct revelation) mechanism[26] **extracts the full surplus** if it induces an equilibrium $P^*$ such that:

$$\sum_{i\in N}\sum_{\omega\in\Omega}\mu_0(\omega)\mathbb{E}_{P_i^*,P_{-i}^*}[t_i(\mu_i,\mu_{-i})] = \text{ Max. Total Surplus},$$

while satisfying incentive and individual rationality constraints. Incentive constraints are of two sorts here: agents should be incentivized to reveal their private information to the designer at the interim stage, and should find it optimal to acquire the socially efficient level of information at the ex-ante stage. The former is the standard IC constraint in mechanism design, and from now on, suppose it holds. The latter, which is the one limiting the possibility of full

[25] Indeed, if $P^\dagger$ is informationally simple, then we can design a mechanism which is informationally simple and not de facto separable, by setting $M_i = \operatorname{supp} P_i^\dagger$, $x(\mu_i,\mu_{-i}) = x^*(\mu)$ and transfers such that $\mathbb{E}_{\mu_0,P_{-i}^\dagger(\cdot|\omega)}[t_i(\mu_i,\mu_{-i})] = -\mathbb{E}_{\mu_0,P_{-i}^\dagger(\cdot|\omega)}[\sum_{j\neq i} x_j^*(\mu)u_j(\omega_j)]$. Such a mechanism can however exist only for a non-generic set of preferences.

[26] The standard Revelation Principle applies in our setting: If $(P_i^*)_{i\in N}$, or equivalently $(\pi_i^*,\sigma_i^*)_{i\in N}$, is an equilibrium of $\Gamma$ then there is an outcome-equivalent direct revelation mechanism $\widehat{\Gamma}$ in which the principal elicits agents' beliefs $\widehat{M}_i = \operatorname{supp}\pi_i^*$ and commits to implementing their equilibrium strategy $(\hat{x}(\mu_i,\mu_{-i})),\hat{t}(\mu_i,\mu_{-i})) = (x(\sigma^*(\mu_i,\mu_{-i})),t(\sigma^*(\mu_i,\mu_{-i})))$.

surplus extraction in this setting, writes:

$$\sum_{\omega\in\Omega}\mu_0(\omega)\mathbb{E}_{P_i^\dagger,P_{-i}^\dagger}\left[x_i^*(\mu)u_i(\omega_i)-t_i(\mu_i,\mu_{-i})\right]-c(P_i^\dagger)$$
$$\geq\sum_{\omega\in\Omega}\mu_0(\omega)\mathbb{E}_{P_i,P_{-i}^\dagger}\left[x_i^*(\mu)u_i(\omega_i)-t_i(\mu_i,\mu_{-i})\right]-c(P_i)\quad\text{for all } P_i, i.$$

Finally, the mechanism should satisfy the following ex-ante and interim individual rationality constraints:

$$\sum_{\omega\in\Omega}\mu_0(\omega)\mathbb{E}_{P_i^\dagger(\cdot|\omega),P_{-i}^\dagger(\cdot|\omega)}\left[x_i^*(\mu)u_i(\omega_i)-t_i(\mu_i,\mu_{-i})\right]-c(P_i^\dagger)\geq 0\quad\text{for all } i.$$
$$\sum_{\omega\in\Omega}\mu_i(\omega)\mathbb{E}_{P_{-i}^\dagger(\cdot|\omega)}\left[x_i^*(\mu)u_i(\omega_i)-t_i(\mu_i,\mu_{-i})\right]\geq 0\quad\text{for all } i, \mu_i\in\operatorname{supp}P_i^\dagger.$$

Hence to extract the full surplus, the mechanism must (i) induce agents to acquire the socially efficient level of information, (ii) pick the ex-post efficient allocation given reported posterior beliefs, and (iii) have transfers that extract all surplus net of information acquisition costs. The last two requirements are familiar from Crémer and McLean (1988), whereas the first one is new but necessary to make sense of ex-post efficiency.

Observe that in some extreme cases full surplus extraction is possible. For instance, consider a setting in which the efficient allocation is the same in every state $\omega$. This means that $x^*(\mu)=x^*$ is independent of agents' posterior beliefs and that the efficient information strategy is to acquire no information at all. The mechanism that always selects outcome $x^*$ irrespective of agents' reports, and has transfers $t_i=\sum_\omega \mu_0(\omega)x_i^*u_i(\omega_i)$ is individually rational, incentive compatible, and extracts full surplus.

However informational incentives generically limit the possibility of full surplus extraction whenever it is efficient for agents to acquire some information. To prove this, we show that the three requirements exposed above translate into necessary conditions that are generically mutually incompatible. We first focus on requirements (i) and (ii) of full surplus extraction, namely that the mechanism induces socially efficient information acquisition and implements the ex-post efficient allocation. Conditional on $P^\dagger_{-i}$, agent $i$'s optimal strategy solves:

$$\max_{P_i} \sum_{\omega \in \Omega} \mu_0(\omega) \mathbb{E}_{P_i, P^\dagger_{-i}} \left[ x^*_i(\mu) u_i(\omega_i) - t_i(\mu_i, \mu_{-i}) \right] - c(P_i).$$

The standard approach to incentivize efficient information acquisition is to use VCG transfers, such that an agent's expected transfer upon reporting $\mu_i$ equals others' expected payoff at that report:

$$\mathbb{E}_{\mu_0, P^\dagger_{-i}(\cdot|\omega)} \left[ t_i(\mu_i, \mu_{-i}) \right] = -\mathbb{E}_{\mu_0, P^\dagger_{-i}(\cdot|\omega)} \left[ \sum_{j \neq i} x^*_j(\mu_i, \mu_{-i}) u_j(\omega_j) \right].$$

We show that such transfers are actually the *only* one inducing agents to acquire the socially efficient level of information. This result is reminiscent of Hatfield et al. (2018) who extend the Green–Laffont–Holmström theorem by showing that VCG mechanisms with ex-ante costly investments are the unique efficient and strategy-proof mechanisms. To obtain a Crémer-McLean mechanism and enforce the third requirement of full surplus extraction, the designer would need to add side-bets $b_i : \bigtimes_{j \neq i} \Delta\Omega \longrightarrow \mathbb{R}$. Such side bets, however, generically distort *informational* incentives, that is incentives to acquire the efficient level of information. This reduces the total surplus generated by the mechanism, preventing full surplus extraction.

**Theorem 2.** *Suppose that it is socially efficient for at least one agent to acquire some information. Then full surplus extraction is generically infeasible.*

The feasibility of full surplus extraction with information acquisition received mixed answers in the literature. For instance, Bikhchandani (2010) shows that when the set of signals agents can acquire on others' type is small enough, then full surplus extraction is feasible. Instead, when the set of signals is large enough, then full surplus extraction becomes impossible. Our result confirms that when information acquisition is sufficiently flexible (in our case, fully flexible) then full surplus extraction seems impossible.

In this section, we took an *ex-ante* perspective to full surplus extraction, requiring that the mechanism extracts the maximal surplus that can be generated in the economy. Another approach would be to ask whether full extraction of *ex-post* surplus is possible: Does there exist a mechanism such that, in equilibrium, the ex-post efficient allocation is implemented and all the associated surplus is extracted from agents? Here the answer is always yes: the constant mechanism that always picks the efficient allocation at the prior $x^*(\mu_0)$ and has transfers equal to $t_i = \sum_\omega \mu_0(\omega) x_i^*(\mu_0) u_i(\omega_i)$ induces no information acquisition, and does extract all ex-post surplus in equilibrium. Even if we restrict attention to equilibria in which agents acquire some information, it seems to be always possible to find a mechanism extracting all ex-post surplus in equilibrium. This, however, is no guarantee on the magnitude of the surplus that is extracted by the seller: it can very well be that the generated surplus in equilibrium is very small.

# 6 CONCLUSION

In this paper, we investigate players' *informational incentives* in mechanism design, namely how the choice of the mechanism impacts what information players acquire in equilibrium. A mechanism is informationally simple if players have no incentives to acquire information on others' preferences. Our main result is that, for any smooth technology of information acquisition satisfying an Inada condition, a mechanism is Informationally Simple if and only if it is *de facto* separable. Separability means that agents' report cannot interact with one another in the allocation function, which rules out many economically meaningful mechanisms, and in particular all standard auction formats. This result holds generically, that is for an open set of preferences that has full measure. The intuition is that the outcomes a player can bring about in a mechanism depend on others' report, which makes it optimal to acquire information on them before investing in information acquisition on her own preferences.

