Collective Information Acquisition

Ran Eilat* and Kfir Eliaz†

August 2022

Abstract

We consider the problem faced by a group of players who bargain over what public signal to acquire before deciding on a collective action. The players differ in their privately known state-dependent payoffs from taking the action, and therefore differ also in their willingness to pay for the public signal. We take a mechanism design approach to characterizing the efficient frontier of outcomes achievable via bargaining over information. We identify novel distortions in the optimal information structure that arise from the information asymmetry and from the fact that after the signal is realized, the outcome is determined in equilibrium of a subsequent voting game.

Keywords: Collective decision-making, Mechanism design, Information design, Rational inattention, Public good provision.

1 Introduction

There are many situations in which a group of individuals needs to take a collective decision in the face of uncertainty. In such situations, the group members often want to have some information presented to them prior to taking the decision. However, collecting and processing information is costly in terms of time, effort, or money, and the group members typically have different preferences over the final outcomes of their decision. How should the group members decide what information to acquire and how to distribute its cost? Despite being ubiquitous, this form of "collective bargaining" over information is largely underexplored in the literature. This paper takes a preliminary step toward understanding the outcomes of such bargaining.

Consider for example a household that needs to make an important decision such as whether to have a child, whether to send their child to a non-standard educational environment, or whether to relocate. Such decisions typically depend on many unknown factors and hence the couple may want to invest resources in acquiring some information about them. While both partners typically want to take the "right decision," each may not necessarily know how

*Department of Economics, Ben Gurion University, eilatr@bgu.ac.il.

†School of Economics, Tel-Aviv University and Eccles School of Business, University of Utah, kfire@tauex.tau.ac.il.

intensely the other feels about making the "wrong" decision. How informed would the couple choose to be? Would they choose to acquire the optimal information structure? How would they divide the burden of collecting and processing the information?

As another example, consider a group of division heads in some firm who need to decide whether to undertake some project, develop a new product, or enter a new market. Because of the uncertainty regarding the right decision, the division heads may want to carry out a pilot project that involves all the divisions (e.g., software, hardware, product, marketing, etc.). While all division heads want to make the right decision, they typically differ in the price they pay in the case of a failure (which oftentimes is known only to each devision head). Because conducting a pilot is costly, the devision heads need to agree on its scale and goals and how to divide the labor and costs among them. What characterizes the information that the optimal pilot reveals? How does the rule for deciding on the collective action (e.g., whether undertaking the project requires the consent of all devision heads or just a majority of them) interact with the devision heads' decision on the pilot characteristics?

Other examples with similar features include partners in a firm who need to vote on a merger or an acquisition and hence need to agree on which consulting firm to hire for market research (and what research it will conduct), or committee members who need to vote on whether to hire a candidate and hence deliberate over what information to collect about the candidate. In these situations and others, it is natural to ask what characterizes the information that the group acquires? How is the acquired information affected by the fact that the group members will base their collective decision on it? What does the optimal signal look like, compared to the case in which the players' preferences are commonly known?

To address these questions, we propose the following stylized model. A group of players is faced with a binary decision: whether or not to depart from a status quo (the "default action"). There are two states of nature, and all players would like the action to match the state. However, they differ in their disutility from a mismatch, and this disutility is privately observed. Prior to making the binary decision, the players have the opportunity to collectively acquire a costly public signal about the state. The players then proceed in two steps. First, they bargain over which signal to acquire and how to distribute its cost. Second, they all observe the signal realization and vote on the binary decision, where a supermajority is required to depart from the status quo.[1] If no information is acquired, all players prefer the status quo.

Our analysis abstracts from the particular protocol of bargaining over information by following Myerson and Satterthwaite (1983) and taking a mechanism design approach to exploring the bounds on the "constrained" social surplus that the group can achieve. That is, we characterize the optimal feasible mechanism for deciding which signal to acquire, taking into account

[1]In the paper the term "supermajority" refers to a minimal threshold of m votes that is required to depart from the status quo, where $m \geq 1$.

the incentive and participation constraints as well as the second-stage voting game.

Within this framework, we draw the following insights. Even if the group members' preferences are known, the fact that a supermajority vote is required to depart from the status quo leads to a distortion of the information that is acquired (relative to the unconstrained socially optimal signal). Furthermore, the higher the supermajority requirement, the lower the social welfare. When the participants' disutility from making the wrong decision is unobservable, an incentive to free ride arises in the sense that an individual will want to behave as if the information is not that important to him so that others will bear its cost. On the one hand, this further lowers the likelihood of acquiring information (relative to the case in which preferences are known) and makes the group even more conservative in departing from the status quo. On the other hand, when the group does decide to leave the status quo, it is more confident in its decision (i.e., the probability that this is the right decision is higher relative to the case of complete information).

The socially optimal bargaining outcome exhibits the following features. In one subset of the type space, no signal is acquired. In a second subset, the acquired signal is at its optimal "interior" solution: its structure is optimal given the cost, it attenuates the incentives to free-ride, and it is always instrumental (that is, it always has at least one realization for which the collective action is different than if no signal is acquired). In a third subset of the type space the acquired signal is distorted to be minimally "supermajority persuasive": it is chosen such that one of its realizations just ensures supermajority support for the non-default action. This last subset illustrates the distortion caused by the presence of a second-stage voting game, which is beyond the control of the designer.

In addition to characterizing the solution to the collective bargaining over information, this paper also contributes to three different strands of literature. First, our group decision problem may be viewed as a variant of rational inattention à la Sims (2003). While this literature has focused exclusively on individual decision makers, in many of the applications the decision is inherently made by a group (a management team, a household, etc.). Our paper introduces a framework of *collective* rational inattention: a group needs to agree on which signal to acquire, taking into account the trade-off between the cost and benefit of more precise information. There are three key differences between the problem we study and the problem of individual rational inattention. First, in our setting, the final decision following a signal realization is determined by an *equilibrium* in a game. Second, the group members may disagree on the benefit from each signal. Finally, in order to aggregate the individuals' willingness to pay for signals, the individuals need to disclose their private information.

Second, we expand the scope of the information design literature by introducing a new design problem with the following features: (i) the "receivers" themselves have to choose the optimal information structure (i.e., there is no "sender"), (*ii*) the optimal information structure depends on the receivers' private types, and (*iii*) information is costly.

Finally, this paper introduces a new class of public good provision problems. In our setup the public good is information: each player benefits from the public signal, but prefers others to bear its cost. However, unlike a standard public good problem, here the public good (the signal) does not directly produce utility for the agents, but it is instrumental in making a more informed choice in a subsequent game (the voting game). Additionally, in contrast to a "standard" public good, in our model the players do not necessarily agree on the ranking of (noisy) signals, even when the cost is ignored (although they all agree that full information is the best signal). This is because they may disagree on the optimal collective action for each realized posterior belief. Lastly, the public good in our framework is a multidimensional object. Nevertheless, we manage to "map" the problem back to one that can be solved using Myersonian techniques.

Related literature. As mentioned above, our analysis combines information design with mechanism design. In a linear environment with a single player, Kolotilin et al. (2017) show that the optimal signal can be implemented without relying on the player's private information. However, it is well known that in environments with multiple interacting players (e.g., Bergemann and Morris, 2013; Alonso and Câmara, 2016; Taneva, 2019; Mathevet, Perego, and Taneva, 2020) ignoring the players' private information is suboptimal. Candogan and Strack (2021) show that when there are more than two possible actions, ignoring the player's private information is suboptimal even if there is only one receiver.

Several recent works have addressed the problem of designing information for a group of voters. Notable papers include Wang (2013), Schnakenberg (2015), Alonso and Câmara (2016), Bardhi and Guo (2018), Chan et al. (2019), and Arieli and Babichenko (2019). These studies characterize the signal that maximizes the probability that in equilibrium voters vote on the outcome favorable to the sender. They differ in whether the designed signals are private or public, and in the class of voting rules that are considered. There are two key differences between these papers and ours. First, in these papers the voters' state-dependent utilities are *commonly known* (i.e., voters have no private information) and hence, in order to design the optimal signal, there is no need to elicit information from the voters. Second, in these papers signals are *costless*, and the problem is to find the signal that induces voters to coordinate on an equilibrium that is favorable for the sender.

The question we study is also related to the problem of designing voting rules that incentivize the voters to acquire costly information. Persico (2004) characterizes the optimal size and voting threshold that efficiently aggregates information when each voter needs to pay a cost to acquire a private binary signal. Gershkov and Szentes (2009) extend the analysis to a broader class of voting mechanisms. Our approach differs in that voters' willingness-to-pay for information is private and the signal they acquire is public. We fix the voting rule and look for the optimal signal, taking into account that this signal depends on the voters' private information, and that the signal realization affects voting behavior. Relatedly, Godefroy and Perez-Richet (2013) study a model where in the first stage a group of asymmetrically informed individuals vote on whether to acquire full information on their payoffs from a proposal and, if information is acquired, they proceed to vote on the proposal. They show that the likelihood of remaining with the status quo increases with the supermajority requirement in the first stage and decreases with the supermajority requirement in the second stage.

An alternative approach to the study of collective information acquisition is analyzed by Chan et al. (2018). They consider a dynamic model where at each point in time a group receives an exogenous signal and needs to vote on whether to stop and vote on a binary action, or to continue and receive additional signals. Unlike us, they study a stopping problem in which the signal is exogenously given and the players' preferences are commonly known.[2] Relatedly, Gersbach (2000) considers a group with known preferences who can either accept a policy with no information or defer the vote on the policy until after the state is realized.

Finally, our paper contributes to the literature that examines how strategic players free-ride on the information acquisition of others (see, e.g., Bolton and Harris, 1999; Bergemann and Välimäki, 2000; Décamp and Mariotti, 2004; Aghamolla and Hashimoto, 2020). The key difference is that in all of these works, agents privately decide either to acquire costly information (for example by investing in R&D or assessing the profitability of an industry), or to wait and learn from the actions of other players. By contrast, in our work the players jointly decide on what public signal to acquire and free-riding occurs by pretending that information is less valuable. In addition, our innovation is that we characterize the socially optimal information structure, and identify the optimal implementable information structures when players' types are private.

Outline. The remainder of the paper is organized as follows. Section 2 presents the model, solves the central-planner benchmark, and provides an illustrative example. The mechanism-design problem is presented in Section 3 and solved in Section 4. The latter section contains the lion's share of the theoretical analysis. We provide an intuitive summary of the main steps of the analysis at the beginning of the section, to allow readers who are more interested in the qualitative form of the solution to skip directly to Section 5, which characterizes and discusses properties of the optimal signal structure. In Section 6 we discuss some key ingredients of our model and the challenges involved in extending our framework in several directions. Concluding remarks are presented in Section 7. All proofs are relegated to the appendix.

[2]For additional related works that take a collective search approach to sequential information gathering by a group, see the references in Chan et al. (2018).

2 Model

Our model consists of the following components.

Players and payoffs. There are n players who have to jointly agree on a decision $a \in \{0, 1\}$. Following the literature on strategic voting (most notably, Feddersen and Pesendorfer, 1998), we assume that each player's payoff from the collective action, u_i, depends on the joint action, on his type $\theta_i \in \Theta \subseteq [0, 1]$, and on the state of the world $\omega \in \{0, 1\}$ as follows:

$$u_i(a, \omega, \theta_i) = \begin{cases} 1 & if \quad a = \omega \\ \theta_i & if \quad a = 1, \omega = 0 \\ 1 - \theta_i & if \quad a = 0, \omega = 1 \end{cases}$$

Each player has quasilinear preferences over the collective action and any additional costs he incurs.

We assume that the players do not observe the realization of ω and have the common prior belief that the probability of $\omega = 1$ is $p \in (0, 1)$. In addition, each player i privately and independently draws a type θ_i from a common distribution F on the interval $\Theta = [0, 1 - p]$ (we explain below why we assume that $\theta_i \leq 1 - p$). We assume that F admits a density f that is strictly positive, continuously differentiable, and bounded over $[0, 1 - p]$. Let $v(\theta_i) \equiv \theta_i - (1 - F(\theta_i)) / f(\theta_i)$ denote the virtual valuation of the player's type θ_i. We assume that F is regular, i.e., $v(\theta_i)$ is increasing in[3] θ_i.

Our specification of the utility function u_i implies that player i weakly prefers the joint decision $a = 1$ if and only if, given any information he has, his posterior belief on $\omega = 1$ is at least $1 - \theta_i$. This follows from observing that if the posterior belief on $\omega = 1$ is r, then the action $a = 1$ yields an expected payoff of $r \cdot 1 + (1 - r) \cdot \theta_i$ whereas the action $a = 0$ yields an expected payoff of $r \cdot (1 - \theta_i) + (1 - r) \cdot 1$. From our assumption that $p \leq 1 - \theta_i$ for every θ_i, it follows that without further information on the state each player prefers the action $a = 0$. This gives a clean benchmark that, without additional information, the group remains with the status quo.

Costly signals. Before making the joint decision (in a manner described below), the players have the opportunity to acquire a *public* signal on the state ω. A signal can be represented by a probability distribution over posterior beliefs on $\omega = 1$, such that the expected posterior belief on $\omega = 1$ equals the prior p.[4] To simplify the exposition we assume that the distribution

[3] The assumption that F is regular simplifies the analysis of the mechanism design problem that we introduce later. It is a standard assumption in the mechanism design literature, where it guarantees that the solution to the design problem satisfies a monotonicity condition that ensures incentive compatibility. In our setting, regularity helps us prove that the solution to our design problem indeed satisfies a similar monotonicity condition.

[4] In our collective decision-making setting, there is no loss of generality in representing a signal as a distribution

is discrete, with finitely many possible realizations.[5] We denote by q_j the probability that the posterior belief on the state $\omega = 1$ is r_j and by J the total number of posteriors. We then have

$$\sum_{j \in \{1,\dots,J\}} q_j \cdot r_j = p, \tag{1}$$

where $0 < q_j \leq 1$ and $0 \leq r_j \leq 1$ for all $j \in \{1, ..., J\}$, and $\sum_{j\in\{1,\dots,J\}} q_j = 1$. The players can decide to acquire no information. This option is equivalent to choosing the degenerate signal $J = 1$, $q_1 = 1$, and $r_1 = p$.