This result has two implications for mechanism design. First, we show that a mechanism is ex-ante dominance solvable only if it has an informationally simple equilibrium, hence only if the mechanism is *de facto* separable. This points to a limitation of strategy-proofness as a concept of strategic simplicity. Indeed, even interim strategy-proof mechanisms incentivize players to acquire information about others and to best respond to beliefs about opponents' play at the ex-ante stage. Second, our result suggests that the independent private value assumption is unlikely to arise endogenously. This, however, does not mean full surplus extraction is possible using side bets as in Crémer and McLean (1988), as these would distort players' incentives when acquiring information.

There are several avenues for future research. One source of information acquisition that we ignored is communication among players. On one hand, our result suggests that some players would benefit from information aggregation in a communication stage after the information acquisition stage, as players endogenously hold information relevant to others. On the other hand, adding such a communication stage would modify informational incentives and free-riding may arise in the information acquisition stage. This raises an interesting question: Under what conditions does communication facilitate implementation and would arise endogenously from a coalition of players?

These considerations suggest that informational incentives may have important and concrete implications for the design of institutions—many of which remain unexplored to this day.

# APPENDIX A PROOFS

*Proof of Theorem 1.* We start by the proof of necessity, which is more involved than that of sufficiency. The general proof does not leverage Assumption 4, and easily extends to more general settings than allocation problems with a single good. There however exists a much more straightforward proof of necessity if we impose Assumption 4, and we give this shorter proof at the end.

**Proof of Necessity.** By contradiction, suppose that there exists an IS equilibrium $P^*$ of $\Gamma$ but $\Gamma$ is not de facto separable under $P^*$.

First we show that, because $\Gamma$ is not de facto separable, at least two players must acquire information in equilibrium. By definition, if $\Gamma$ is not de facto separable, then there exist no mechanism $\widehat{\Gamma}$ that is separable and induces the same state-dependent outcome as $\Gamma$. In particular, the direct revelation mechanism $\widehat{\Gamma}$ associated with equilibrium $P^*$ of $\Gamma$ is non-separable. Let $M_i^* \equiv \{m_i \mid \sum_\omega P_i^*(m_i|\omega) > 0\}$ be the set of messages $i$ sends with positive probability in equilibrium and $\mu_i^{m_i}(\cdot) \equiv (P_i(m_i|\cdot)\mu_0(\cdot))/(\sum_{\omega'} P_i^*(m_i|\omega'))$ her belief when she sends $m_i$. The direct revelation mechanism asks agents to report their equilibrium beliefs $\widehat{M}_i \equiv \{\mu_i^{m_i} | m_i \in M_i^*\}$ and implements the same outcome as $\Gamma$: $\hat{x}(\mu_i^{m_i}, \mu_{-i}^{m_{-i}}) = x(m_i, m_{-i})$.[27] By construction, there exists an equilibrium $\widehat{P}$ of $\widehat{\Gamma}$ that replicates $P^*$ in the following sense: $\operatorname{supp} \widehat{P}_i^* = \widehat{M}_i$ and $\widehat{P}_i^*(\mu_i^{m_i}|\omega) = P_i^*(m_i|\omega)$ for all $m_i, \omega$. Hence if $P^*$ is Informationally Simple, then so is $\widehat{P}^*$.

For this direct revelation mechanism not to be separable, there must exist an agent $i$ such that $|\widehat{M}_i| \geq 2$ and $|\widehat{M}_{-i}| \geq 2$, and

$$\hat{x}_i(\mu_i^{m_i}, \mu_{-i}^{m_{-i}}) - \hat{x}_i(\mu_i^{m_i'}, \mu_{-i}^{m_{-i}}) \neq \hat{x}_i(\mu_i^{m_i}, \mu_{-i}^{m_{-i}'}) - \hat{x}_i(\mu_i^{m_i'}, \mu_{-i}^{m_{-i}'})$$

for some $\mu_i^{m_i}, \mu_i^{m_i'} \in \widehat{M}_i$, $\mu_{-i}^{m_{-i}}, \mu_{-i}^{m_{-i}'} \in \widehat{M}_{-i}$. This has several implications. First, it must be that $\mu_i^{m_i} \neq \mu_i^{m_i'}$, i.e. that agent $i$'s belief when reporting $m_i$ is different from her belief when reporting $m_i'$ in equilibrium $P_i^*$. This means $P_i^*(m_i|\cdot) \neq P_i^*(m_i'|\cdot)$ and ensures that $i$ does acquire some information in equilibrium. Similarly, it must be that $\mu_{-i}^{m_{-i}} \neq \mu_{-i}^{m_{-i}'}$ and hence that $P_{-i}^*(m_{-i}|\cdot) \neq P_{-i}^*(m_{-i}'|\cdot)$. This ensures that other agents also acquire information, and that the way they do so impacts how much agent $i$ can influence the

[27] It can be that $\mu_i^{m_i} = \mu_i^{m_i'}$ for some $m_i', m_i \in M_i^*$. This is the case if $i$ randomizes over several messages at some posterior belief. In that case, the direct revelation mechanism implements the same randomization:

$$\hat{x}(\mu_i^{m_i}, \mu_{-i}^{m_{-i}}) = \sum_{m_i' \in M_i^*(m_i)} \frac{P_i^*(m_i')}{\sum_{m_i'' \in M_i^*(m_i)} P_i^*(m_i'')} x(m_i', m_{-i}),$$

where $M_i^*(m_i) = \{m_i' \in M_i^* \mid \mu_i^{m_i'} = \mu_i^{m_i}\}$ is the set of messages that $i$ sends with positive probability at posterior $\mu_i^{m_i}$.

outcome. From now on, we focus on the incentives of this particular agent $i$ and work with the direct revelation mechanism.

Second, we show that, for almost all preferences of $i$ in $\mathcal{U}_i$, $i$'s optimal strategy is not informationally simple. That is, generically, there does not exist transfers $\hat{t}_i \in \mathbb{R}^{\widehat{M}}$ such that the strategy $\widehat{P}_i^*$ that solves $i$'s system of FOCs ($\star$) is informationally simple. Since what matters for agent $i$ is how her preferences compare from one state to another, we fix agent $i$'s preferences in some arbitrarily chosen state $u_i(\omega_i^0)$ and show that for almost all $(u_i(\omega_i))_{\omega_i \neq \omega_i^0} \in \mathcal{U}_i^{-u_i(\omega_i^0)} \equiv \{(u_i(\omega_i))_{\omega_i \neq \omega_i^0} | (u_i(\omega_i^0), (u_i(\omega_i))_{\omega_i \neq \omega_i^0}) \in \mathcal{U}_i\}$, $i$'s optimal strategy is not IS. To do so, consider the FOCs ($\star$) corresponding to agent $i$ and messages in $\widehat{M}_i$. Since we know that these messages are sent with positive probability in equilibrium, we can ignore the non-negativity constraints on equilibrium probabilities. By the Inada condition we furthermore know that the equilibrium stochastic choice rule $\widehat{P}_i^*$ must be interior, and hence that these FOCs must hold with equality. Note that the endogenous variables in the FOCs are not only the agent's choice rule $\widehat{P}_i$ but also the Lagrange multipliers $\gamma_i$. To avoid carrying the multipliers around in the analysis, we substitute them out by choosing an arbitrarily message $m_i^0 \in \widehat{M}_i$, and subtracting the FOC for message $m_i^0$ to the FOCs for messages $m_i \in \widehat{M}_i \setminus \{m_i^0\}$. Agent $i$'s FOCs then write

($\star$)

$$\mathbb{E}_{\widehat{P}^*_{-i}(\cdot|\omega)}\Big[(\hat{x}_i(m_i, m_{-i}) - \hat{x}_i(m_i^0, m_{-i}))u_i(\omega_i) - (\hat{t}_i(m_i, m_{-i}) - \hat{t}_i(m_i^0, m_{-i}))\Big] - \frac{\partial c(\widehat{P}_i)}{\partial \widehat{P}_i(m_i|\omega_i)\mu_0(\omega)} + \frac{\partial c(\widehat{P}_i)}{\partial \widehat{P}_i(m_i^0|\omega_i)\mu_0(\omega)} = 0,$$

for all $\omega$, all $m_i \in \widehat{M}_i \setminus \{m_i^0\}$.[28]

One can think of a choice rule $\widehat{P}^*_{-i}$ as a matrix $[\widehat{P}^*_{-i}(m_{-i}|\omega)]_{(m_{-i},\omega)}$ where a row corresponds to the probability of sending a particular message $m_{-i} \in \widehat{M}_{-i}$

[28]We denote a generic element of $\widehat{M}_i$ by $m_i$ (and not $\mu_i^{m_i}$ as in the first part of the proof) to keep the notation uncluttered. Since we are working with the direct revelation mechanism, a message $m_i \in \widehat{M}_i$ is a belief.

in each state. Since we know that some $j \neq i$ acquires some information in equilibrium, and hence has at least two possible posterior beliefs, it must be that rank $\widehat{P}^*_{-i} \geq 2$. Furthermore, given that the equilibrium is IS, it has to be that rank $\widehat{P}^*_{-i} \leq |\Omega_{-i}|$. It might however be that rank $\widehat{P}^*_{-i} < |\Omega_{-i}|$, if for instance some $j$ acquires no information about $\omega_j$. If this is the case, there must exist $\Omega^*_{-i} \subset \Omega_{-i}$ such that $|\Omega^*_{-i}| =$ rank $\widehat{P}^*_{-i}$, and for all $\omega_{-i} \notin \Omega^*_{-i}$ there exist $(\alpha(\omega'_{-i}))_{\omega'_{-i} \in \Omega^*_{-i}}$ such that

$$\widehat{P}^*_{-i}(\cdot|\omega_{-i}) = \sum_{\omega'_{-i} \in \Omega^*_{-i}} \alpha(\omega'_{-i}) \widehat{P}^*_{-i}(\cdot|\omega'_{-i}).$$

Intuitively, $\Omega^*_{-i}$ is the relevant subset of states over which others' equilibrium strategy vary—i.e., that others' strategy can span. Restrict attention to the FOCs associated with some state of others $\omega_{-i} \in \Omega^*_{-i}$—if no IS solution exists to that restricted system of equations, then no solution exists to the overall system.

The designer has full flexibility in designing the transfers $\hat{t}_i$, but we show that this is not enough to ensure that the solution $\widehat{P}^*_i$ to the system of FOCs ($\star$) is IS. Note that transfers enter agent $i$'s FOCs only through

$$\mathbb{E}_{\widehat{P}^*_{-i}(\cdot|\omega_{-i})}\left[(\hat{t}_i(m_i, m_{-i}) - \hat{t}_i(m^0_i, m_{-i}))\right] \equiv \widehat{T}_i(m_i, \omega_{-i}).$$

This highlights two things. First, we can normalize $i$'s transfers associated with one particular message, for instance $\hat{t}_i(m^0_i, \cdot)$, as only the relative payoff between sending one message instead of another matters for $i$'s optimal strategy. Second, by tweaking the transfers, the designer can never make them depend more on $\omega_{-i}$ than $\widehat{P}^*_{-i}$, since they only depend on others' preferences through their equilibrium strategy. Hence the vector of (expected) transfers is effectively an element of $\mathbb{R}^{(|\widehat{M}_i|-1)\times|\Omega^*_{-i}|} \equiv \mathcal{T}$.

Let $\mathcal{P} \equiv (\Delta\widehat{M}_i)^{\Omega_i}$ and define $\Phi : \mathcal{P} \times \mathcal{T} \times \mathcal{U}_i^{-u_i(\omega^0_i)} \longrightarrow \mathbb{R}^{(|\widehat{M}_i|-1)\times|\Omega_i|\times|\Omega^*_{-i}|}$ as the function that maps stochastic choice rules with support $\widehat{M}_i$, and transfers

$(\widehat{P}_i, \widehat{T}_i) \in \mathcal{P} \times \mathcal{T}$ together with preferences $u_i \in \mathcal{U}_i^{-\omega_i^0}$ to the following vector:

$$\mathbb{E}_{\widehat{P}^*_{-i}(\cdot|\omega_{-i})}\Big[(\hat{x}_i(m_i, m_{-i}) - \hat{x}_i(m_i^0, m_{-i}))u_i(\omega_i)\Big] - \widehat{T}_i(m_i, \omega_{-i}) \\ - \frac{\partial c(\widehat{P}_i)}{\partial \widehat{P}_i(m_i|\omega_i)\mu_0(\omega)} + \frac{\partial c(\widehat{P}_i)}{\partial \widehat{P}_i(m_i^0|\omega_i)\mu_0(\omega)}$$

for all $m_i \in \widehat{M_i}\backslash\{m_i^0\}, \omega_i \in \Omega_i, \omega_{-i} \in \Omega^*_{-i}$. Importantly, $i$'s stochastic choice rule is informationally simple by assumption, and thus belongs to $\mathbb{R}^{(|\widehat{M}_i|-1)\times|\Omega_i|}$. Therefore we have

$$\dim \mathcal{P} \times \mathcal{T} = (|\widehat{M}_i| - 1) \times (|\Omega_i| + |\Omega^*_{-i}|).$$

The FOCs for agent $i$ can be written as $\Phi(\widehat{P}_i^*, \widehat{T}_i; u_i) = \mathbf{0}$. Hence, the set of IS stochastic choice rules (together with transfers) which solve the agent's FOCs is $\Phi^{-1}(\mathbf{0}; u_i)$. We show that this set is a manifold of negative dimension, and hence is empty, for almost all $u_i \in \mathcal{U}_i^{-u_i(\omega_i^0)}$. Since this is true irrespective of the normalization we choose for $u_i(\omega_i^0)$ and because $\mathcal{U}_i = \bigcup_{\omega_i^0} u_i(\omega_i^0) \times \mathcal{U}_i^{-u_i(\omega_i^0)}$, this implies that there exists no IS solution to $i$'s system of FOCs for almost all preferences in $i$'s overall set of possible preferences $\mathcal{U}_i$. This is done by successively applying the Transversality theorem (to show that the non-linear equations in this system are locally linearly independent at $\mathbf{0}$ for almost all $u_i$), and the Regular Value theorem (to show that the solution set is a manifold of negative dimension).[29]

In order to apply the Transversality theorem we need to show that $\mathbf{0}$ is a regular value of $\Phi$, i.e. that the Jacobian of $\Phi$ at $\mathbf{0}$ has full rank: $\Phi(\widehat{P}_i, \widehat{T}_i; u_i) = \mathbf{0} \implies \operatorname{rank} D\Phi(\widehat{P}_i, \widehat{T}_i; u_i) = \min\{(|\widehat{M}_i| - 1) \times |\Omega_i| \times |\Omega^*_{-i}|, \dim \mathcal{P} \times \mathcal{T} + \dim \mathcal{U}_i^{-u_i(\omega_i^0)}\}$ where $D$ is the Jacobian. Intuitively, this is equivalent to showing that the number of locally linearly independent equations of the system evaluated at $\mathbf{0}$ is maximal. Note that $D\Phi$ has $(|\widehat{M}_i| - 1) \times |\Omega_i| \times |\Omega^*_{-i}|$ rows—one for each FOC, so one for each $m_i \in \widehat{M}_i\backslash\{m_i^0\}$, $\omega_i \in \Omega_i$, and $\omega_{-i} \in \Omega^*_{-i}$—and

[29] Mas-Colell (1989) Chapter 1 (section H) and especially Chapter 8 provide an introduction to differential topology. A formal statement of the results we use here can be found on page 320.

$\dim \mathcal{P} \times \mathcal{T} + \dim \mathcal{U}_i^{-u_i(\omega_i^0)}$ columns—each corresponding to the derivative of $\Phi$ with respect to one element of $(\widehat{P}_i, \widehat{T}_i; u_i)$. We show that the columns of $D\Phi$ are linearly independent.