Acquiring and/or analyzing signals is costly. Following the rational inattention literature (in particular, the posterior-based approach of Caplin, Dean, and Leahy, 2020), we assume that a signal's cost is a function of the induced distribution over posterior beliefs. For now we assume only that given J, the cost function $c\left(\{(q_j, r_j)\}_{j=1}^J\right)$ is twice continuously differentiable and monotone with respect to the Blackwell ordering, and that acquiring no information is costless (for our main characterization result we will impose additional structure). For concreteness, and to be able to give illustrative examples, we will occasionally use a cost specification that is proportional to the mutual information between the signal realization and the state. That is, the cost of the signal $\{(q_j, r_j)\}_{j=1}^J$ will be given by the expected KL-divergence (or relative entropy) between the posteriors and the prior:

$$c\left(\{(q_j, r_j)\}_{j=1}^J\right) = \kappa \cdot \textstyle\sum_{j=1}^J q_j D_{KL}(r_j \mid\mid p), \tag{2}$$

where κ is some positive constant, and[6]

$$D_{KL}(r \mid\mid r') \equiv r \log \frac{r}{r'} + (1-r) \log \frac{1-r}{1-r'}. \tag{3}$$

This specification captures situations where there is an overwhelming amount of information available and the difficulty is in processing and understanding that information (see, e.g., Maćkowiak, Matějka, and Wiederholt, 2018). As is well known, the cost function given by (2) is monotone with respect to the Blackwell ordering. However, our analysis applies to a

over posteriors that average to the prior belief. First, as long as the mean posterior equals the prior, there is some signal that generates the distribution over posteriors. Second, our preference specification implies that voting for $a = 1$ is dominant for type θ_i if and only if the realized posterior is at least $1 - \theta_i$. Thus, for a given tie-breaking rule, each player's decision, and hence the final outcome, is pinned down by the realized public posterior.

[5] We can extend the analysis to signals with infinitely many realizations (posteriors). Apart from affecting notation, this requires a change only in the proof of Lemma 1, where instead of merging a pair of posteriors, we need to merge all the posteriors above $1 - \theta^{(n-m+1)}$, and all the posteriors below $1 - \theta^{(n-m+1)}$.

[6] Since there are only two states, we represent a distribution over the states by the probability on $\omega = 1$. Thus, the divergence between the two distributions can be written as a function of the probability that each distribution assigns to the state $\omega = 1$.

broader class of cost functions (P1–P3 in Section 4.2 give the general sufficient conditions for the cost function). For example, it is easy to verify that these conditions are also satisfied by a cost function that is proportional to the variance of the induced posteriors on the high state.

The cost of the signal has to be covered by the players. We denote by t_i the cost borne by player i, so that[7] $\sum_{i=1}^{n} t_i = c\left(\{(q_j, r_j)\}_{j=1}^{J}\right)$. Thus, the net payoff of type θ_i from action a in state ω is given by $u_i(a, \omega, \theta_i) - t_i$.

The assumption that the cost of information can be shared among the players (or that players can compensate each other using another sort of transferable utility) is crucial for our analysis. Sharing this cost can be interpreted, for example, as sharing the monetary cost of the signal (e.g., when different departments in a firm use their budgets to pitch in for the cost of hiring a consultant); or as reallocating chores, consumption, or resources (e.g., in a household or a firm); or as sharing the collective effort of processing the acquired information (e.g., the amount of documents that need to summarized, or the time involved in organizing the data).

Agreeing on a signal. The group members decide which signal to acquire and how to distribute the costs through some form of bargaining. In doing so they take into account that each of them has private information and that the signal realization will affect the decision on the collective action. As we explain in the next section, our analysis is not tied to any particular bargaining protocol. Instead, we take a mechanism design approach to characterizing the bargaining outcomes that maximize the constrained social surplus.

We assume that when a player refuses to take part in the discussion of what signal to acquire, he prevents the group from making a decision on the signal. This assumption, which is typical in almost all public good settings (e.g., Mailath and Postlewaite, 1990; Hellwig, 2003), effectively means that each player has veto power with respect to acquiring the public signal. Notice that, with no additional information, the group will stick with the status quo. For instance, recall the example of several department heads that need to agree on a pilot study. If the pilot requires experts from every department (software, hardware, etc.) then any department head can block the pilot by refusing to allocate manpower. With no pilot, the firm will not undertake the project. Hence, by opting out of bargaining a player expects to get the payoff of the status-quo option. This assumption is helpful for tractability since it fixes the payoff of the outside option of a player in a way that is independent of the other participants. We further discuss this assumption in Section 6.2.

Voting. After the players observe the realization of the public signal (if one is acquired), they vote on the collective decision using an m-majority rule: the action $a = 1$ is chosen if and only

[7] While we require that the players exactly cover the cost of the signal, the analysis would remain unchanged if instead we were to impose the weaker requirement that the cost of the signal be at most the sum of costs borne by the players.

if at least m players vote for this option. Otherwise, the default action $a = 0$ is chosen. We assume that the players do not choose weakly dominated strategies. In addition, whenever a player is indifferent between $a = 1$ and $a = 0$, he breaks ties in favor of $a = 1$. Thus, player i votes for $a = 1$ if and only if the realized posterior belief that the state is $\omega = 1$ is at least $1-\theta_i$. Consequently, the alternative $a = 1$ is chosen if and only if the realized posterior belief that the state is $\omega = 1$ is at least $1 - \theta^{(n-m+1)}$, where $\theta^{(k)}$ is the k^{th} smallest element in $\theta = (\theta_1, ..., \theta_n)$. For example, if choosing the non-default action $a = 1$ requires unanimity, i.e., $m = n$, then for this action to be chosen the realized posterior belief has to be larger than $1 - \theta^{(1)}$, where $\theta^{(1)}$ is the smallest element in θ. Note that, given θ, a signal induces a probability distribution over the outcomes of the vote.

We assume the group cannot make any transfers that are conditional on their votes. This can follow from institutional constraints that prohibit such vote buying, or because the votes are secret, or because such contractual arrangements cannot be enforced (the case of contractible votes is covered by the central-planner benchmark below). In light of this, the group can only bargain over what information to acquire (and not simultaneously on both the information and the ultimate action). We therefore take the voting stage as given, while allowing for any supermajority requirement.

Optimal bargaining outcome. Note that the players' preferences are quasilinear and that the signal realization fully determines the voting outcome. It follows that the ex-ante social surplus that is induced by a mapping from types to signals equals the sum of the players' ex-ante expected utilities, where utilities are evaluated according to the equilibrium outcome in the ensuing voting game. We say that a bargaining protocol is socially optimal if the mapping it induces from types to signals maximizes the ex-ante social surplus.

Our model strikes a balance between being sufficiently simple to analyze and being sufficiently general to accommodate a broad range of situations.[8] In Section 6 we discuss the role

[8]For instance, as we mention in the Introduction, our model accommodates situations of household decisions that are made in the face of uncertainty, where the partners differ in their attitudes toward "mistakes" and their bargaining involves some sort of "transfers." These assumptions are consistent with an extensive literature on intra-household bargaining; see, e.g., Doepke and Kindermann (2019) and Ashraf et al. (2021). Also, our model fits both the case in which partners are symmetrically informed about each other's preferences (that is, attitudes toward mistakes) and the case in which they are not. Assuming that spouses who share a household may be asymmetrically informed about preferences is consistent with Ashraf et al. (2021) who argue, in the context of fertility decisions, that spouses may have different perceptions of costs and benefits of decisions and that these perceptions are private information.

In our department heads example, our model is consistent with situations in which the firm will undertake some project only if a majority (or all) of the department heads support it. Conducting a pilot project is oftentimes costly, requires a joint effort from all departments, and its outcome is publicly observed. Indeed, a department head may be skeptical about the benefit of the pilot, but she may be willing to go ahead with it if she is appropriately compensated, for example by temporarily allocating manpower or other resources to her department. However, if after a pilot is run, the department head is still not convinced that the project is worth

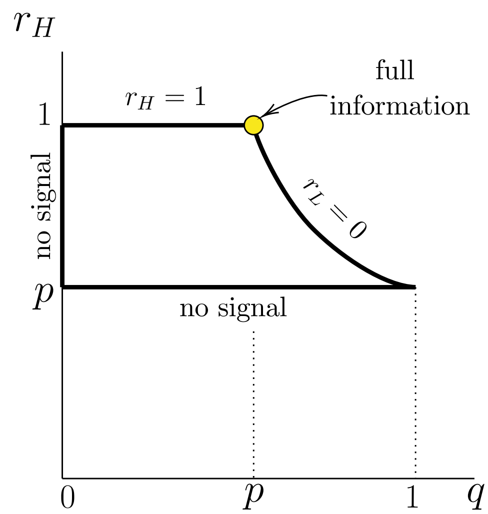

Figure 1: Feasible signals

of some key features of our model in the analysis, suggest possible directions for extending the model, and explain the technical challenges involved in pursuing them.

2.1 Simplifying the domain of signals

Our first observation is that signals that induce only two posterior beliefs on $\omega = 1$ (one on either side of the prior belief) dominate signals with more posterior beliefs on $\omega = 1$.

Lemma 1 *For any signal that induces more than two posterior beliefs, there exists a signal that induces only two posterior beliefs, generates the same distribution over actions for each realization of state and types, and has a strictly lower cost.*

This result is straightforward in standard information design problems where signals are costless. In such settings all that matters is the distribution over the actions in each state, and this distribution can be replicated by signals that induce two posterior beliefs when there are only two actions. In our setup, signals are costly and the cost depends on the entire distribution of posteriors. However, our assumption that the cost increases with Blackwell informativeness implies that the optimal mechanism does not need to employ signals with more than two posteriors (for an analogous result in a model of individual rational inattention see, e.g., Lemma 1 in Matějka and McKay, 2015).

In light of this observation, we restrict attention to signals that induce at most two posterior beliefs.[9] Thus, a signal can be represented by a pair (q, r_H), where $q \in [0, 1]$ is the probability

undertaking, it is unlikely that any transfer will persuade her to risk her reputation for a project that she thinks will turn out to be a failure.

[9]Because the players are risk neutral, if we switch from a signal with more than two posteriors to a cheaper signal that induces only two posteriors, it is possible to readjust the transfers between the players such that the

that the posterior belief on $\omega = 1$ is $r_H \geq p$. Equation (1) then implies that with probability $1-q$ the other posterior belief induced by the signal is $r_L \equiv (p - qr_H)/(1-q) \leq p$. Thus, when the realized posterior belief is r_L all players agree that the optimal action is $a = 0$. When the realized posterior belief is r_H there are m players who prefer $a = 1$ to $a = 0$ if and only if $r_H \geq 1 - \theta^{(n-m+1)}$. Notice that, since $r_L \geq 0$, it must be the case that $p \geq qr_H$. Figure 1 illustrates the set of all possible signals, depicted on the plane of q and r_H.

Choosing $q = 0$ or $q = 1$ is equivalent to purchasing no signal. The cost in this case is 0 by assumption. We say that a signal is *informative* if $q \in (0, p/r_H]$ and $r_H > p$. Given m, we say that a signal is *instrumental* for a type profile θ if for at least one of the signal's realizations there is an m-majority for the non-default action $a = 1$. This means that a signal (q, r_H) is instrumental for θ if $r_H \geq 1 - \theta^{(n-m+1)}$ and $q \in (0, p/r_H]$.

2.2 A central-planner benchmark

To better understand the implications of the players' private information and the supermajority requirement for departing from the status quo, we begin by analyzing the benchmark where a central planner, who knows the players' types, chooses both the socially optimal signal and the collective action as a function of the signal realization.

Given a type profile θ, let $\overline{\theta} = \frac{1}{n}\sum_{i=1}^{n} \theta_i$ denote the average type in θ. Clearly, if the planner acquires a signal, it is necessarily instrumental: the posterior r_H has to lead to the action $a = 1$ and the posterior r_L has to lead to $a = 0$ (otherwise, the players incur a cost but ignore the information). The expected per-capita social welfare from purchasing an instrumental signal (q, r_H) for the type profile θ is given by

$$q \cdot \left[1 \cdot r_H + \overline{\theta} \cdot (1 - r_H)\right] + (1-q) \cdot \left[\left(1 - \overline{\theta}\right) r_L + (1 - r_L)\right] - \frac{1}{n} c\left(q, r_H\right). \tag{4}$$

This expression is a sum of three terms. The first term is the total per-capita welfare when the posterior r_H is realized, which occurs with probability q. Since the planner takes the action $a = 1$ in this event, the per-capita payoff is 1 with probability r_H, and is $\overline{\theta}$ with the complementary probability. The second term is the total per-capita welfare when the posterior r_L is realized, which occurs with probability $1-q$. In this case, the action $a = 0$ is taken. The third term is the per-capita cost.

Alternatively, if the planner does not acquire any signal, the players remain at the prior beliefs and the planner takes the action $a = 0$. In this case, the per-capita social welfare is $1 - p\overline{\theta}$. The planner acquires information if the signal (q, r_H) that maximizes (4) achieves a social welfare above $1 - p\overline{\theta}$.

---

decrease in cost is translated into a constant reduction in the interim payment of each player. This modification maintains ex-post budget balancedness, incentive compatibility, and interim individual rationality.

Assume that the cost function is given by Equation (2); then the planner's solution is[10]

$$r_H = \frac{e^{\frac{n}{\kappa}} - e^{\frac{n}{\kappa}\overline{\theta}}}{e^{\frac{n}{\kappa}} - 1}, \qquad r_L = \min\left\{\frac{e^{\frac{n}{\kappa}\left(1-\overline{\theta}\right)} - 1}{e^{\frac{n}{\kappa}} - 1}, p\right\}, \qquad q = \frac{p - r_L}{r_H - r_L}. \tag{5}$$

This implies that a signal is acquired whenever $r_L < p$, which occurs if $\overline{\theta} > 1 - \frac{\kappa}{n} \ln\left[p\left(e^{\frac{n}{\kappa}} - 1\right) + 1\right]$.