The Jacobian of $\Phi$ has some simplifying structure, as many of its entries are zero. First, the columns associated with the derivatives w.r.t. $\widehat{P}_i$ correspond to the Hessian of the cost of information, as $\widehat{P}_i$ only enter the FOCs through the marginal cost:[30]

$$D_{\widehat{P}_i}\Phi = \left[\frac{\partial^2 c(\widehat{P}_i)}{\partial \widehat{P}_i(m_i^0|\omega)\mu_0(\omega)\partial \widehat{P}_i(m_i'|\omega')\mu_0(\omega')} - \frac{\partial^2 c(\widehat{P}_i)}{\partial \widehat{P}_i(m_i|\omega)\mu_0(\omega)\partial \widehat{P}_i(m_i'|\omega')\mu_0(\omega')}\right]_{((m_i,\omega),(m_i',\omega'))}.$$

Second, since each $\widehat{T}_i(m_i, \omega_{-i})$ only enters the FOCs of agent $i$ associated with sending message $m_i$ when others' fundamentals are $\omega_{-i}$, the columns associated with the derivatives w.r.t. $\widehat{T}_i$ form a block diagonal matrix with each block corresponding to one message $m_i$ for agent $i$ and one state of others $\omega_{-i}$:

$$D_{\widehat{T}_i}\Phi = \begin{bmatrix} B_{\widehat{T}_i}(m_i, \omega_{-i}) & \mathbf{0} & & \dots & & \\ & \mathbf{0} & \ddots & & & \\ & \vdots & & B_{\widehat{T}_i}(m_i', \omega_{-i}) & \mathbf{0} & \\ & & & \mathbf{0} & \ddots & \end{bmatrix}$$

Each block $B_{\widehat{T}_i}(m_i, \omega_{-i})$ is simply a $|\Omega_i|$-vector of ones. Indeed, each row corresponds to a possible state $\omega_i \in \Omega_i$ and the derivative of agent $i$'s FOC w.r.t. $\widehat{T}_i(m_i, \omega_{-i})$ is one.

Similarly, since each $u_i(\omega_i)$ only enters the FOCs of agent $i$ in state $\omega_i$, the columns associated with its derivative have non-zero entries only for rows that correspond to FOCs in state $\omega_i$. For these rows, the derivative equal

$$\sum_{m_{-i} \in \widehat{M}_{-i}} \widehat{P}^*_{-i}(m_{-i} \mid \omega_{-i})(\hat{x}_i(m_i, m_{-i}) - \hat{x}_i(m_i^0, m_{-i})).$$

[30] $D_{\widehat{P}_i}$ denotes the restriction of the Jacobian corresponding to the derivative w.r.t. $P_i$.

We first argue that the columns of $D_{\widehat{P}_i,\widehat{T}_i}$ are linearly independent. Note that the columns corresponding to derivatives w.r.t. $\widehat{T}_i$ are independent of $\omega_i$, and thus are constant across rows that differ only by $\omega_i$. On the contrary, $\partial^2 c(\widehat{P}_i^*)/\partial \widehat{P}_i^*(m_i|\omega_i)\mu_0(\omega)^2$ varies with $\omega_i$ since $i$ acquires information in equilibrium, and it is thus impossible to express the first sets of columns (corresponding to derivatives w.r.t. $\widehat{P}_i$) in terms of the second (corresponding to derivatives w.r.t. $\widehat{T}_i$). Furthermore, the columns corresponding to derivatives w.r.t. $\widehat{T}_i$ are also linearly independent since they form a block-diagonal matrix, with each block being composed of a single column.

We now show that the columns of $D_{u_i}$ are linearly independent from those of $D_{\widehat{P}_i,\widehat{T}_i}$. Using a similar argument as above, derivatives w.r.t. $u_i$ must be linearly independent from those w.r.t. $\widehat{P}_i$ as the former depend on $\omega_{-i}$ whereas the latter do not. Indeed, and as discussed above, the fact that $\widehat{\Gamma}$ is not de facto non-separable implies $\sum_{m_{-i}\in\widehat{M}_{-i}} \widehat{P}_{-i}^*(m_{-i} \mid \omega_{-i})(\hat{x}_i(m_i, m_{-i})-\hat{x}_i(m_i^0, m_{-i}))$ must vary with $\omega_{-i} \in \Omega_{-i}^*$. The main thing to prove is that the columns of $D_{u_i}$ are linearly independent from the columns corresponding to derivatives w.r.t. $\widehat{T}_i$. Recall that only $(u_i(\omega_i))_{\omega_i\neq\omega_i^0}$ are parameters here, as $u_i(\omega_i^0)$ is normalized to some fixed and arbitrary value. Hence all rows corresponding to FOCs in state $\omega_i^0$ must have zero entries in $D_{u_i}\Phi$. All other entries equal $\sum_{m_{-i}\in\widehat{M}_{-i}} \widehat{P}_{-i}^*(m_{-i} \mid \omega_{-i})(\hat{x}_i(m_i, m_{-i}) - \hat{x}_i(m_i^0, m_{-i}))$, and could be replicated using the $D_{\widehat{T}_i}\Phi$ columns by weighting by $\sum_{m_{-i}\in\widehat{M}_{-i}} \widehat{P}_{-i}^*(m_{-i} \mid \omega_{-i})(\hat{x}_i(m_i, m_{-i})-\hat{x}_i(m_i^0, m_{-i}))$ the column corresponding to the derivative w.r.t. $\widehat{T}_i(m_i, \omega_{-i})$. However, this would need to generate a zero entry for state $\omega_i^0$ which is possible only if $\sum_{m_{-i}\in\widehat{M}_{-i}} \widehat{P}_{-i}^*(m_{-i} \mid \omega_{-i})(\hat{x}_i(m_i, m_{-i}) - \hat{x}_i(m_i^0, m_{-i})) = 0$ for all $m_i$, which cannot be true in a non-separable mechanism.[31]

Thus, if $\widehat{\Gamma}$ is not de facto separable, $D\Phi(\widehat{P}_i, \widehat{T}_i; u_i)$ has full rank and $\mathbf{0}$ is a regular value of $\Phi$. The Parametric Transversality theorem states that, except for a nullset $\overline{\mathcal{U}}_i^{-u_i(\omega_i^0)} \subset \mathcal{U}_i^{-u_i(\omega_i^0)}$ of preferences, $\mathbf{0}$ is a regular value of $\Phi(\cdot; u_i)$. Then by the Regular Value theorem, $\Phi^{-1}(\mathbf{0}; u_i)$ is a smooth manifold of dimen-

[31] Indeed, if that were true, then $i$'s message would effectively have no effect on her outcome in equilibrium, and she could not find it optimal to acquire information.

sion

$$\dim \Phi^{-1}(\mathbf{0}; u_i) = (|\widehat{M}_i| - 1) \times (|\Omega_i| + |\Omega^*_{-i}|) - (|\widehat{M}_i| - 1) \times |\Omega_i| \times |\Omega^*_{-i}| < 0$$

whenever $|\Omega_i| > 2$. Therefore we conclude that, for a full measure set of preferences $\mathcal{U}_i^{-u_i(\omega_i^0)} \setminus \overline{\mathcal{U}}_i^{-u_i(\omega_i^0)}$, the set of IS stochastic choice rules (together with transfers) solving the FOCs is empty. Let $\mathcal{U}_i^0 \equiv \cup_{\omega_i^0} u_i(\omega_i^0) \times \overline{\mathcal{U}}_i^{-u_i(\omega_i^0)}$ be the overall set of preferences for $i$ for which there is an IS solution to $i$'s system of FOCs. Since $\overline{\mathcal{U}}_i^{-u_i(\omega_i^0)}$ has Lebesgue measure zero for each possible normalization of $u_i(\omega_i^0)$, the overall set of preferences $\mathcal{U}_i^0 \subset \mathcal{U}_i$ for which $i$ has an informationally simple optimal strategy is null as well.

We have left to show that $\mathcal{U}_i^0$ is closed, or equivalently that $\mathcal{U}_i \setminus \mathcal{U}_i^0$ is open. Take any $u_i \in \mathcal{U}_i \setminus \mathcal{U}_i^0$. By definition, for these preferences, there does not exist transfers that make $i$'s optimal strategy Informationally Simple. That is, there does not exist $(\widehat{P}_i, \widehat{T}_i)$ such that $\Phi(\widehat{P}_i, \widehat{T}_i; u_i) = 0$. Let $\|\cdot\|$ denote the Euclidean distance, and note that the minimum of $\|\Phi(\cdot; u_i)\|$ is reached for some $(\widehat{P}_i, \widehat{T}_i) \in \mathcal{P} \times \mathcal{T}$. Indeed, any large enough $\widehat{T}_i$ or boundary choice rule $\widehat{P}_i$ sends $\|\Phi(\cdot; u_i)\|$ to infinity, and so we can restrict attention to a compact subset of $\mathcal{P} \times \mathcal{T}$ to find a minimizer of $\|\Phi(\cdot; u_i)\|$. Since $\|\Phi(\cdot; u_i)\|$ is continuous on such compact subset, it must reach a minimum. Let $\delta \equiv \min_{(\widehat{P}_i, \widehat{T}_i)} \|\Phi(\widehat{P}_i, \widehat{T}_i; u_i)\|$, with $\delta > 0$ by assumption. Take any $\varepsilon \in (0, \delta(|\widehat{M}_i|)^{-1/2})$, and consider any $u_i' \in \mathcal{U}_i$ such that $\|u_i - u_i'\| < \varepsilon$. Then, for any $(\widehat{P}_i, \widehat{T}_i)$,

$$\begin{aligned}
&\|\Phi(\widehat{P}_i, \widehat{T}_i; u_i) - \Phi(\widehat{P}_i, \widehat{T}_i; u_i')\| \\
&\qquad = \left( \sum_{\omega, m_i} (\mathbb{E}_{\widehat{P}^*_{-i}(\cdot|\omega)}[(\hat{x}_i(m_i, m_{-i}) - \hat{x}_i(m_i^0, m_{-i}))(u_i(\omega_i) - u_i'(\omega_i))])^2 \right)^{\frac{1}{2}} \\
&\qquad = \left( \sum_{\omega} \left( \sum_{m_i} \mathbb{E}_{\widehat{P}^*_{-i}(\cdot|\omega)}[\hat{x}_i(m_i, m_{-i}) - \hat{x}_i(m_i^0, m_{-i})]^2 \right) (u_i(\omega_i) - u_i'(\omega_i))^2 \right)^{\frac{1}{2}} \\
&\qquad \leq \sqrt{|\widehat{M}_i|} \left( \sum_{\omega} (u_i(\omega_i) - u_i'(\omega_i))^2 \right)^{\frac{1}{2}} < \sqrt{|\widehat{M}_i|}\varepsilon,
\end{aligned}$$

where the inequality follows from $|\mathbb{E}_{\widehat{P}^*_{-i}(\cdot|\omega)}[\hat{x}_i(m_i, m_{-i}) - \hat{x}_i(m_i^0, m_{-i})]| \leq 1$, and thus $\sum_{m_i} \mathbb{E}_{\widehat{P}^*_{-i}(\cdot|\omega)}[\hat{x}_i(m_i, m_{-i}) - \hat{x}_i(m_i^0, m_{-i})]^2 \leq |\widehat{M_i}|$. From the reverse triangle inequality, we know that $|\|\Phi(\widehat{P}_i, \widehat{T}_i; u_i)\| - \|\Phi(\widehat{P}_i, \widehat{T}_i; u_i')\|| \leq \|\Phi(\widehat{P}_i, \widehat{T}_i; u_i) - \Phi(\widehat{P}_i, \widehat{T}_i; u_i')\| < \sqrt{|\widehat{M_i}|}\varepsilon < \delta$, where the last inequality comes from the definition of $\varepsilon$. This rewrites as $|\delta - \|\Phi(\widehat{P}_i, \widehat{T}_i; u_i')\|| < \delta$, which implies $\|\Phi(\widehat{P}_i, \widehat{T}_i; u_i')\| > 0$ and $u_i' \in \mathcal{U}_i \backslash \mathcal{U}_i^0$. Hence $\mathcal{U}_i \backslash \mathcal{U}_i^0$ is open, and the system of FOCs for $i$ has an IS solution only for a set of preferences $\mathcal{U}_i^0$ whose closure has Lebesgue measure zero.

**Proof of Necessity with Assumption 4.** Suppose that Assumption 4 holds (as in the entropic case), and consider again the system of FOCs ($\star$). Under Assumption 4, the marginal cost

$$\frac{\partial c(\widehat{P}_i)}{\partial \widehat{P}_i(m_i|\omega_i)\mu_0(\omega)}$$

must be independent of $\omega_{-i}$ at an IS choice rule $\widehat{P}_i(\cdot|\omega_i)$.[32] If that were not the case, an optimal choice rule might still depend on $\omega_{-i}$ even if an agent's gross payoff does not, neither directly nor indirectly.

Subtracting the FOC associated with message $m_i$ in state $(\omega_i, \omega_{-i})$ from the one associated with same message $m_i$ but in state $(\omega_i, \omega_{-i}')$ yields

$$\begin{aligned} &\mathbb{E}_{\widehat{P}^*_{-i}(\cdot|\omega_{-i}) - \widehat{P}^*_{-i}(\cdot|\omega_{-i}')} \left[\hat{x}_i(m_i, m_{-i}) - \hat{x}_i(m_i^0, m_{-i})\right] u_i(\omega_i) \\ &\qquad - \mathbb{E}_{\widehat{P}^*_{-i}(\cdot|\omega_{-i}) - \widehat{P}^*_{-i}(\cdot|\omega_{-i}')} \left[\hat{t}_i(m_i, m_{-i}) - \hat{t}_i(m_i^0, m_{-i})\right] = 0. \end{aligned}$$

Now consider the same equation but associated with state $\omega_i'$. Subtracting it

[32] For instance, under the entropic cost,

$$\frac{\partial c(\widehat{P}_i)}{\partial \widehat{P}_i(m_i|\omega_i)\mu_0(\omega)} = \log \left[\frac{\widehat{P}_i(m_i|\omega_i)}{\sum_{\omega_i'} \widehat{P}_i(m_i|\omega_i')}\right].$$

from the previous expression yields

$$\mathbb{E}_{\widehat{P}^*_{-i}(\cdot|\omega_{-i})-\widehat{P}^*_{-i}(\cdot|\omega'_{-i})}\left[(\hat{x}_i(m_i,m_{-i})-\hat{x}_i(m_i^0,m_{-i}))\right](u_i(\omega_i)-u_i(\omega'_i))=0.$$

Since the mechanism is not de facto separable, we know that the first term is different from zero for at least one message $m_i$. Hence the solution to agent $i$'s system of FOCs is IS only if $u_i(\omega_i)$ is constant.

**Proof of Sufficiency.** We now prove that if both the outcome and transfer functions of $\Gamma$ are separable, then all equilibria of $\Gamma$ are Informationally Simple. Outcome and transfer functions being separable means that the way an agent $i$ impacts her outcome/transfer only depends on her messages. Formally, there exist mappings $X_i : M_i \times M_i \longrightarrow [-1,1]$ and $T_i : M_i \times M_i \longrightarrow \mathbb{R}$ for all $i$ such that

$$\begin{aligned} x_i(m_i,m_{-i})-x_i(m'_i,m_{-i}) &= X_i(m_i,m'_i) \\ t_i(m_i,m_{-i})-t_i(m'_i,m_{-i}) &= T_i(m_i,m'_i) \end{aligned}$$

for all $m_{-i}$. Consider some agent $i$, who takes as given others' strategy $P^*_{-i}$. Her objective is

$$\sum_{\omega\in\Omega}\mu_0(\omega)\mathbb{E}_{P_i(\cdot|\omega),P^*_{-i}(\cdot|\omega)}\Big[x_i(m_i,m_{-i})u_i(\omega_i)-t_i(m_i,m_{-i})\Big]-c(P_i).$$

Since in each state $\omega$ her choice rule must sum to one $\sum_{m_i}P_i(m_i|\omega)=1$, we can normalize agent $i$'s utility by her expected utility from sending some arbitrarily chosen message $m_i^0\in M_i$:

$$\begin{aligned} &\sum_{\omega\in\Omega}\mu_0(\omega)\mathbb{E}_{P_i,P^*_{-i}}\Big[(x_i(m_i,m_{-i})-x_i(m_i^0,m_{-i}))u_i(\omega_i)-(t_i(m_i,m_{-i})-t_i(m_i^0,m_{-i}))\Big] \\ &\qquad\qquad+\sum_{\omega\in\Omega}\mu_0(\omega)\mathbb{E}_{P^*_{-i}}\Big[x_i(m_i^0,m_{-i})u_i(\omega_i)-t_i(m_i^0,m_{-i})\Big]-c(P_i). \end{aligned}$$