The socially optimal signal structure exhibits the property that both r_H and r_L are decreasing in $\overline{\theta}$, whereas q is increasing in $\overline{\theta}$. To get some intuition for this, note that when the average type is higher, the social disutility from taking the action $a = 0$ when the state is $\omega = 1$ is higher. At the same time, the disutility from making the opposite mistake is lower. Therefore, when the average type is higher, the action $a = 1$ is socially more desirable, which suggests that q should be higher. For the same reason, higher types, who are biased toward action $a = 1$, are willing to suffer a reduction in r_H (which lowers the confidence that $a = 1$ is the right action when this posterior is realized) in exchange for a reduction in r_L (which raises the confidence that $a = 0$ is the right action when this posterior is realized). Of course, the precise argument for the comparative statics of the signal with respect to the type profile also depends on the properties of the cost function. Sufficient conditions for these comparative statics are provided in Section 4.2.

The socially optimal signal also changes in an intuitive way as κ, the scaler of the cost function, changes.

Proposition 1 *Assume that the cost function is given by Equation (2). Let θ be a type profile and $\kappa_L < \kappa_H$ be two levels of the scaler of the cost function such that a signal is acquired for θ at both scaler levels. Then the signal for κ_L is more Blackwell informative than the signal for κ_H. Furthermore, at the limit, as $\kappa \to 0$, the socially optimal signal that is acquired for any type profile converges to the fully informative one.*

2.3 An illustrative example

To illustrate what forces are at work in the absence of a central planner, consider the following example. Suppose that there are only two players (i.e., $n = 2$), $p = 0.5$, and the cost function is given by Equation (2) with $\kappa = 0.5$.

The central planner benchmark. Suppose first that there's a central planner, as described above. For the type profile $\theta = (0, 0)$, the planner's decision is straightforward: since the action $a = 0$ dominates the action $a = 1$ for both players in each state of the world, implementing

[10]These equations are directly obtained from the first-order conditions of the planner's problem, which are also given by (FOCq) and (FOCr) in Section 4.2 for the case of $w = \overline{\theta}$. Also, notice that since $\overline{\theta} \leq 1 - p$ it follows that $r_H \in (p, 1)$.

$a = 0$ without acquiring any information is optimal. By contrast, for the type profile $\theta = (0.5, 0.5)$, acquiring some information is beneficial for both players. A computation shows that the signal that is given by $q = 0.5$ and $r_H \approx 0.88$ maximizes (4) and is better than acquiring no information.

Now suppose that the type profile is $\theta = (0, 0.5)$. In this case, player 1 (weakly) prefers the action $a = 0$ regardless of the state of the world, whereas player 2 prefers the action $a = 1$ when his posterior belief on the state $\omega = 1$ is greater than 0.5. The planner, who is not concerned with the preferences of each player individually but rather with the aggregate social surplus, chooses the signal that maximizes (4), provided that it is better than acquiring no information. A computation shows that this signal is given by $q \approx 0.24$ and $r_H \approx 0.97$. Intuitively, the social harm from a mismatch when $a = 1$ is greater for the type profile $(0, 0.5)$ than for the type profile $(0.5, 0.5)$. Thus, the optimal signal for $(0, 0.5)$ leads to adopting $a = 1$ with a lower probability $(0.24 < 0.5)$, but whenever $a = 1$ is taken, it is done with greater confidence $(0.97 > 0.88)$.

A supermajority voting rule. Suppose now that there's no planner and the players vote on the collective action. Suppose further that $a = 1$ requires a unanimous agreement $(m = 2)$. For now, retain the assumption that types are commonly known. It is easy to verify that for the type profiles $(0, 0)$ and $(0.5, 0.5)$, where the players' interests are perfectly aligned, the optimal signals are the same as those purchased by the planner.

Consider now the asymmetric profile $(0, 0.5)$. In this case, the socially optimal signal ($q = 0.24, r_H = 0.97$) is too weak to persuade player 1 to vote for $a = 1$. In other words, this signal is not instrumental and is therefore useless: player 1 will thwart any attempt to deviate from the status quo, regardless of the signal's realization, even when $a = 1$ is the socially optimal action. Indeed, when $\theta = (0, 0.5)$ and a unanimous agreement is required, only signals with $r_H \geq 1 - \min(0, 0.5) = 1$ are sufficiently strong to persuade the two players to vote for $a = 1$ when the posterior r_H is realized. The optimal signal in this case is $q \approx 0.21$, $r_H = 1$. Notice that this signal is distorted compared to the socially optimal one in order for the signal to be instrumental. This observation raises a question: how does the voting stage affect the optimal signal in the more general case, and how does the distortion depend on the required supermajority?

The asymmetric information distortion. Assume now that the players' types are private information. Would the players agree to disclose their types if they knew that the optimal signal will be acquired? While addressing this question requires some more involved computations, it is not too difficult to show that the answer is negative. Perhaps not surprisingly, the players may be tempted to free-ride on each other by pretending to be a lower type, in order to reduce their share in the signal's cost. Indeed, as in the case of "standard" mechanism design, it is possible to weaken the players' incentives to free-ride on each other, by distorting outcomes away from their efficient level (that is beyond the distortion required to make the signal instrumental). Given the parameters of the example, assuming that the distribution F is uniform and letting $m = 2$, it can be shown that it is (ex-ante) optimal not to acquire any signal when the type profile is[11] $\theta = (0, 0.5)$. But what is generally the optimal way to distort a signal that is a multidimensional object?

3 A mechanism for information acquisition

As explained in the Introduction, to characterize the constrained efficient frontier of signals that can be acquired through bargaining, we take a mechanism design approach that abstracts from any particular bargaining protocol.[12] By "constrained efficient" we mean that we take as given the ensuing voting game, which may lead to outcomes that are not socially efficient ex post.

In our setting, there is no loss of generality in restricting attention to direct revelation mechanisms in the first stage of the players' interaction, i.e., when they decide on which signal to acquire.[13] We define an *actual* direct mechanism to be a vector of functions $(q, r_H, t_1, ..., t_n)$, where $q : \Theta^n \to [0,1]$, $r_H : \Theta^n \to [p, 1]$, and $t_i : \Theta^n \to \mathbb{R}$ for every $i \in \{1, \dots, n\}$ such that $\sum_{i=1}^{n} t_i(\theta) = c(q(\theta), r_H(\theta))$. Thus, following a profile of reports $\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_n)$, with probability $q(\hat{\theta})$ the players end up with the posterior probability $r_H(\hat{\theta})$ on the state $\omega = 1$ and with probability $1 - q(\hat{\theta})$ they end up with the posterior probability $r_L(\hat{\theta})$ on that state, where $r_L(\hat{\theta}) \equiv (p - q(\hat{\theta}) \cdot r_H(\hat{\theta}))/(1 - q(\hat{\theta}))$. In addition, each player i pays $t_i(\hat{\theta})$. We say that an actual mechanism is *optimal* if it maximizes the expected social surplus, taking into account the equilibrium of the subsequent voting game.

In the actual mechanism the designer cannot directly control the outcome of the second-stage voting game because the group members cannot commit in advance to how they would vote following any possible realization of the signal. Thus, a player who misreports his true type to the mechanism (say, in order to reduce his share in the cost) retains his ability to vote according to his true preferences in the second stage. As a step toward characterizing the optimal mechanism, we proceed by considering *auxiliary* (direct) mechanisms in which, in addition to choosing which signal to acquire and how to distribute the costs, the mechanism also votes on behalf of the players in the second stage. Thus, an auxiliary mechanism effectively chooses the collective action $a = 1$ whenever $r_H \geq 1 - \hat{\theta}^{(n-m+1)}$, and the collective action $a = 0$ otherwise. In other words, we assume that the players commit to vote according to their

[11] This observation foreshadows a result that will be formally presented in Corollary 2 below.

[12] That is, there is no "real" designer. The mechanism here is used as a method for characterizing the limits of what the players can achieve by any protocol of bargaining.

[13] The mechanism is used only to select the public signal, and not the collective action, and hence the revelation principle follows from standard arguments.

reported types and not their true types.[14] Our focus on direct auxiliary mechanisms follows from the revelation principle, which holds in this environment.[15]

Formally, an auxiliary mechanism is an actual mechanism augmented by two decision functions, $a_H(\hat{\theta})$ and $a_L(\hat{\theta})$, which are the collective actions chosen by the mechanism when the posterior beliefs r_H and r_L are realized, respectively. Thus,

$$a_H(\hat{\theta}) = \begin{cases} 1 & \text{if} \quad r_H(\hat{\theta}) \geq 1 - \hat{\theta}^{(n-m+1)} \\ 0 & \quad\quad \text{otherwise} \end{cases} \tag{6}$$

$$a_L(\hat{\theta}) = 0. \tag{7}$$

The highest expected surplus achievable by an auxiliary mechanism is weakly higher than the highest expected surplus achievable by an actual mechanism. This is because any equilibrium play path in the actual mechanism and the ensuing voting game can be replicated by the auxiliary mechanism. The reason is that deviations from truth-telling are more costly in the auxiliary mechanism than in the actual mechanism. In light of this, we begin by looking for the auxiliary mechanism that attains the highest social surplus. We will then show that the equilibrium that attains this surplus can be replicated by an actual mechanism and the ensuing voting game.

Fix a player i and suppose that the remaining players report their types truthfully. The expected utility of player i of type θ_i who reports $\hat{\theta}_i$ is then given by

$$\begin{aligned} V(\theta_i, \hat{\theta}_i) = \mathbb{E}_{\theta_{-i}} \Big[& q(\hat{\theta}_i, \theta_{-i}) \cdot \Big(r_H(\hat{\theta}_i, \theta_{-i}) \cdot u(a_H(\hat{\theta}_i, \theta_{-i}), 1, \theta_i) + (1 - r_H(\hat{\theta}_i, \theta_{-i})) \cdot u(a_H(\hat{\theta}_i, \theta_{-i}), 0, \theta_i) \Big) \\ & + \; (1 - q(\hat{\theta}_i, \theta_{-i})) \; \cdot \; \Big(r_L(\hat{\theta}_i, \theta_{-i}) \cdot u\,(0, 1, \theta_i) + (1 - r_L(\hat{\theta}_i, \theta_{-i})) \cdot u\,(0, 0, \theta_i) \Big) \\ & - \; t_i(\hat{\theta}_i, \theta_{-i}) \Big] \end{aligned} \tag{8}$$

where $\theta_{-i} \in \Theta^{n-1}$ represents the vector of true types of all players other than i, and $\mathbb{E}_{\theta_{-i}}$ is evaluated according to the probability distribution of the true types θ_{-i}.

Since we are interested in the auxiliary mechanism that maximizes the total surplus (subject to the voting stage), it is useful to represent the players' payoffs as the expected *gains* from information (rather than the utility per se). This gain is computed relative to the case in which the players do not participate in the mechanism and no information is acquired. Note that in

[14] In contrast to the actual mechanism, the auxiliary mechanism can decide on the collective action; however, it is still constrained in terms of how it maps types to collective actions. For example, the auxiliary mechanism cannot force the action $a = 1$: for any signal that is acquired, when r_L is realized, it must choose $a = 0$.

[15] The auxiliary mechanism is a standard mechanism in the sense that it maps reports to final outcomes, and hence the revelation principle applies. In fact, even if there was loss of optimality in restricting attention to direct auxiliary mechanisms, it would not affect our analysis since a solution to the optimal auxiliary mechanism design problem only serves as an upper bound on the optimal actual mechanism.

the latter case, the default action $a = 0$ is chosen and type θ_i's payoff is $p \cdot (1 - \theta_i) + (1 - p)$. Thus, the gain from information of type θ_i of player i who reports $\hat{\theta}_i$ is given by

$$U(\theta_i, \hat{\theta}_i) = V(\theta_i, \hat{\theta}_i) - (p \cdot (1 - \theta_i) + (1 - p)) \tag{9}$$

To simplify the exposition, when all players report truthfully, we denote $U(\theta_i) \equiv U(\theta_i, \theta_i)$.

The objective of the mechanism is to maximize the total ex-ante expected gain from signals under truthful reporting, while taking as given the ensuing voting game:

$$\sum_{i=1}^{n} \mathbb{E}_{\theta_i} U(\theta_i). \tag{OBJ}$$

The auxiliary mechanism has to be ex-post budget balanced: the cost of any signal that is acquired has to be fully covered by the players. In what follows we slightly weaken this requirement and allow the auxiliary mechanism to be balanced only *ex ante*, so that the cost of the acquired signal has to be covered only on average (that is, we allow the mechanism to have a budget deficit in some cases, so long as on average the costs are fully covered):

$$\mathbb{E}_\theta \sum_{i=1}^{n} t_i(\theta) = \mathbb{E}_\theta [c(q(\theta), r_H(\theta))] \tag{BB}$$

However, as is well known (see, e.g., Borgers, 2015, p.47), if a mechanism is ex-ante budget balanced, one can modify the transfers to satisfy ex-post budget balancedness without affecting the interim expected transfers or the incentives for truthful reporting. That is, if a mechanism is incentive compatible, individually rational, and ex-ante budget balanced, then there is another mechanism that achieves the same allocation of types to signals, and which is also incentive compatible and individually rational but is ex-post budget balanced. In light of this, we will focus on ex-ante budget balancedness in the analysis that follows.

The players cannot be forced to participate in the mechanism. Since when a player opts out, he gets the payoff of the status-quo option (evaluated according to the prior belief), it follows that the gain from participation must be nonnegative for any type θ_i of any player i:

$$U(\theta_i) \geq 0. \tag{IR}$$

Finally, to guarantee that truth-telling is indeed an equilibrium, the following incentive compatibility condition must hold:

$$U(\theta_i) \geq U(\theta_i, \hat{\theta}_i) \tag{IC}$$

for any type θ_i of any player i, and for any report $\hat{\theta}_i$.

In sum, we look for an auxiliary mechanism that maximizes (OBJ) subject to the constraints (IR), (IC), and (BB).

4 Solving the design problem

In this section we characterize the mechanism that maximizes the ex-ante social surplus (taking into account how the signal affects the ensuing voting game). While the mechanism design problem we study is not a conventional one, we nevertheless are able to reduce it to a problem that is amenable to familiar methods. In the next paragraphs we summarize the main steps that we take to achieve this, and explain how the reduced problem is solved. Readers who prefer to skip the detailed theoretical analysis, can go directly to Section 5.