So what matters for agent $i$ is the relative payoff she gets under the different messages she can send. Since the mechanism is separable, her objective can be equivalently expressed as

$$\sum_{\omega \in \Omega} \mu_0(\omega) \mathbb{E}_{P_i(\cdot|\omega)} \left[ X_i(m_i, m_i^0) u_i(\omega_i) - T_i(m_i, m_i^0) \right] - c(P_i).$$

This formulation makes it clear that the relative value agent $i$ gets from sending message $m_i$ in state $\omega = (\omega_i, \omega_{-i})$ only depends on $\omega_i$ and not on $\omega_{-i}$. By Assumption 4, agent $i$'s optimal choice rule must be independent of payoff-irrelevant states: $P_i^*(\cdot|\omega_i, \omega_{-i}) = P_i^*(\cdot|\omega_i, \omega'_{-i})$ for all $\omega_i$, $\omega_{-i}$, $\omega'_{-i}$. This holds for all agents, and thus all equilibria of $\Gamma$ must be informationally simple. □

*Proof of Proposition 1.* Let $\mathcal{U}_{IS}(\kappa) \subseteq \bigtimes_i \mathcal{U}_i$ be the set of utility functions for which (i) there exists an IS equilibrium $P^*$ of $\Gamma$, and (ii) $\Gamma$ is not de facto separable under $P^*$.

The case with $\kappa = 0$ is the baseline case considered in this paper, for which Theorem 1 applies: $\mathcal{U}_{IS}(0)$ has Lebesgue measure zero, hence $\rho(0) = 0$.

To show that $\rho$ is increasing, we prove that for any $\kappa$, $\kappa'$ with $\kappa' \geq \kappa$, $\mathcal{U}_{IS}(\kappa) \subseteq \mathcal{U}_{IS}(\kappa')$. Take any $u \in \mathcal{U}_{IS}(\kappa)$. By definition, we know that there exists a non-separable IS equilibrium $P^*$. We need to show that $P^*$ remains an equilibrium if we increase the fixed cost from $\kappa$ to $\kappa'$. Now that we have introduced a discontinuity in the objective function of agents, the FOCs ($\star$) are not sufficient to characterize an equilibrium. There are two possible types of equilibrium strategies for an agent: either she learns about others or not. If she does, then her strategy must satisfy ($\star$). If she does not, then her IS strategy must solve:

$$\mathbb{E}_{\mu_0^{-i}} \left( \mathbb{E}_{P^*_{-i}(\cdot|\omega_{-i})} \Big[ x_i(m_i, m_{-i}) u_i(\omega_i) - t_i(m_i, m_{-i}) \Big] + \frac{\gamma_i(\omega)}{\mu_0(\omega)} - \frac{\partial c(P_i^*)}{\partial P_i^*(m_i|\omega)\mu_0(\omega)} \right) = 0$$

for all $\omega_i$ and all $m_i \in \operatorname{supp} P_i^*$. These two sets of FOCs define two possible equilibrium strategies for agent $i$, yielding two different expected payoffs. In equilibrium, agent $i$ learns about others only if the gap between these two expected payoffs $\Delta(u)$ more than compensate the fixed cost $\kappa$. Since $P^*$ is informationally simple by assumption, we know that this gap is lower than $\kappa$. It is hence also lower than $\kappa'$, and $P^*$ remains a equilibrium under $\kappa'$: $u \in \mathcal{U}_{IS}(\kappa')$, for all $u \in \mathcal{U}_{IS}(\kappa)$.

Finally, we show that $\rho$ is continuous in $\kappa$. By contradiction, suppose it is not: there exists $\kappa^*$ and $\delta > 0$ such that, for all $\varepsilon > 0$, either $\rho(\kappa^*) - \rho(\kappa^* - \varepsilon) > \delta$ or $\rho(\kappa^* + \varepsilon) - \rho(\kappa^*) > \delta$. Consider the latter case[33] — the function $\rho$ discontinuously jumps up at $\kappa^*$ — and pick any $\varepsilon > 0$. By assumption there is a difference of at least $\delta$ between the Lebesgue measure of $\mathcal{U}_{IS}(\kappa^* + \varepsilon)$ and that of $\mathcal{U}_{IS}(\kappa^*)$. Consider any $u \in \mathcal{U}_{IS}(\kappa^* + \varepsilon) \setminus \mathcal{U}_{IS}(\kappa^*)$. For these utility functions, there exists a non-separable IS equilibrium $P^*$ under $\kappa^* + \varepsilon$ but not under $\kappa^*$. Hence $\kappa^* < \Delta(u) < \kappa^* + \varepsilon$: for at least one agent $i$, it is worth learning about others given that they play $P^*_{-i}$ if the associated fixed cost is $\kappa^*$ but not if it is $\kappa^* + \varepsilon$. As $\varepsilon$ tends to zero, this means that any $u \in \mathcal{U}_{IS}(\kappa^* + \varepsilon) \setminus \mathcal{U}_{IS}(\kappa^*)$ must satisfy $\Delta(u) = \kappa^*$. This equality defines a manifold of dimension strictly less than $|\Omega_i|$ in the domain of $i$'s preferences, and hence the Lebesgue measure of the set of utility functions satisfying it is zero. This contradicts the assumption that the measure of $\mathcal{U}_{IS}(\kappa^* + \varepsilon) \setminus \mathcal{U}_{IS}(\kappa^*)$ must be above $\delta$ even for vanishing $\varepsilon$. $\square$

*Proof of Proposition 2.* Let $P^*$ be an equilibrium in dominant strategy of $\Gamma$. That means $P_i^*$ is an optimal strategy for agent $i$, irrespective of other agents' strategy:

$$P_i^* \in \arg\max_{P_i} \sum_{\omega} \mu_0(\omega) \mathbb{E}_{P_i(\cdot|\omega), P_{-i}(\cdot|\omega)}[x_i(m_i, m_{-i}) u_i(\omega_i) - t_i(m_i, m_{-i})] - c(P_i)$$

for all $P_{-i}$. In particular, $P_i^*$ is optimal when others' strategy is independent of the state, i.e. when $P_{-i}(\cdot|\omega) = P_{-i}(\cdot|\omega')$ for all $\omega, \omega'$. This requires

$$P_i^* \in \arg\max_{P_i} \sum_{\omega} \mu_0(\omega) \mathbb{E}_{P_i(\cdot|\omega)} \left( \mathbb{E}_{P_{-i}(\cdot)}[x_i(m_i, m_{-i})] u_i(\omega_i) - \mathbb{E}_{P_{-i}(\cdot)}[t_i(m_i, m_{-i})] \right) - c(P_i).$$

Note however that in such case, the value of reporting a particular message $m_i$ is state $\omega$ equals $\mathbb{E}_{P_{-i}(\cdot)}[x_i(m_i, m_{-i})] u_i(\omega_i) - \mathbb{E}_{P_{-i}(\cdot)}[t_i(m_i, m_{-i})]$, and is always independent of $\omega_{-i}$. Hence, by Assumption 4, agent $i$'s optimal choice rule does not depend on the payoff-irrelevant dimensions $\omega_{-i}$: $P_i^*$ is informa-

[33] The proof is similar for the other case.

tionally simple.

The same argument holds for all agents, and so if $P^*$ is an equilibrium in dominant strategy of $\Gamma$ then $P^*$ is informationally simple. $\square$

*Proof of Theorem 2.* The proof of Theorem 2 uses the same techniques as that of Theorem 1. We find necessary conditions for full surplus extraction that are non generic in the space of preferences. By assumption, there is at least one agent $i$ for whom it is efficient to acquire some information. From now on, we restrict attention to this agent $i$, and take as given that all others play their efficient strategy $P^\dagger_{-i}$. We show that it is generically impossible to induce $i$ to choose her efficient strategy $P^\dagger_i$ while extracting all surplus from her.