In Section 4.1 we reformulate the designer's maximization problem. While mechanisms for selecting multidimensional objects are in principle difficult to solve, we show how to transform our problem into a tractable one. The key step is to show that the auxiliary design problem can be reformulated to look "almost" like a second-best public good provision problem. In this problem, for any type profile, the designer chooses the probability of departing from the status quo (which is the analog of the quantity of the public good), the division of costs, and a new variable, which is the posterior probability that departing from the status quo is the right decision (which does not have an analog in the standard public good provision problem).

In Section 4.2 we introduce a property we call "supermajority persuasiveness" (SP), which captures the requirement that any acquired signal must be instrumental for decision-making: it must have at least one realized posterior for which the collective action is different than if no signal was acquired. We then show that the optimal auxiliary mechanism necessarily satisfies this property. Consequently, supermajority persuasiveness leads to a distortion of the acquired signal relative to the socially optimal one: there are type profiles for which the optimal auxiliary mechanism chooses a signal that is different from the signal that would be chosen by a central planner, and this would be so even if the type profile were commonly known. This occurs when the central planner's signal is such that *none* of its realizations can persuade a supermajority of players to vote against the status quo. We then write the auxiliary design problem as a problem of maximizing a Lagrangian, show that it has a solution, and characterize it (Proposition 2). We show how our solution also applies to two variants of our model: one where types are commonly known (but there is no central planner) and one where players can also bargain over the collective decision (i.e., the voting assumption is relaxed). Finally, we illustrate the solution using the cost specification of mutual information (see Corollary 1).

In Section 4.3 we show that the solution to the optimal auxiliary design problem coincides with the solution to the actual design problem where players vote for the collective action independently of their reports to the mechanism (Proposition 3). This follows from the observation that truth-telling is a dominant strategy equilibrium in the optimal auxiliary mechanism (Corollary 2).

4.1 Reformulating the design problem

Fix a player i and suppose that all other players $-i$ report truthfully $\theta_{-i} \in \Theta^{n-1}$. If player i's report is such that $r_H(\hat{\theta}) \geq 1 - \hat{\theta}^{(n-m+1)}$, where $\hat{\theta} = (\hat{\theta}_i, \theta_{-i})$, then player i's net utility is given by $q(\hat{\theta}) \cdot (\theta_i - (1 - r_H(\hat{\theta}))) - t_i(\hat{\theta})$. If player i's report is such that $r_H(\hat{\theta}) < 1 - \hat{\theta}^{(n-m+1)}$ then no signal is acquired and player i's net utility is $-t_i(\hat{\theta})$. Thus, we can rewrite the net utility of type θ_i of player i who reports $\hat{\theta}_i$ when all other players report truthfully (Equation (9)) as follows:

$$\begin{aligned} U(\theta_i, \hat{\theta}_i) = & \int_{\theta_{-i} | r_H(\hat{\theta}_i, \theta_{-i}) \geq 1 - (\hat{\theta}_i, \theta_{-i})^{(n-m+1)}} q\left(\hat{\theta}_i, \theta_{-i}\right) \cdot \left[\theta_i - \left(1 - r_H\left(\hat{\theta}_i, \theta_{-i}\right)\right)\right] dF^{n-1}\left(\theta_{-i}\right) \\ & - \int_{\theta_{-i}} t_i(\hat{\theta}_i, \theta_{-i}) dF^{n-1}\left(\theta_{-i}\right). \end{aligned}$$

To express $U(\theta_i, \hat{\theta}_i)$ more compactly, we introduce the following notation. Given a report $\hat{\theta}_i$, denote by $Q(\hat{\theta}_i)$ the expected probability that the auxiliary mechanism chooses the action $a = 1$. Denote by $M(\hat{\theta}_i)$ the expected probability that the auxiliary mechanism chooses $a = 1$ but the state is $\omega = 0$ (this is the probability that the auxiliary mechanism deviates from the default action when it shouldn't). Denote by $T_i(\hat{\theta}_i)$ the expected payment of player i. Formally

$$\begin{aligned} Q(\hat{\theta}_i) &= \int_{\theta_{-i} | r_H(\hat{\theta}_i, \theta_{-i}) \geq 1 - (\hat{\theta}_i, \theta_{-i})^{(n-m+1)}} q(\hat{\theta}_i, \theta_{-i}) dF^{n-1}\left(\theta_{-i}\right) \\ M(\hat{\theta}_i) &= \int_{\theta_{-i} | r_H(\hat{\theta}_i, \theta_{-i}) \geq 1 - (\hat{\theta}_i, \theta_{-i})^{(n-m+1)}} q(\hat{\theta}_i, \theta_{-i}) \cdot \left(1 - r_H(\hat{\theta}_i, \theta_{-i})\right) dF^{n-1}\left(\theta_{-i}\right) \\ T_i(\hat{\theta}_i) &= \int_{\theta_{-i}} t_i(\hat{\theta}_i, \theta_{-i}) dF^{n-1}\left(\theta_{-i}\right) \end{aligned}$$

The expected net utility of player i with type θ_i who reports $\hat{\theta}_i$ is then given by:

$$U\left(\theta_i, \hat{\theta}_i\right) = Q(\hat{\theta}_i) \cdot \theta_i - M(\hat{\theta}_i) - T_i(\hat{\theta}_i) \tag{10}$$

Note that our specification of the players' utility has the convenient feature that it is *as if* a player gets a payoff of θ_i every time the collective action 1 is chosen, but he pays a penalty $(M(\hat{\theta}_i))$ that is equal to the probability that this is the *wrong* collective action.

The designer's objective function (OBJ) can therefore be written as

$$\sum_{i=1}^{n} \int_0^{1-p} \left[Q\left(\theta_i\right) \cdot \theta_i - M\left(\theta_i\right) - T_i\left(\theta_i\right)\right] dF(\theta_i) \tag{11}$$

while incentive compatibility (i.e., Equation (IC)) requires

$$Q(\theta_i) \cdot \theta_i - M(\theta_i) - T_i(\theta_i) \geq Q(\hat{\theta}_i) \cdot \theta_i - M(\hat{\theta}_i) - T_i(\hat{\theta}_i)$$

for all $\hat{\theta}_i$ and θ_i and every player i. Note that $U(\theta_i)$ is the upper envelope of a family of affine functions in θ_i, and is therefore convex. It follows that an auxiliary mechanism satisfies incentive compatibility if and only if $Q(\theta_i)$ is nondecreasing and $U(\theta_i) = \int_0^{\theta_i} Q(x)\,dx - M(0) - T_i(0)$ (see, e.g., Krishna, 2010, p. 64).

The above argument allows us to express the auxiliary mechanism design problem in the following compact form:

Lemma 2 *The auxiliary design problem consists of finding $q(\theta)$ and $r_H(\theta)$ that maximize the aggregate surplus,*

$$\sum_{i=1}^{n} \int_0^{1-p} \left[\theta_i \cdot Q(\theta_i) - M(\theta_i)\right] dF(\theta_i) - \int_\theta c(q(\theta), r_H(\theta)) dF^n(\theta), \tag{12}$$

subject to the following constraints: (i) $Q(\theta_i)$ is monotone and (ii) the aggregate virtual surplus *is nonnegative,*

$$\sum_{i=1}^{n} \int_0^{1-p} \left[v(\theta_i) \cdot Q(\theta_i) - M(\theta_i)\right] dF(\theta_i) - \int_\theta c(q(\theta), r_H(\theta)) dF^n(\theta) \geq 0. \tag{13}$$

This inequality is both necessary and sufficient for individual rationality and ex-ante budget balancedness.

The monotonicity of $Q(\theta_i)$ is necessary for incentive compatibility, while the nonnegativity of the aggregate virtual surplus follows from the IR constraint after we impose budget balancedness and employ transfers that induce incentive compatibility. For further details, see the proof in the appendix.

We refer to an auxiliary mechanism that employs the functions $q(\theta)$ and $r_H(\theta)$, which solve the design problem described in the lemma, as an *optimal auxiliary mechanism.*

We have therefore transformed the design problem of acquiring the (ex-ante) welfare-maximizing signal and covering its cost into a problem of choosing a welfare-maximizing public good and covering its cost, but with the following "twists". First, the public good is multidimensional: it is a distribution over posterior beliefs that can be summarized by a pair of numbers, the high posterior r_H, and the probability q of realizing it. Second, unlike in a conventional public good provision problem, the characteristics of the public good affect the players' actions in a game that is played after the good is provided. Third, also unlike in a conventional public good provision problem, the players do not necessarily agree on the ranking of (noisy) signals, even when the cost is ignored. This follows from the players' possible disagreement on the optimal collective action for each realized posterior belief. Finally, in a conventional public good provision problem, the cost of the optimal level of the public good increases in types. This is not necessarily true in our setup: even when the types are known, the cost of the optimal signal is not necessarily monotone in the types.

4.2 Characterizing an optimal auxiliary mechanism

Assigning a type profile θ to an informative signal that will not result in a supermajority vote for $a = 1$ no matter what posterior is realized is wasteful: the players incur a cost, but do not change their behavior relative to having no signal. We therefore introduce the following property:

Definition 1 (supermajority persuasiveness, SP) *Given m, a signal (q, r_H) is supermajority persuasive (SP) for the type profile θ if $q \in (0, p/r_H]$ and $r_H \geq 1 - \theta^{(n-m+1)}$. An auxiliary mechanism is SP if almost every informative signal that it acquires is SP.*

We then have that

Lemma 3 *Every optimal auxiliary mechanism is SP.*

While it may seem intuitive that an optimal mechanism should be SP, note that, in principle, a mechanism can achieve ex-ante optimality by committing to suboptimal interim actions (as in Myerson and Satterthwaite, 1983). This is because suboptimal actions can lower the players' incentives to misreport their types, and hence lower the information rents required to support truth-telling. However, in our model, information rents depend solely on the distribution over outcomes and not on the (cost of the) acquired signal. Thus, not buying a signal induces the same distribution over outcomes as buying a noninstrumental signal. It follows that buying a noninstrumental signal wastes resources but is not helpful in decreasing information rents.

By Lemma 3, an optimal auxiliary mechanism solves the constrained optimization problem defined in Lemma 2 subject to an additional constraint that the mechanism is SP. If for a particular type profile θ the acquired signal satisfies that $r_H(\theta) = 1 - \theta^{(n-m+1)}$ (i.e., the "marginal" decisive voter is indifferent between the two collective actions), we say that the signal is *minimally supermajority persuasive* (MSP) for that profile, or that at the profile θ, the SP constraint is binding.

Since in an optimal auxiliary mechanism $q(\theta) = 0$ whenever $r_H(\theta) < 1 - \theta^{(n-m+1)}$, we can simplify the expressions of $Q(\theta_i)$ and $M(\theta_i)$ as follows:

$$Q(\theta_i) = \int_{\theta_{-i}} q(\theta_i, \theta_{-i}) dF^{n-1}(\theta_{-i}) \tag{14}$$

$$M(\theta_i) = \int_{\theta_{-i}} (1 - r_H(\theta_i, \theta_{-i})) \cdot q(\theta_i, \theta_{-i}) dF^{n-1}(\theta_{-i}) \tag{15}$$

To solve the design problem given in Lemma 2, we start by ignoring the monotonicity constraint on $Q(\cdot)$, and later verify that the solution to this relaxed problem actually satisfies this ignored constraint.

Finding the two functions $q(\cdot)$ and $r_H(\cdot)$ that maximize the social surplus (12) is, formally, a problem of finding the maximum of a mapping from the domain of (pairs of) integrable functions to the reals. In addition, the problem includes a constraint that the aggregate surplus be nonnegative (13). Since this constraint takes the form of an integral, the problem can be solved as an isoperimetric one, using techniques from the calculus of variations (see, e.g., Kamien and Schwartz, 2012, part I, section 7, p.48). Specifically, it is possible to append the constraint to the objective with a Lagrange multiplier, and derive the necessary conditions for the optimum by maximizing the augmented integral (just as in a "standard" calculus optimization problem). By the Lagrange sufficiency theorem, these conditions are also sufficient, provided that a feasible solution exists (for details see the proof of Proposition 2).

The Lagrangian associated with maximizing Equation (12) under the constraint in Equation (13) is given by[16]

$$\mathcal{L} = \int_{\theta} \left[w(\theta, \lambda) \cdot q(\theta, \lambda) - (1 - r_H(\theta, \lambda)) \cdot q(\theta, \lambda) - \frac{1}{n} \cdot c(q(\theta, \lambda), r_H(\theta, \lambda)) \right] dF^n(\theta) \tag{16}$$

where

$$w(\theta, \lambda) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{1+\lambda} \theta_i + \frac{\lambda}{1+\lambda} v(\theta_i) \right). \tag{17}$$

Our assumption that the distribution F is regular ensures that for all λ, the function $w(\theta, \lambda)$ is increasing in each component of θ.

Notice that the Lagrangian in Equation (16) is independent of the derivatives of $q(\cdot)$ and $r_H(\cdot)$. Thus, the Euler equation that solves the problem of maximizing (16) is not a differential equation, and therefore the between-type-profiles aspects of the optimization problem are

[16] To obtain the Lagrangian $\mathcal{L}(\lambda)$, write the aggregate surplus given by Equation (12) plus λ times the aggregate virtual surplus given by Equation (13). Now plug in the expressions for $Q(\theta_i)$ and $M(\theta_i)$ given by Equations (14) and (15) and divide by $(1+\lambda) \cdot n$.

degenerate.[17] Thus, the problem can be solved pointwise, i.e. by maximizing the integrand at each θ (see also Kamien and Schwartz, 2012, part I, section 5, p. 34).

Fix a profile of types θ and a multiplier λ. The part of the Lagrangian (16) that is affected by $q(\theta,\lambda)$ and $r_H(\theta,\lambda)$ is

$$\hat{\mathcal{L}}\left(q, r_H\ ;\ w\right) = q(r_H - (1-w)) - \frac{1}{n} \cdot c(q, r_H) \tag{18}$$

where r_H, w and q are used for brevity instead of $r_H\left(\theta,\lambda\right)$, $w(\theta,\lambda)$, and $q\left(\theta,\lambda\right)$. Note that λ and θ affect the values of the maximizers q and r_H only through w. Differentiating $\hat{\mathcal{L}}\left(q, r_H\ ;\ w\right)$ with respect to q and r_H and equating to zero yields

$$\hat{\mathcal{L}}_1\left(q, r_H\ ;\ w\right) = r_H - (1-w) - \frac{1}{n} \cdot c_1\left(q, r_H\right) = 0, \tag{FOCq}$$

$$\hat{\mathcal{L}}_2\left(q, r_H\ ;\ w\right) = q - \frac{1}{n} \cdot c_2\left(q, r_H\right) = 0, \tag{FOCr}$$

where $c_1(q, r_H)$ is the derivative of the function $c(q, r_H)$ with respect to its first argument q, and $c_2(q, r_H)$ is the derivative of $c(q, r_H)$ with respect to its second argument r_H.

Given θ and λ, denote by $(\tilde{q}(\theta,\lambda), \tilde{r}_H(\theta,\lambda))$ the signal that solves (FOCq) and (FOCr), if such a signal exists. We say that $(\tilde{q}(\theta,\lambda), \tilde{r}_H(\theta,\lambda))$ is *interior* if $\tilde{q}(\theta,\lambda) \in (0, p/r_H)$ and $\tilde{r}_H(\theta,\lambda) \in (p, 1)$. Recall that $(\tilde{q}(\theta,\lambda), \tilde{r}_H(\theta,\lambda))$ is SP if $\tilde{r}_H(\theta,\lambda) \geq 1 - \theta^{(n-m+1)}$ and $\tilde{q}(\theta,\lambda) \in (0, p/r_H)$.

We provide a characterization of optimal mechanisms for a class of cost functions $c\left(q, r_H\right)$ that satisfy the following properties:[18]

P1 The cost and marginal costs are all increasing in both q and r_H. The marginal cost of r_H is convex in q.

P2 The marginal cost of achieving certainty in either $\omega = 0$ or $\omega = 1$ is at least n.

P3 For any w there is at most one solution to (FOCq) and (FOCr).

[17] In the language of the calculus of variations, the problem is not a dynamic one.

[18] Formally, we require the following properties: (P1) $c_1 > 0, c_2 > 0, c_{11} > 0, c_{22} > 0, c_{12} > 0, c_{211} > 0$, (P2) $c_1\left(\frac{p}{r_H}, r_H\right) > n$ for any $r_H > p$, and $c_2\left(q, 1\right) > n$ for any $q \in (0,1)$, and (P3) as stated in the text. A sufficient condition for (P3) to hold is that $c_{11}c_{22} - \left(1 - c_{12}\right)^2 > 0$ whenever $q = c_2\left(q, r_H\right)$, which is the (FOCq) condition. Alternatively, this condition holds if the right-hand side of Equation (18)) is quasiconcave. If $c\left(\{q_j, r_j\}_{j=1}^J\right) = \sum_j q_j h\left(r_j\right)$ for some $h : [0,1] \to \mathbb{R}^+$, then (P3) is satisfied when h is increasing and convex.

It is easy to verify that these properties are satisfied by the cost function that is proportional to the mutual information between the signal and the state (given by Equation (2)) or to the expected variance of the induced posterior beliefs.

Proposition 2 *Assume that the cost function c satisfies $(P1)$-$(P3)$. Then there exists $\lambda^* \geq 0$ for which an optimal auxiliary mechanism $q^*(\theta), r_H^*(\theta)$ is characterized as follows:*

- *If $(\tilde{q}(\theta,\lambda^*), \tilde{r}_H(\theta,\lambda^*))$ is not interior, then no information is acquired for the profile θ.*
- *If $(\tilde{q}(\theta,\lambda^*), \tilde{r}_H(\theta,\lambda^*))$ is interior and SP, then $q^*(\theta) = \tilde{q}(\theta,\lambda^*)$ and $r_H^*(\theta) = \tilde{r}(\theta,\lambda^*)$.*
- *If $(\tilde{q}(\theta,\lambda^*), \tilde{r}_H(\theta,\lambda^*))$ is interior but is not SP then $r_H^*(\theta) = 1 - \theta^{(n-m+1)}$ and $q^*(\theta)$ is determined according to (FOCq), provided that the solution is interior. Otherwise, no signal is acquired.*

Finally, $r_H^*(\theta)$ is decreasing in each player's type, and $q^*(\theta)$ is increasing in each player's type.

The proposition addresses the following special cases.

The case of commonly known types and/or no participation constraints. The optimal auxiliary mechanism when types are commonly known is obtained by setting $\lambda^* = 0$. To see this, note that when $\lambda = 0$ the problem of maximizing the Lagrangian in Equation (16) reduces to maximizing the aggregate surplus as given in (12) subject only to the SP constraint. Under complete information, a signal is purchased only when it creates a positive social surplus and, therefore, the cost of the signal can always be covered, and the excess surplus can be redistributed to satisfy all the players' participation constraints.

The solution that is obtained when $\lambda^* = 0$ is optimal also if types are private and players are obligated to participate in the bargaining. The reason is that under incomplete information, it follows from standard arguments (see, e.g., Borgers, 2015) that if there are no participation constraints, there exists a payment schedule that induces truth-telling and allows the group to implement the socially optimal outcome of the complete information case.[19]

The case of a contractible collective action. The case in which the mechanism can choose the signal and the action with no voting is equivalent to the case in which a single vote for $a = 1$ is enough to make that decision (i.e., $m = 1$). To see this, notice that by (FOCq), if an interior signal is purchased, then it is always the case that at least one player prefers the

[19] Specifically, $T(\theta_i) = T(0) + Q(\theta_i) \cdot \theta_i + M(0) - M(\theta_i) - \int_0^{\theta_i} Q(x)\, dx$, where Q and M are given by (14) and (15), respectively, where $q(\cdot)$ and $r_H(\cdot)$ are characterized by Proposition 2 when $\lambda^* = 0$, and $T(0)$ is chosen to satisfy budget balancedness (since we don't need to worry about participation constraints).

action $a = 1$ when the high posterior r_H is realized.[20] Under $m = 1$, this means that every interior solution is supermajority persuasive and therefore the SP constraint is never binding at the optimum. Thus, when $m = 1$ the SP constraint can be ignored. It is easy to verify that solving the problem without the SP constraint is effectively the same as solving the problem when actions are contractible (in this case, too, any optimal signal is always instrumental, but the mechanism can choose the action $a = 1$ after a realization of the posterior r_H, regardless of its value.)

An immediate corollary of Proposition 2 is that when the cost function is proportional to the mutual information between the signal and the state, an optimal auxiliary mechanism can be characterized as follows.[21]

Corollary 1 *Suppose that the cost function is defined by Equation (2). Let λ^* be the constant defined in Proposition 2. Then, in the corresponding optimal auxiliary mechanism, we have*[22]

$$r_H^*(\theta) = \max\left\{\frac{e^{\frac{n}{\kappa}} - e^{\frac{n}{\kappa}w(\theta,\lambda^*)}}{e^{\frac{n}{\kappa}} - 1}, 1 - \theta^{(n-m+1)}\right\}. \tag{19}$$

Next, $r_L^*(\theta)$ is determined such that $D_{KL}(r_H^*(\theta), r_L^*(\theta)) = \frac{n}{\kappa}[r_H^*(\theta) - (1 - w(\theta,\lambda^*))]$ provided that a solution exists and is in $(0, p)$; otherwise, $r_L^*(\theta) = p$. If $r_H^*(\theta) > 1 - \theta^{(n-m+1)}$ then $r_L^*(\theta)$ is given by

$$r_L^*(\theta,\lambda^*) = \min\left\{\frac{e^{\frac{n}{\kappa}(1-w(\theta,\lambda^*))} - 1}{e^{\frac{n}{\kappa}} - 1}, p\right\}. \tag{20}$$

Finally,

$$q^*(\theta) = \frac{p - r_L^*(\theta)}{r_H^*(\theta) - r_L^*(\theta)}. \tag{21}$$

[20] To see this formally, note that when $m = 1$, then if $(\tilde{q}(\theta,\lambda^*), \tilde{r}_H(\theta,\lambda^*))$ is interior, it follows from (FOCq) that $\tilde{r}_H(\theta,\lambda^*) > 1 - w(\theta,\lambda^*)$. However, because $w(\theta,\lambda^*) \leq w(\theta,0) \leq \theta^{(n)}$ it follows that $\tilde{r}_H(\theta,\lambda^*) > 1 - \theta^{(n)}$, implying that SP is never violated when a signal that is socially optimal is acquired.

[21] As in the central-planner benchmark, it is still true that as $\kappa \to 0$, the optimal signal to acquire for each type profile converges to the fully informative signal. However, since the value of the Lagrange multiplier λ^* changes with κ, it remains an open question whether the acquired signal for a type profile necessarily becomes more Blackwell informative as κ decreases.

[22] We do not restrict $r_H(\theta)$ to be at most one if no signal is acquired (i.e., if $q(\theta) = 0$). Indeed, $r_H^*(\theta) > 1$ whenever $w(\theta,\lambda) < 0$. But in this case, $r_L^*(\theta) = p$, and hence $q(\theta) = 0$. Also notice that since $\theta_i \leq 1 - p$, it follows that $1 - \theta^{(n-m+1)} \geq p$.

Proposition 2 established that $q^*(\theta)$ is increasing in each of its components. An immediate corollary of this is the following (Mookherjee and Reichelstein, 1992):[23]

Corollary 2 *There exists an optimal auxiliary mechanism that solves the problem stated in Lemma 2 in which truth-telling is a dominant strategy equilibrium.*

4.3 From the auxiliary mechanism to the actual mechanism

Up to now we have analyzed an auxiliary mechanism, that operates as if the players commit to vote on the collective action according to their reported types. In such a mechanism, when a player considers misreporting, he takes into account that in the subsequent voting game his vote will not be cast according to his true preferences but rather according to his reported ones. For instance, if player i of type θ_i reports that his type is $\theta_i' > \theta_i$, and a posterior r_H between $1 - \theta_i'$ and $1 - \theta_i$ is realized, then the mechanism votes for $a = 1$ on behalf of the player even though the player actually prefers the action $a = 0$ according to the information he possesses and his true type. This fact may discourage players from misreporting their types in the auxiliary mechanism. By contrast, in an actual mechanism a player is free to vote according to his true preferences. In that case, a player may have an incentive to affect the choice of the signal, knowing that he can vote in favor of his truly preferred action in the ensuing voting game.

Consider now an actual mechanism that employs the functions $q^*(\cdot)$ and $r_H^*(\cdot)$ that were characterized in Proposition 2. Does the mechanism remain incentive compatible, budget balanced, and individually rational even though the players vote on the collective action according to their true preferences? Our next result shows that this is indeed the case.

Proposition 3 *The solution to the actual mechanism design problem coincides with the solution to the auxiliary mechanism design problem.*

The proof hinges on the fact that there exists an optimal auxiliary mechanism, characterized by Proposition 2, that is SP and for which truth-telling is a dominant strategy. We show that if a player wants to misreport in the actual mechanism, but not in the auxiliary mechanism, it must be the case that after r_H is realized the player finds it beneficial to vote for the status quo ($a = 0$), whereas the auxiliary mechanism would have voted on his behalf for taking the action ($a = 1$). However, it can be shown that in this case the player can profitably deviate in the auxiliary mechanism by reporting that he is of the lowest type.

[23] The equivalence between Bayesian and dominant incentive compatibility that we use here is defined in terms of the *ex-post allocation*. Mookherjee and Reichelstein (1992) show that this equivalence fails unless the ex-post allocation rule (which corresponds to q in our environment) is monotone in each of its coordinates. Gershkov et al. (2013) use a more permissive notion of equivalence, which considers the interim expected utilities of the players. They then show that this form of equivalence between Bayesian and dominant incentive compatible mechanisms holds whenever players have linear utilities and independent, one-dimensional, private types.

5 Properties of the optimal signal structure

Proposition 2 in Section 4.2 characterizes the optimal signal acquired by the group. This characterization allows us to describe features of the optimal signal structure, and also to explain the effect of incomplete information and a supermajority requirement on the voting stage.

The mapping from types to signals. The characterization in Proposition 2 highlights three types of outcomes of the bargaining over information: acquiring no signal, acquiring an "interior" signal, and acquiring an MSP signal. As defined in Section 4, an interior signal solves the first-order conditions of the design problem (given by Equations (FOCq) and (FOCr) in Section 4.2). That is, this signal equates the marginal social benefit from increasing q and r_H with their marginal costs. A signal is MSP (minimally supermajority persuasive) for the type profile θ if the high posterior r_H just ensures a supermajority support for $a = 1$. That is, r_H is equal to $1 - \theta^{(n-m+1)}$. Note that such a signal does not necessarily equate the marginal benefit with the marginal cost.

These outcomes are summarized in the following proposition. To simplify the statement of the proposition, we slightly strengthen the assumptions on the cost function, and require that the marginal cost with respect to r_H be infinite when $r_H = 1$ and $q > 0$, and that the marginal cost with respect to q be finite when $q = 0$ and $r_H = 1$.[24]

Proposition 4 *Suppose that the cost function is $c(q,r) = \kappa \cdot \hat{c}(q,r)$ for some fixed function $\hat{c}(\cdot,\cdot)$, and that it satisfies $(P1)$-$(P3)$ for all κ. Suppose further that $c_2(q,1)$ is infinite for every $q > 0$ and that $c_1(0,1)$ is finite. Then, for any type distribution F and for any $m > 1$, there exists κ^* such that for all $\kappa < \kappa^*$, an optimal mechanism partitions the set of all type profiles into three subsets:*

i. No signal*: a positive measure subset of type profiles for which no signal is acquired,*

ii. Interior signal*: a positive measure subset of type profiles for which the signal is given by the solutions to Equations (FOCq) and (FOCr),*

iii. MSP signal*: a positive measure subset of type profiles for which an MSP signal is acquired.*

If $\kappa > \kappa^*$, subset (i) is of positive measure, while subsets (ii) and (iii) may be empty.