Agent $i$'s efficient strategy $P^\dagger_i$ solves

$$\max_{P_i \in \mathcal{P}_i} \sum_{\omega \in \Omega} \mu_0(\omega) \sum_{\mu_i \in \operatorname{supp} P_i} P_i(\mu_i|\omega) \mathbb{E}_{P^\dagger_{-i}(\cdot|\omega)} \left[ \sum_{j \in N} x^*_j(\mu) u_j(\omega_j) \right] - c(P_i).$$

Recall that beliefs $\mu_i$ in the support of a choice rule $P_i$ need to be consistent with the choice rule. Hence, a marginal change in $P_i(\mu_i|\omega)\mu_0(\omega)$ also marginally changes $\mu_i$. However such small change in belief has no effect on the ex post efficient allocation $x^*(\mu_i, \mu_{-i})$ at belief profiles that are part of the efficient strategy $(\mu_i, \mu_{-i}) \in \bigtimes_i \operatorname{supp} P^\dagger_i$. Indeed, for a marginal change in belief $\mu_i$ to change the efficient allocation $x^*(\mu_i, \mu_{-i})$, it has to be that at that belief profile, two agents have the exact same expected valuation for the good. But then some agent could acquire a bit less information—that is, we could garble that agent's choice rule with $\varepsilon$ noise—so as to strictly reduce her information cost without affecting overall gross welfare.

The FOC with respect to $P_i(\mu_i|\omega)\mu_0(\omega)$ then writes

$$(1) \qquad \mathbb{E}_{P^\dagger_{-i}(\cdot|\omega)} \left[ \sum_{j \in N} x^*_j(\mu_i, \mu_{-i}) u_j(\omega_j) \right] - \frac{\partial c(P_i)}{\partial P_i(\mu_i|\omega)\mu_0(\omega)} + \frac{\zeta_i(\omega)}{\mu_0(\omega)} = 0,$$

for all $\mu_i \in \operatorname{supp} P_i$ and for all $\omega \in \Omega$, where $\zeta_i(\omega)$ is the Lagrange multiplier associated with the constraint that $P_i(\cdot|\omega)$ sums to one. Hence agent $i$'s efficient stochastic choice rule $P^\dagger_i$ must satisfy the above system of equations, as well as the constraints that

$$(1') \qquad \sum_{\mu_i} P_i(\mu_i|\omega) - 1 = 0 \quad \forall \omega.$$

Now suppose there exists a mechanism that extracts full surplus. Such mechanism must lead each agent to acquire the efficient level of information $P_i^\dagger$. Without loss, this mechanism can be written down as a direct revelation mechanism in which agents report their information $M_i = \operatorname{supp} P_i^\dagger$, and the mechanism implements the efficient allocation given reported beliefs. The FOCs for the individual decision problem write

$$(2) \qquad \mathbb{E}_{P^\dagger_{-i}(\cdot|\omega)}\Big[x_i^*(\mu)u_i(\omega_i) - t_i(\mu_i,\mu_{-i})\Big] - \frac{\partial c(P_i)}{\partial P_i(\mu_i|\omega)\mu_0(\omega)} + \frac{\gamma_i(\omega)}{\mu_0(\omega)} = 0.$$

Surplus extraction requires that the information strategy chosen by the agent coincides with $P_i^\dagger$. Hence $P_i^\dagger$ must solve both (1) and (2). Subtracting (2) from (1) yields:

$$\sum_{\mu_{-i}} P^\dagger_{-i}(\mu_{-i}|\omega)t_i(\mu_i,\mu_{-i}) = -\sum_{\mu_{-i}} P^\dagger_{-i}(\mu_{-i}|\omega)\sum_{j\neq i} x_j^*(\mu)u_j(\omega_j) - \frac{\zeta_i(\omega)-\gamma_i(\omega)}{\mu_0(\omega)}$$

which implies:

$$(3) \quad \mathbb{E}_{\mu_0,P^\dagger_{-i}(\cdot|\omega)}\left[t_i(\mu_i,\mu_{-i})\right] = -\mathbb{E}_{\mu_0,P^\dagger_{-i}(\cdot|\omega)}\left[\sum_{j\neq i} x_j^*(\mu)u_j(\omega_j)\right] - \sum_{\omega\in\Omega}\big(\zeta_i(\omega)-\gamma_i(\omega)\big).$$

Hence efficient information acquisition requires a VCG mechanism, in which agent $i$'s expected transfer given her report $\mu_i$ equals the expected payoff that all other agents get when $i$ reports $\mu_i$. Since, with these transfers, the solutions to both systems of FOCs coincide and equal $P_i^\dagger$, the Lagrange multipliers also coincide: $\zeta_i(\omega) = \gamma_i(\omega)$ for all $\omega$. To extract full surplus from agent $i$, her expected transfer must sum to her net utility:

$$(4) \qquad \mathbb{E}_{\mu_0,P^\dagger(\cdot|\omega)}[t_i(\mu_i,\mu_{-i})] = \mathbb{E}_{\mu_0,P^\dagger(\cdot|\omega)}\left[x_i^*(\mu)u_i(\omega_i)\right] - c(P_i^\dagger).$$

Transfers must extract the agent's expected utility given her type while compensating her for the ex-ante investment in information acquisition. Not compensating for these costs would violate the ex-ante IR constraint. Combining (3) and (4), $(P_i^\dagger)_i$ must solve:

$$\mathbb{E}_{\mu_0, P^\dagger(\cdot|\omega)}\left[\sum_{j \in N} x_j^*(\mu) u_j(\omega_j)\right] = c(P_i^\dagger).$$

Finally taking expectations over $\mu_i$ and $\omega$ in equation (1) yields:

$$\mathbb{E}_{\mu_0, P^\dagger(\cdot|\omega)}\left[\sum_{j \in N} x_j^*(\mu) u_j(\omega_j)\right] = \mathbb{E}_{\mu_0, P_i^\dagger(\cdot|\omega)}\left[\frac{\partial c(P_i^\dagger)}{\partial P_i^\dagger(\mu_i|\omega)\mu_0(\omega)}\right] - \sum_{\omega \in \Omega} \zeta_i(\omega).$$

Combining the above two equations entails that $P_i^\dagger$ must solve:

$$\mathbb{E}_{\mu_0, P_i^\dagger(\cdot|\omega)}\left[\frac{\partial c(P_i^\dagger)}{\partial P_i^\dagger(\mu_i|\omega)\mu_0(\omega)}\right] - \sum_{\omega \in \Omega} \zeta_i(\omega) - c(P_i^\dagger) = 0. \tag{5}$$

That is, they together require that the total cost of information equals the expected marginal cost at the efficient solution. We show that this condition, however, is non-generic.

Let $\hat{\Phi}$ be the functional which maps a choice rule for $i$, Lagrange multipliers and preferences to the LHS of the system of equations (1) and constraints (1′), as well as to the LHS of equation (5). Hence the necessary conditions (1), (1′) and (5) for full extraction of agent $i$'s surplus are jointly written as $\hat{\Phi}(P_i, \zeta_i; u) = \mathbf{0}$. As in the proof of Theorem 1, we leverage the Transversality theorem and Regular Value theorem to show that the set $\hat{\Phi}^{-1}(\mathbf{0}; u)$ is empty for almost all $u \in \mathcal{U}$.

In order to apply the Transversality theorem we need to show that $\mathbf{0}$ is a regular value of $\hat{\Phi}$, i.e., that the number of locally linearly independent equations of the system evaluated at $\mathbf{0}$ is maximal:

$$\hat{\Phi}(P_i, \zeta_i; u) = \mathbf{0} \implies \operatorname{rank} D\hat{\Phi}(P_i, \zeta_i; u) = 1 + |\Omega| \times (|\operatorname{supp} P_i^\dagger| + 1),$$

where $D\hat{\Phi}(P, \zeta; u)$ is the Jacobian, and has as many rows as there are equations in the systems (1), (1′) and (5). We need to show that all its rows are linearly independent. The Jacobian has $|\Omega| \times (|\operatorname{supp} P_i^\dagger| + 1) + \sum_j |\Omega_j|$ columns, each corresponding to the derivative w.r.t. each element of $(P_i, \zeta_i; u)$. Ignoring the row that corresponds to equation (5) for now, it equals