[24] These stronger assumptions guarantee that the set of type profiles for which MSP signals are acquired has a non-zero measure – signals with $r_H = 1$ are not too costly, nor too cheap, to acquire. These conditions are satisfied by our leading cost specification given by Equation (2), in which case $c_1(0,1) = -\kappa \ln(p)$ and $c_2(q,1) = \infty$ for all $q > 0$.

To illustrate the projection of the above three cases onto the players' type space, suppose that there are two players who need a unanimous vote in order to depart from the status quo and choose $a = 1$ (i.e., $m = n = 2$). Proposition 4 highlights the following features of an optimal signal, which are also depicted in Figure (2b) below. When the sum of the players' types is small, no signal is purchased. Indeed, if the aggregate social harm from incorrectly keeping the status quo is small, purchasing a costly signal is inefficient. Conversely, when the sum of types is large, it is optimal to purchase a signal. Intuitively, this is because the social harm of incorrectly keeping the status quo is large and because convincing the players that the action $a = 1$ is optimal is relatively easy. When one type is high and the other is low, a signal is purchased but its characteristics must be distorted in order to just pass the unanimity requirement. This is because the high posterior r_H must be sufficiently large so as to convince the low type to agree to vote for $a = 1$ upon the realization of r_H.

The effect of asymmetric information. To understand the distortions that are due to the fact that players' types are privately known, it is instructive to compare the characterization of optimal signals in Proposition 2 in the two cases of complete and incomplete information. Let $(q^c(\theta), r^c_H(\theta), r^c_L(\theta))$ denote the mapping from commonly known type profiles to signals that maximizes the total ex-ante social surplus (OBJ) such that the acquired signal is SP. We refer to this as the optimal acquisition rule under complete information. Recall that this acquisition rule is obtained by letting $\lambda^* = 0$. Since $w(\theta, \lambda)$ (as defined in Equation (17)) is increasing in each θ_i, decreasing in λ, and $w((1-p, ..., 1-p), \lambda) = 1 - p$ for any λ, it follows that

$$w(\theta, \lambda^*) < w(\theta, 0) \leq w((1-p, ..., 1-p), 0) = 1 - p$$

Since $w(\theta, \lambda^*)$ is continuous in each θ_i, there exists $\theta' > \theta$ (i.e., $\theta'_i \geq \theta_i$ for all i, with at least one strict inequality) such that $w(\theta', \lambda^*) = w(\theta, 0)$. This has the following implications.

Observation 1. If θ is such that under both complete and incomplete information a signal is acquired and the SP constraint is slack, then $q^c(\theta) > q^*(\theta)$ and $r^c_H(\theta) < r^*_H(\theta)$.

Since $q(\theta)$ is the probability of taking the non-default action, this observation means that under incomplete information this action will be taken with a *lower* probability. However, since $r^c_H(\theta) < r^*_H(\theta)$, it follows that whenever the non-default action is taken under incomplete information, it is taken with *greater* confidence.

An intuition for the above observation stems from how a signal is distorted to incentivize players to reveal their type. From the literature on trade with asymmetric information, we know that to incentivize agents to reveal their types, the probability of trade must be distorted downward. Since a signal is a multidimensional object, a priori it is not clear how information will be distorted to incentivize truth-telling. Our characterization shows that the distortion occurs by lowering the probability that $a = 1$ will be taken, i.e., by lowering q.

However, in contrast to a standard mechanism design problem of allocating an asset, since the mechanism in our environment (or the bargaining) decides on a signal, it can "compensate" for the distortion in the probability of taking $a = 1$ (i.e., q) by changing the likelihood that this is the correct action whenever it is taken (i.e., by changing r_H). In the family of cost functions that we focus on, both the marginal cost and the marginal benefit of r_H decrease when q is distorted downward relative to the unconstrained socially optimal value. However, since the marginal cost of r_H is convex in q (see (P1) above), whereas the marginal benefit is linear in q, in order to satisfy the first-order condition (FOCr), r_H must increase (relative to the complete information case).

A related intuition concerns the ability of low types to veto information acquisition. Lowering the probability that the signal will be acquired, and raising the confidence that $a = 1$ is the right action (conditional on acquisition), makes the low types more inclined to participate in the bargaining over information.

Observation 2. *If $q^c(\theta) = 0$ then $q^*(\theta) = 0$, but the converse is not true.*

Put differently, there are realizations of θ for which information is acquired under complete information but not under incomplete information. Hence, the fact that players do not observe each other's type can lead to *under-provision of information* for the collective decision, which implies greater "conservatism" in the sense of being less likely to depart from the status quo. This again follows from the downward distortion in q that is necessary to induce players to reveal their type. Consequently, for some type profiles θ for which $q^c(\theta)$ is relatively low, this probability will decrease to zero when types are private.

Observation 3. *A type profile is assigned an MSP signal under incomplete information, only if it is assigned one under complete information.*

The fact that players vote after they observe the realization of the acquired signal introduces an ex-ante distortion even when players' types are commonly known. This occurs when the signal (q, r_H) that maximizes ex-ante welfare satisfies $q > 0$ and $p < r_H < 1 - \theta^{(n-m+1)}$. In this case, the acquired signal will be distorted such that r_H will increase to $1 - \theta^{(n-m+1)}$. Observation 3 establishes that introducing private types does *not* exacerbate this distortion. There are two reasons for this. First, some type profiles that are assigned a signal that just satisfies the SP constraint when the profile is commonly known may be assigned no signal when types are privately known. This follows from the downward distortion in q (relative to the unconstrained socially optimal signal), which arises under incomplete information in order to induce truth-telling. Since the SP constraint typically has bite when some players have low types, it means that for these profiles, q^c is relatively low (since it increases in each player's

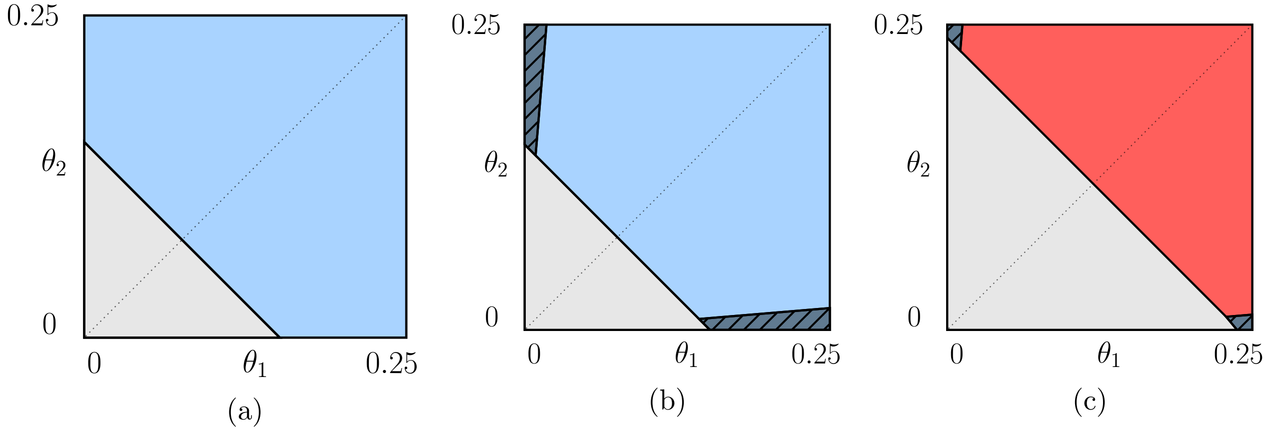

Figure 2: Optimal signals for two players. (a) The case of complete information and $m = 1$. (b) Complete information and $m = 2$. (c) Privately known types, uniformly distributed over [0,0.25] and $m = 2$. For type profiles in grey regions, no signal is acquired; in blue regions, a socially optimal signal is acquired; in hatched regions, the signal is distorted to be minimally supermajority-persuasive; in the red region, the signal is distorted due to the asymmetric information.

type). Hence, for sufficiently low q^c, the downward distortion introduced by private types may drive this probability all the way down to zero. The second reason is a consequence of Observation 1: for each type profile $r_H^* > r_H^c$ and, therefore, for some profiles where the SP constraint is binding under complete information, the increase in r_H slackens the constraint under incomplete information.

We illustrate the above differences between the complete and incomplete information environments for the case of two players who need to agree unanimously on $a = 1$ (i.e., $n = m = 2$), whose types are independently drawn from a uniform distribution on $[0, 0.25]$, and whose prior probability p is 0.75. The cost function is assumed to be proportional to the mutual information between the state and the signal, as defined in Equation (2), with $\kappa = 0.55$. We use this example also to illustrate the effect of allowing players to commit to whatever collective action they choose for each signal realization, which, as explained above, is obtained by setting $m = 1$.

Figure (2a) illustrates the complete information case with $m = 1$. There are two regions in this figure. The grey region represents the type profiles for which no signal is acquired, while the blue region represents the types profiles for which the socially optimal signal is purchased. Notice that since $m = 1$, supermajority persuasiveness has no bite. It follows that the case represented by this figure exhibits no distortions relative to the (unconstrained) socially optimum.

Figure (2b) illustrates the complete information case with $m = 2$. Here, a third region emerges: the hatched area represents the type profiles for which the (unconstrained) socially optimal signal is not supermajority persuasive, but acquiring an MSP signal is better than no information at all. Notice that this occurs for profiles in which the average type is above some threshold but the minimal type is small.

Finally, Figure (2c) corresponds to the case of asymmetric information with $m = 2$. This figure depicts the observations described above: the region in which no signal is acquired under asymmetric information contains the corresponding region under complete information. The regions with MSP signals under asymmetric information are contained in the corresponding regions under complete information. The red region represents type profiles for which the signal is distorted due to asymmetric information; i.e., relative to the case of complete information, q is lower and r_H is higher.

The diagnostic odds ratio of the acquired signal. In our environment, a signal is essentially a classification test: either a reform is needed ($\omega = 1$) or not ($\omega = 0$). As in any imperfect classification test, there are false positives (choosing the reform when it should not be chosen), and there are false negatives (sticking with the status quo when a reform is required). How should one evaluate the "quality" of the signal as a diagnostic tool for checking when the group should adopt the reform? The literature on clinical testing (see, e.g., Glas et al., 2003) has proposed the Diagnostic Odds Ratio (DOR) as a possible one-dimensional indicator of diagnostic performance that takes into account both forms of false results. More specifically, the DOR measures the ratio between the odds of positive results when a reform should be chosen (r_H/r_L) and the odds of positive results when a reform should not be chosen ($(1 - r_H)/(1 - r_L)$).

Should the quality of a test depend on the particular stochastic realization of the types? In our leading specification, where the cost is measured as the mutual information between the state and signal, as defined in Equation (2), Corollary 1 implies that for interior solutions the signal's quality (as measured by DOR) is *independent* of the type profile and is equal to $e^{n/\kappa}$. Thus, the more members, the better the quality. And, the greater the cost scaler κ, the lower the quality.

The effect of the supermajority rule. In our model, players do not have private information on the state ω (see also Section 6.3). Thus, voting on the collective action does not serve as a means of aggregating information (notice the difference from settings in which the players are privately informed about the state, in which case more votes typically contribute to the probability that the right decision is taken). Consequently, a higher supermajority requirement only toughens the supermajority-persuasiveness constraint. In particular, as explained above, when $m = 1$, this constraint is never binding. This makes it more difficult for the group to purchase a persuasive signal. Hence, in some cases the group foregoes opportunities to depart from the status quo, which could have raised its welfare. This implies the following:

Observation 4. *The socially optimal gain from information is nonincreasing in the supermajority requirement m.*

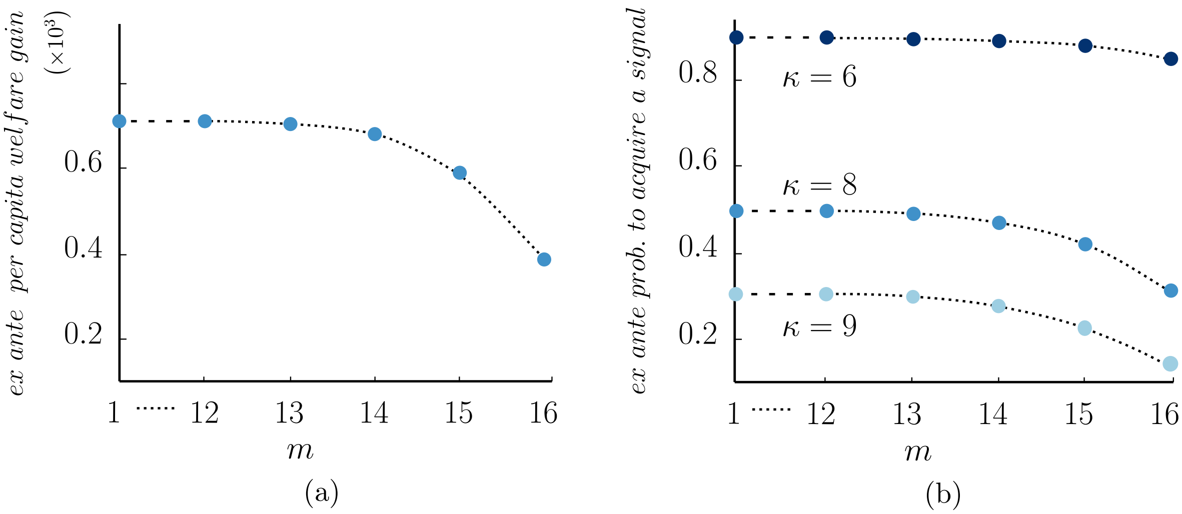

Figure 3: Effects of the cost scaler κ and the supermajority requirement m

This observation is illustrated for the case of commonly known types in panel (a) of Figure 3. The figure assumes $n = 16$ players whose types are uniformly distributed over $[0, 0.3]$, a prior probability $p = 0.7$ that the state is $\omega = 1$, and a cost function given by Equation (2) with $\kappa = 8$. Panel (b) depicts the ex-ante probability of acquiring information (i.e., the probability mass of type profiles for which $q^c(\theta) > 0$) as a function of m for different values of κ . As is evident from the figure, for larger m and for larger κ, the ex-ante probability of acquiring information decreases.