$$\begin{array}{c|cccccccccc}
 & & \partial P_i & & \partial\zeta_i(\omega) & \partial\zeta_i(\omega') & \dots & \partial u_i(\omega) & \partial u_j(\omega) & \dots & \partial u_i(\omega') \\
\hline
(1)_{\omega,\mu_i} & & & & \frac{1}{\mu_0(\omega)} & 0 & \dots & \mathbb{E}_{P^\dagger_{-i}(\cdot|\omega)}(x_i^*(\mu_i, \mu_{-i})) & \mathbb{E}_{P^\dagger_{-i}(\cdot|\omega)}(x_j^*(\mu_i, \mu_{-i})) & \dots & \\
(1)_{\omega,\mu_i'} & & & & \frac{1}{\mu_0(\omega)} & 0 & & \mathbb{E}_{P^\dagger_{-i}(\cdot|\omega)}(x_i^*(\mu_i', \mu_{-i})) & \mathbb{E}_{P^\dagger_{-i}(\cdot|\omega)}(x_j^*(\mu_i', \mu_{-i})) & & \\
\vdots & & -H & & \vdots & & & \vdots & & \ddots & \\
(1)_{\omega',\mu_i} & & & & 0 & \frac{1}{\mu_0(\omega')} & \dots & 0 & 0 & & \mathbb{E}_{P^\dagger_{-i}(\cdot|\omega')}(x_i^*(\mu_i, \mu_{-i})) \\
(1)_{\omega',\mu_i'} & & & & 0 & \frac{1}{\mu_0(\omega')} & & 0 & 0 & & \mathbb{E}_{P^\dagger_{-i}(\cdot|\omega')}(x_i^*(\mu_i', \mu_{-i})) \\
\vdots & & & & \vdots & & & \vdots & & & \vdots \\
(1')_{\omega} & \mathbf{1} & \mathbf{0} & \dots & 0 & \dots & & & & & \\
(1')_{\omega'} & \mathbf{0} & \mathbf{1} & & 0 & \dots & & & & & \\
\vdots & \vdots & & \ddots & & & & & & &
\end{array}$$

The columns to the left, which correspond to derivatives w.r.t. $P_i$, equal the Hessian $H$ of the cost of information as $P_i$ only enters agent $i$'s FOCs through the marginal cost:

$$H = \left[\frac{\partial^2 c(P_i)}{\partial P_i(\mu_i|\omega)\mu_0(\omega)\partial P_i(\mu_i'|\omega')\mu_0(\omega')}\right]_{((\mu_i,\omega),(\mu_i',\omega'))}.$$

Note that columns corresponding to the derivatives w.r.t. $u_j(\omega)$ for some agent $j$ and state $\omega$ equal the probability that $j$ gets the good in state $\omega$ given $P^\dagger_{-i}$, for each possible report of agent $i$. It follows directly from Blackwell's principle of irrelevant information that the rows of the above matrix are linearly independent: the efficient allocation must vary with $i$'s report if it is efficient for $i$ to acquire some information.

The key element to prove is that the full surplus extraction condition (5) imposes additional restrictions on $P_i^\dagger$. That is, we need to prove that the derivative of the LHS of (5) w.r.t. $(P_i, \zeta_i, u)$ is linearly independent from the rows in the above matrix. The derivative of the LHS of (5) w.r.t. $P_i(\mu_i|\omega)$, $\zeta_i(\omega)$ and $u_i(\omega)$ equal

$$\sum_{\omega'} \mu_0(\omega') \sum_{\mu_i'} P_i(\mu_i'|\omega') \frac{\partial^2 c(P_i)}{\partial P_i(\mu_i'|\omega)\mu_0(\omega')\partial P_i(\mu_i|\omega)\mu_0(\omega)}, \ -1, \text{ and } 0, \text{ respectively.}$$

To replicate the derivative w.r.t. $P_i$ from a linear combination of the above matrix, we would need to sum all rows corresponding to the system of equations (1), weighting each row $(1)_{\omega,\mu_i}$ by $\mu_0(\omega)P_i^\dagger(\mu_i|\omega)$. However, this linear combination also replicates the columns corresponding to derivatives w.r.t. $u$ only if $\mathbb{E}_{\mu_0,P_i^\dagger,P_{-i}^\dagger}[x_j^*(\mu)] = 0$ for all $j$, i.e., only if no agent gets the good with positive probability under the efficient solution. That cannot be true if it is socially efficient for agent $i$ to acquire some information.

Thus the Jacobian of $\hat{\Phi}$ at the efficient solution has full rank, and $\mathbf{0}$ is a regular value of $\hat{\Phi}$. The Transversality theorem states that, except for a nullset $\mathcal{U}^0 \subset \mathcal{U}$ of preferences, $\mathbf{0}$ is a regular value of $\hat{\Phi}(\cdot; u)$. Then by the Regular Value theorem, $\hat{\Phi}^{-1}(\mathbf{0}; u)$ is a smooth manifold of dimension

$$\dim \hat{\Phi}^{-1}(\mathbf{0}; u) = |\Omega| \times (|\operatorname{supp} P_i^\dagger| + 1) - \Big(1 + |\Omega| \times (|\operatorname{supp} P_i^\dagger| + 1)\Big) < 0.$$

Therefore we conclude that for a full measure set of preferences $u \in \mathcal{U} \setminus \mathcal{U}^0$, the set of choice rules for $i$ solving (1), (1′) and (5) is empty. In other words, for all preferences in $\mathcal{U} \setminus \mathcal{U}^0$, it is impossible for the designer to both incentivize agent $i$ to choose the efficient strategy and extract all surplus from $i$.

We have left to show that the set of preferences $\mathcal{U}^0$ for which full surplus extraction is feasible, is closed. This can be done using the same argument as in the proof of Theorem 1, and we omit the formal proof for the sake of brevity. □

# REFERENCES

Bergemann, D., Shi, X., and Välimäki, J. (2009). Information acquisition in interdependent value auctions. *Journal of the European Economic Association*, 7(1):61–89.

Bergemann, D. and Välimäki, J. (2002). Information acquisition and efficient mechanism design. *Econometrica*, 70(3):1007–1033.

Bikhchandani, S. (2010). Information acquisition and full surplus extraction. *Journal of Economic Theory*, 145(6):2282–2308.

Bobkova, N. (2019). Information Choice in Auctions. Working paper.

Caplin, A., Dean, M., and Leahy, J. (2022). Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy. *Journal of Political Economy*, 130(6):1676–1715.

Crémer, J. and McLean, R. P. (1988). Full extraction of the surplus in bayesian and dominant strategy auctions. *Econometrica: Journal of the Econometric Society*, pages 1247–1257.

Denti, T. (2018). Unrestricted Information Acquisition. Working Paper.

Hatfield, J. W., Kojima, F., and Kominers, S. D. (2018). Strategy-proofness, investment efficiency, and marginal returns: An equivalence. *Becker Friedman Institute for Research in Economics Working Paper*.

Hébert, B. M. and La'O, J. (2020). Information acquisition, efficiency, and nonfundamental volatility. Technical report, National Bureau of Economic Research.

Immorlica, N. S., Leshno, J. D., Lo, I. Y., and Lucier, B. J. (2018). Costly Information Acquisition and Stable Matching Mechanisms. Working Paper.

Jehiel, P. and Moldovanu, B. (2001). Efficient design with interdependent valuations. *Econometrica*, 69(5):1237–1259.

Kamenica, E. and Gentzkow, M. (2011). Bayesian Persuasion. *American Economic Review*, 101(6):2590–2615.

Larson, K. and Sandholm, T. (2001). Costly Valuation Computation in Auctions. *Theoretical Aspects of Rationality and Knowledge (TARK VIII)*, pages 169–182.

Mas-Colell, A. (1989). *The Theory of General Economic Equilibrium: A Differentiable Approach*. Cambridge University Press.

Matějka, F. and McKay, A. (2015). Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model. *American Economic Review*, 105(1):272–98.

Mensch, J. (2022). Screening Inattentive Buyers. *American Economic Review*, 112(6):1949–84.

Myerson, R. B. (1981). Optimal auction design. *Mathematics of operations research*, 6(1):58–73.

Persico, N. (2000). Information Acquisition in Auctions. *Econometrica*, 68(1):135–148.

Ravid, D., Roesler, A.-K., and Szentes, B. (2022). Learning Before Trading: On the Inefficiency of Ignoring Free Information. *Journal of Political Economy*, 130(2):346–387.

Roesler, A.-K. and Szentes, B. (2017). Buyer-Optimal Learning and Monopoly Pricing. *American Economic Review*, 107(7):2072–80.

Sims, C. A. (2003). Implications of Rational Inattention. *Journal of Monetary Economics*, 50(3):665–690.

Zhong, W. and Bloedel, A. (2020). The Cost of Optimally Acquired Information. Working Paper.