While the monotonicity exhibited in panel (a) remains even if types are not commonly known, the monotonicity in panel (b) is not guaranteed for intermediate values of m if types are private information. This is because changes in the required supermajority m may affect the probability of acquiring information through the value of λ^* in subtler ways. It is noteworthy, however, that at the extreme, if unanimity is required ($m = n$) and n is sufficiently large, then for almost all type profiles the minimally supermajority-persuasive signal requires r_H to be close 1. If the cost function is such that purchasing a signal that fully reveals $\omega = 1$ (i.e., $r_H = 1$) is too costly, then no signal will be acquired for almost all type profiles. In our leading cost specification (Equation (2)), this is true for κ sufficiently large.

Finally, the supermajority requirement affects the likelihood that deviating from the status quo is the right decision, conditional on making that decision. By Proposition 2, when types are commonly known (i.e., $\lambda^* = 0$), a tougher SP constraint immediately implies that either information is acquired with a weakly higher r_H, or information is not acquired if the higher r_H that the SP constraint requires is too costly. When types are not commonly known, the effect is again more subtle and depends on the distribution of types because a change in m can also affect the value of λ^*.

6 Discussion

In this section we discuss some key features of our model.

6.1 Key features of the preference specification

We start by examining which features of our preference specification are necessary for tractability, and which play a key role in the analysis of optimal mechanisms. Virtually all works on voting that involve costs or transfers assume binary actions and states, unidimensional types, and quasilinear preferences (see, e.g., Dal Bó, 2007; Dekel, Jackson, and Wolinksy, 2008; Casella, Llorente-Saguer, and Palfrey, 2012; Drexl and Kleiner, 2018). In particular, unidimensionality is important for our mechanism design approach to the bargaining over information since mechanism-design with multidimensional types is notoriously difficult. Since there are four different utility numbers (corresponding to all possible combinations of the action and the state), some normalization is needed to reduce the types to a single dimension.

Our normalization involves two assumptions: (i) the payoff from matching the state is independent of the state (and higher than the payoff from a mismatch) and (ii) the payoffs from a mismatch sum up to a constant. It is the second assumption that allows us to apply Myersonian techniques to characterize the socially optimal (budget balanced and individually rational) bargaining outcomes. To see this, denote a player's payoff from $(a = 1, \omega = 0)$ by x, and the payoff from $(a = 0, \omega = 1)$ by y. As we showed in Section 4.1, a player's gain from participating in the bargaining, relative to having no information, is equal to $q[x - 1 + (2 - x - y)r_H]$, where q is the probability of choosing $a = 1$. If $x + y$ is a constant, then x can represent a player's type, and we obtain that in the expression for the player's gain, the type multiplies only q. This implies that the mapping from types to signals and cost shares is incentive compatible only if the interim expected value of q is monotone in the player's type.

Our analysis can also be carried out with alternative normalizations that yield a unidimensional type space. For example, suppose that all players get a payoff of 1 when the action matches the state, and all players get a payoff of zero when $a = 1$ but $\omega = 0$. Define a player's type t to be his payoff when $a = 0$ but $\omega = 1$, where $t \in (-\infty, 1)$. Then type t's gain from the bargaining, relative to having no information, is equal to $q[-1 + (2 - t)r_H]$. By making a change of variables such that $x := q \cdot r_H$ and $y := q$ (and letting $z := (p - x)/(1 - q)$ in order to capture r_L), the expression for type t's gain becomes $-tx + 2x - y$. Note that in this formulation, a player's type multiplies the variable x. Incentive compatibility would then require the interim expected value of x to be monotone in a player's type. It can be shown that under our leading cost specification (2), at an optimal solution, r_L is decreasing in each player's type, while q is increasing, and since $r_L = (p - q \cdot r_H)/(1 - q)$, it follows that x is monotone in types. Hence, we can also apply the Myersonian techniques to this payoff specification.

6.2 The participation constraint

Recall that in our model, a player who opts out of bargaining effectively vetoes the provision of a public signal. We make this assumption for two reasons. First, this veto-power assumption is typical in almost all public good settings (e.g., Mailath and Postlewaite, 1990; Hellwig, 2003). Second, from a more technical point of view, this assumption implies that the value of the outside option for each player does not depend on the types and actions of other players (notice, however, that the value does depend on the type of the player who opts out). While obviously there are situations that do not fit this assumption, there are many others that do.

One can of course think of other assumptions about the players' outside options. However, when considering such alternative assumptions, it is important to notice that in the novel bargaining problem that we study, it is not obvious what a player should expect to get if he does not participate in the bargaining over the signal. For example, if a quitter cannot prevent the others from acquiring a signal, does he observe the realization? Does he vote on the collective decision? Can he be excluded from the consequences of the ultimate collective decision? While each of these alternatives may be a reasonable description of some "real-world" scenario, none of them seem to be universally true. However, each of these alternatives introduces new challenges to the analysis. For example, even if the quitter cannot prevent the others from acquiring information and cannot participate in the voting (but he does enjoy the consequences of the decision), then when considering whether to quit he needs to take into account the equilibrium outcome of the game without him. This means that the participation constraint is determined endogenously in equilibrium. Thus, our paper opens the door to many interesting questions regarding the effect of outside options in situations of bargaining over public information.

An interesting feature of our framework is that by opting out, a player reveals information about his type. Our veto-power assumption implies that this learning has no effect because the game essentially ends with the group choosing the status quo. Notice that even if we assume that a player who quits the bargaining cannot prevent others from voting on the signal, the information that is leaked about the quitter type still does not affect the voting strategies. This follows from the fact that each player has a weakly dominant strategy that depends solely on his own type and on the public signal.[25]

[25] By contrast, there is a literature that looks at second-stage trading that follows a first-stage auction. In these environments, the information learned about a player in the first stage – in particular, whether that player opted out – affects the strategic behavior in the second stage. See, for example, Zheng (2002), Haile (2003), Hafalir and Krishna (2008), Zhang and Wang (2013), and Dworczak (2020).

6.3 Public vs. private information

This paper focuses on the acquisition of *public* information by a group before it makes a collective action. Clearly, there are situations where group members can also decide on costly acquisition of *private* information. Extending the model by allowing players to collect private information introduces some nontrivial challenges.

First, it is no longer the case that each player's type has a weakly dominant strategy that depends solely on the public signal. Second, determining the effect of tightening the supermajority requirement is more difficult: on the one hand, the supermajority rule affects the players' incentives to acquire private information (see, e.g., Persico, 2004); on the other hand, the supermajority rule serves to aggregate the private information collected by the players. Finally, when the decision to acquire private information is made before the decision to acquire public information, there is a more complicated learning process in which each player tries to infer the other players' types. In particular, here the inference that each player makes about the other group members can affect his voting behavior because he may also update his beliefs about the state. Hopefully our work will inspire future research to explore these questions.

7 Concluding remarks

This paper is concerned with the question of how groups who want to make informed collective decisions bargain over which information to acquire. Instead of committing to a particular bargaining protocol, we took a mechanism design approach that looks for the signal that maximizes the players' expected sum of utilities, taking into account that (*i*) players must be willing to participate in the mechanism, (*ii*) players must be willing to disclose their private willingness-to-pay for information, and (*iii*) players vote on the outcome after they jointly observe the realization of the acquired signal.

An optimal mechanism exhibits two types of distortions in information acquisition. First, the fact that the group members vote on the basis of the signal realization implies that the signal that maximizes the net expected surplus is not necessarily the signal that is acquired (even when types are commonly known). This stems from the fact that it is wasteful to purchase a signal that will not persuade a supermajority to vote against the default action. Second, the fact that players need to be incentivized to disclose their types, as this determines what the optimal signal is, further distorts the type of information that is acquired: the probability of departing from the status quo *decreases* while the induced posterior belief that this is the right decision *increases* (i.e., when the players vote for $a = 1$ they do so with *higher* confidence). Thus, our analysis suggests that groups who rely on collective public information to make collective decisions are more conservative in departing from a status quo relative to the case of commonly known types (with or without a central planner).

In addition, our characterization of the optimal information structure has the potential to inform future empirical studies on decision problems involving collective information acquisition. First, our characterization suggests some testable implications. For instance, in the context of our household example, an important collective decision that households make is on whether or not to send a child to a nonstandard educational environment (e.g., a gifted-child program, a special education school, a competitive sports team, etc.).[26] Checking whether the new environment fits the child's needs and abilities is costly in terms of time, effort, and money. While both parents obviously want to make the "right" decision, they sometimes differ in their attitude toward a "wrong" one. For example, one parent may be worried that sending the child to a stressful environment might harm her, whereas the other parent may be more concerned about the possibility that the child will not fulfill her potential. Since the parents' preferences over information in this case differ, they have to agree first on what information to acquire (if any) and how to share the burden of collecting it.[27]

Suppose that households decide optimally (as is often assumed in the empirical literature). Suppose also that data can be collected on which type of information households acquire (e.g., which experts they consulted with, which preparatory courses they completed, which psychological or medical tests they took, etc.), what their subsequent decision was, what the preference intensities of each spouse regarding the decision were and to what extent these preference intensities were known to each spouse (perhaps using sophisticated questionnaires). One testable implication is that couples with polarized preference intensities will not acquire information even though according to their preference intensities it is sometimes socially optimal to do so (and this is true even if preference intensities are commonly known). Another testable implication is that spouses who are less informed about each other's preferences have a greater tendency to seek tests that lean toward the status quo. That is, they would perform tests with a higher false positive rate (i.e., higher probability of falsely identifying a match) when the default action is to apply to the nonstandard educational environment, and tests with a higher false negative rate (i.e., with a higher probability of falsely identifying a mismatch) when the default action is to not apply to the nonstandard educational environment.

Second, in the context of a specific application, our results allow one to measure how far in terms of welfare the observed information choices and/or the collective actions are from the theoretical optimum. For example, suppose data can be collected on decisions that require a majority consent among managers in an organization, and on the information they choose to

[26] Of course, sometimes after the key yes/no decision is made, the household has to decide on the specific program/school/team to apply to. Oftentimes, however, this decision is of lower importance, and involves less polarization of the spouses' preferences.

[27] For example, while one spouse spends time taking the child to the doctor/psychologist/test, the other spouse does other chores in return.

collect prior to making the decisions. One can then measure the proportion of decisions that lead to deviations from a status quo. If the proportion of such decisions differ significantly from the theoretical optimum implied by our model, this may suggest that the process of bargaining over information in the organization is suboptimal. Of course, in order to do that one has to adapt the abstract model to the specific application in mind.

References

[1] **Aghamolla, Cyrus and Tadashi Hashimoto** (2020): "Information Arrival, Delay, and Clustering in Financial Markets
with Dynamic Freeriding." *Journal of Financial Economics* 138, 27-52.

[2] **Alonso, Ricardo and Odilon Câmara** (2016): "Persuading Voters." *American Economic Review* 106(11), 3590-3605.

[3] **Arieli, Itai and Yakov Babichenko** (2019): "Private Bayesian Persuasion." *Journal of Economic Theory* 182, 185-217.

[4] **Ashraf, Nava, Erica Field, and Jean Lee** (2014): "Household bargaining and excess fertility: an experimental study in Zambia." *American Economic Review*, 104(7), 2210–37.

[5] **Ashraf, Nava, Erica Field, Alessandra Voena and Roberta Ziparo**. (2021): "Maternal Mortality Risk and Spousal Differences in the Demand for Children." Working Paper, London School of Economics.

[6] **Bardhi, Arjada and Yingni Guo** (2018): "Modes of Persuasion Toward Unanimous Consent." *Theoretical Economics* 13, 1111–1149.

[7] **Bergemann, Dirk and Stephen Morris** (2013): "Robust Predictions in Games of Incomplete Information." *Econometrica* 81(4), 1251-1308.

[8] **Bergemann, Dirk and Jusso Välimäki** (2000): "Experimentation in Markets." *Review of Economic Studies* 67, 213–234.

[9] **Bolton, Partick and Christopher Harris** (1999): "Strategic experimentation." *Econometrica* 67, 349–374.

[10] **Borgers, Tilman** (2015): An Introduction to Theory of Mechanism Design, Oxford University Press.

[11] **Candogan, Ozan and Philipp Strack** (2021): "Optimal Disclosure of Information to a Privately Informed Receiver." Working paper.

[12] **Caplin, Andrew, Mark Dean and John Leahy** (2019): "Rational Inattention, Optimal Consideration Sets and Stochastic Choice." *Review of Economic Studies* 86(3), 1061–1094.

[13] **Casella, Alessandra, Aniol Llorente-Saguer and Thomas R. Palfrey** (2012): "Competitive Equilibrium in Markets for Votes." *Journal of Political Economy* 120(4), 593-658.

[14] **Chan, Jimmy, Seher Gupta, Fei Li and Yun Wang** (2019): "Pivotal Persuasion." *Journal of Economic Theory* 180, 178-202.

[15] **Chan, Jimmy, Alessandro Lizzeri, Wing Suen and Leeat Yariv** (2018): "Deliberating Collective Decisions." *The Review of Economic Studies*, 85(2), 929–963.

[16] **Dal Bó, Ernesto** (2007): "Bribing Voters." *American Journal of Political Science* 51(4), 789–803.

[17] **Décamp, Jean-Paul and Thomas Mariotti** (2004): "Investment Timing and Learning Externality." *Journal of Economic Theory* 18, 80–102

[18] **Doepke, Matthias, and Fabian Kindermann** (2019): "Bargaining over Babies: Theory, Evidence, and Policy Implications." *American Economic Review*, 109(9), 3264–3306.

[19] **Dekel, Eddie, Matthew O. Jackson and Asher Wolinsky** (2008). "Vote Buying: General Elections." Journal of Political Economy 116(2), 351-380.

[20] **Drexl, Moritz and Andreas Kleiner** (2018): "Why Voting? A Welfare Analysis." *American Economic Journal: Microeconomics*, 10(3), 253-271.

[21] **Dworczak, Piotr** (2020): "Mechanism Design With Aftermarkets: Cutoff Mechanisms." *Econometrica* 88, 2629-2661.

[22] **Feddersen, Timothy and Wolfgang Pesendorfer** (1998): "Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting." *American Political Science Review*, 92(1), 23–35.

[23] **Gersbach, Hans** (2000): "Size and distributional uncertainty, public information and the information paradox.", *Social Choice and Welfare*, 17, 241–246.

[24] **Gershkov, Alex and Balázs Szentes** (2009): "Optimal Voting Schemes with Costly Information Acquisition." *Journal of Economic Theory*, 144(1), 36–68.

[25] **Gershkov, Alex, Jacob K. Goeree, Alexey Kushnir, Benny Moldovanu and Xianwen Shi** (2013): "On the Equivalence of Bayesian and Dominant Incentive Compatible Mechanisms." *Econometrica*, Volume 81(1), 197–220.

[26] **Glas Afina S., Jeroen G. Lijmer, Martin H. Prins, Gouke J. Bonsel and Patrick M. M. Bossuyt** (2003): "The Diagnostic Odds Ratio: A Single Indicator of Test Performance." *Journal of Clinical Epidemiolgy* 56(11), 1129-1135.

[27] **Godefroy, Raphael and Eduardo Perez-Richet** (2013): "Choosing Choices: Agenda Selection with Uncertain Issues." *Econometrica,* 81(1), 221-253

[28] **Hafalir, Isa and Vijay Krishna** (2008): "Asymmetric Auctions with Resale." *American Economic Review*, 98 (1), 87-112.

[29] **Haile, Philip A.** (2003): "Auctions with Private Uncertainty and Resale Opportunities." *Journal of Economic Theory*, 108, 72–110.

[30] **Hellwig, Martin F.** (2003): "Public-Good Provision with Many Participants." *The Review of Economic Studies*, 70(3), 589–614.

[31] **Kamien, Morton I. and Nancy L. Schwartz** (2012): Dynamic Optimization, Second Edition, Dover Publications.

[32] **Kelly, Frank and Yudovina, Elena** (2014): Stochastic Networks, Cambridge University Press, USA.

[33] **Kolotilin, Anton, Tymofiy Mylovanov, Andriy Zapechelnyuk and Ming Li** (2017): "Persuasion of a Privately Informed Receiver." *Econometrica* 85(6), 1949–1964.

[34] **Krishna, Vijay** (2010): Auction Theory*,* Academic Press.

[35] **Lundberg, Shelly and Robert A. Pollak** (1996): "Bargaining and Distribution in Marriage." *Journal of Economic Perspectives,* 10(4), 139-158.

[36] **Maćkowiak, Bartosz, Filip Matějka and Mirko Wiederholt** (2018): "Rational Inattention: A Disciplined Behavioral Model." Working paper.

[37] **Mailath, George J. and Andrew Postlewaite** (1990): "Asymmetric Information Bargaining Problems with Many Agents." *The Review of Economic Studies*, Volume 57(3), 351–367.

[38] **Matějka, Filip and Alisdair McKay** (2015): "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model." *American Economic Review*, 105(1), 272–298.

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

[40] **Mookherjee, Dilip and Stefan Reichelstein** (1992): "Dominant Strategy Implementation of Bayesian Incentive Compatible Allocation Rules." *Journal of Economic Theory* 56, 378-399.

[41] **Myerson, Roger B. and Mark A. Satterthwaite** (1983). "Efficient mechanisms for bilateral trading." *Journal of Economic Theory*, 29(2), 265-281.

[42] **Persico, Nicola** (2004): "Committee Design with Endogenous Information." *The Review of Economic Studies*, 71(1), 165–191.

[43] **Schnakenberg, Keith E.** (2015): "Expert Advice to a Voting Body." *Journal of Economic Theory* 160, 102-113.

[44] **Sims, Christopher A.** (2003): "Implications of Rational Inattention." *Journal of Monetary Economics* 50(3), 665-690.

[45] **Taneva, Ina** (2019): "Information Design." *American Economic Journal: Microeconomics* 11(4), 2019, 151-185.

[46] **Wang, Yun** (2013): "Bayesian Persuasion with Multiple Receivers." Working Paper.

[47] **Zhang, Jun and Ruqu Wang** (2013): "Optimal Mechanism Design with Resale via Bargaining." *Journal of Economic Theory*, 148(5), 2096-2123.

[48] **Zheng, Zhoucheng C.** (2002): "Optimal Auction with Resale." *Econometrica*, 70, 2197-2224.

8 Appendix: Proofs

Proof of Lemma 1

Let $\theta \in [0, 1-p]^n$ be the players' types. Consider a signal that induces a probability distribution q over a set $R \in [0,1]^J$ of posterior beliefs (on state $\omega = 1$) such that the expected posterior equals p, i.e., $\sum_{r \in R} q(r) \cdot r = p$. Let $\bar{R}$ (respectively, $\underline{R}$) be the set of posterior beliefs above (respectively, below) $1 - \theta^{(n-m+1)}$. Suppose that $\bar{R}$ contains (at least) two distinct elements r' and r'', where $r' < r''$. Both r' and r'' lead to the same collective action $a = 1$ in the voting game.

Consider now a modified signal that induces a distribution $\hat{q}$ over a set of posterior beliefs $\hat{R}$. The set $\hat{R}$ is identical to R, with one difference: the posteriors r' and r'' are replaced by the posterior $\hat{r} \equiv \frac{q(r')}{q(r')+q(r'')} r' + \frac{q(r'')}{q(r')+q(r'')} r''$. The distribution $\hat{q}$ is defined such that $\hat{q}(r) = q(r)$ for all $r \in R \backslash \{r', r''\}$, while $\hat{q}(\hat{r}) = q(r') + q(r'')$. Note that since $\hat{r} \in (r', r'')$, $\hat{r}$ is above $1 - \theta^{(n-m+1)}$ and so it induces the collective action $a = 1$, which is the same as the collective action induced by r' and r''. Thus, the modified signal $\hat{q}$ (over $\hat{R}$) induces the same distribution over outcomes as the original signal q (over R). By construction, the modified signal also satisfies $\sum_{r \in \hat{R}} \hat{q}(r) \cdot r = p$. Since the modified signal is strictly less informative in the Blackwell sense, it is cheaper than the original one. The proof for the case in which there are more than two elements in $\underline{R}$ is analogous. ■

Proof of Proposition 1

Fix n and $\overline{\theta}$. The derivative of r_H in the planner's solution (as defined in Equation (5)) with respect to κ is given by:

$$\frac{d}{d\kappa}\left(\frac{e^{\frac{n}{\kappa}} - e^{\frac{n}{\kappa}\overline{\theta}}}{e^{\frac{n}{\kappa}} - 1}\right) = -\left(\left(\overline{\theta} + \left(1-\overline{\theta}\right) e^{\frac{n}{\kappa}}\right) e^{-\frac{n}{\kappa}\left(1-\overline{\theta}\right)} - 1\right) \cdot \frac{n e^{\frac{n}{\kappa}}}{\kappa^2 \left(e^{\frac{n}{\kappa}} - 1\right)^2}.$$

This expression is negative. This is because $\left(\overline{\theta} + \left(1-\overline{\theta}\right) e^{\frac{n}{\kappa}}\right) e^{-\frac{n}{\kappa}\left(1-\overline{\theta}\right)} - 1$ is positive for all for all κ and $\overline{\theta} < 1$ (to see this, let $g(z) = \left(\overline{\theta} + \left(1-\overline{\theta}\right) z\right) z^{\left(\overline{\theta}-1\right)} - 1$, and notice that $g(1) = 0$, and that $dg/dz > 0$ for all $z > 1$ and $\overline{\theta} < 1$). Thus, r_H is decreasing in κ.

Similarly, the derivative of r_L in the planner's solution (as defined in Equation (5)) with respect to κ is

$$\frac{d}{d\kappa}\left(\frac{e^{\frac{n}{\kappa}\left(1-\overline{\theta}\right)} - 1}{e^{\frac{n}{\kappa}} - 1}\right) = \left(\left(\overline{\theta} + \left(1-\overline{\theta}\right) e^{-\frac{n}{\kappa}}\right) e^{\frac{n}{\kappa}\left(1-\overline{\theta}\right)} - 1\right) \cdot \frac{n}{\kappa^2} \frac{e^{-\frac{n}{\kappa}}}{\left(e^{-\frac{n}{\kappa}} - 1\right)^2}.$$

This expression is positive, because $\left(\overline{\theta}+\left(1-\overline{\theta}\right)e^{-\frac{n}{\kappa}}\right)e^{\frac{n}{\kappa}\left(1-\overline{\theta}\right)}-1$ is positive for all for all κ and $\overline{\theta}<1$ (to see this, define g as before, and notice that $dg/dz<0$ for all $z<1$ and $\overline{\theta}<1$). Thus, r_L is increasing in κ.

Taken together, these observations imply that for all n and $\overline{\theta}$, as κ decreases, the value of r_H in the planner's solution increases whereas r_L decreases, making the signal purchased by the planner more Blackwell informative. Finally, it is easy to verify that as $\kappa\to 0$ the values of r_H and r_L, as given by (5), converge to 1 and 0, respectively, while q converges to p. Thus, as the cost scaler goes to zero, the socially optimal signal that the planner acquires for any type profile converges to the fully informative one.■

Proof of Lemma 2

From $U\left(\theta_i\right)=\int_0^{\theta}Q\left(x\right)dx-M(0)-T_i(0)$ we obtain

$$T_i(\theta_i)=Q(\theta_i)\cdot\theta_i-M(\theta_i)-\int_0^{\theta}Q\left(x\right)dx+M(0)+T_i(0). \tag{22}$$

Player i's ex-ante expected net utility is given by $\int_0^{1-p}U(\theta_i)dF(\theta_i)$. Applying integration by parts we obtain

$$\int_0^{1-p}U(\theta_i)dF(\theta_i)=\int_0^{1-p}Q(\theta_i)[\frac{1-F(\theta_i)}{f(\theta_i)}]dF\left(\theta_i\right)-T_i(0)-M(0)$$

Plugging in $U\left(\theta_i\right)=Q(\theta_i)\cdot\theta_i-M(\theta_i)-T_i(\theta_i)$ and rearranging yields:

$$\int_0^{1-p}T_i(\theta_i)dF(\theta_i)=T_i(0)+M(0)+\int_0^{1-p}\left[v(\theta_i)\cdot Q(\theta_i)-M(\theta_i)\right]dF(\theta_i) \tag{23}$$

where $v(\theta_i)$ is the virtual valuation of type θ_i.

Substituting Equation (23) into Equation (11) yields that the designer's problem is to maximize

$$\sum_{i=1}^{n}\int_0^{1-p}\left[\frac{1-F(\theta_i)}{f(\theta_i)}Q(\theta_i)\right]dF(\theta_i)-\sum_{i=1}^{n}T_i(0)-\sum_{i=1}^{n}M(0) \tag{24}$$

subject to the following ex-ante budget balancedness constraint (which is obtained by plugging Equation (23) into Equation (BB)):

$$\int_{\theta}c(q(\theta),r_H(\theta))dF^n(\theta)=\sum_{i=1}^{n}T_i(0)+\sum_{i=1}^{n}M(0)+\sum_{i=1}^{n}\int_0^{1-p}\left[v(\theta_i)Q(\theta_i)-M(\theta_i)\right]dF(\theta_i) \tag{25}$$

where individual rationality requires $-M(0)-T_i(0)\geq 0$ for every player i, and therefore

$0 \leq -\sum_{i=1}^{n}[T_i(0)+M(0)]$. Since the constants $T_1(0), \ldots, T_n(0)$ enter the objective function and the constraint only through the aggregate $\sum_{i=1}^{n} T_i(0)$, we can assume that they are all equal. We therefore denote $T(0) = T_1(0) = \ldots = T_n(0)$. We then use the ex-ante budget-balancedness constraint to substitute for $-\sum_{i=1}^{n} T_i(0) - \sum_{i=1}^{n} M(0)$ in Equation (24) and obtain the objective function and the conclusion that inequality (13) is a necessary condition for individual rationality and ex-ante budget-balancedness.

To show that inequality (13) is a sufficient condition for individual rationality and ex-ante budget-balancedness, first denote by q^* and r_H^* the solution to the optimization problem stated in the lemma. Let Ψ^* denote the aggregate virtual surplus (the left-hand side of Equation (13)) evaluated at q^* and r_H^*. Second, compute $M^*(0)$ using q^* and r_H^*. Third, set $T^*(0) = -M^*(0) - \frac{1}{n}\Psi^*$. This guarantees ex-ante budget balancedness according to Equation (25). Since the aggregate virtual surplus Ψ^* is nonnegative by Equation (13), individual rationality is satisfied (i.e. $-T^*(0) - M^*(0) \geq 0$). Finally, to complete the description of the mechanism it remains to define the transfer functions $(t_i^*(\theta))_{i=1}^{n}$ such that for each player i, $\mathbb{E}_{\theta_{-i}}(t_i^*(\theta_i, \theta_{-i})) = T_i^*(\theta_i)$. One way to do this is simply to let $t_i^*(\theta_i, \theta_{-i}) = T_i^*(\theta_i)$. ■

Proof of Lemma 3

Suppose that $\langle q, r_H, t_1, ..., t_n \rangle$ is an auxiliary mechanism that satisfies incentive compatibility, individual rationality, and ex-ante budget balancedness, but does not satisfy SP. We show a modification that increases the expected payoff to the players without affecting the constraints. Therefore, the given mechanism is not optimal.

Since the mechanism does not satisfy SP, there exists a non-zero measure of type realizations (θ_i, θ_{-i}) for which $q(\theta) > 0$ and $r_H(\theta) < 1 - \theta^{(n-m+1)}$. Suppose we modify q into q' as follows:

$$q'(\theta) = \begin{cases} q(\theta) & if \quad r_H(\theta) \geq 1 - \theta^{(n-m+1)} \\ 0 & if \quad r_H(\theta) < 1 - \theta^{(n-m+1)} \end{cases} .$$

That is, whenever the original mechanism purchases a noninstrumental signal, the modified mechanism does not purchase a signal. Notice that

$$
\begin{aligned}
Q'(\theta_i) &= \int_{\theta_{-i}|r_H(\theta_i,\theta_{-i})>1-(\theta_i,\theta_{-i})^{(n-m+1)}} q'(\theta_i,\theta_{-i})dF\left(\theta_{-i}\right) \\
&= \int_{\theta_{-i}|r_H(\theta_i,\theta_{-i})>1-(\theta_i,\theta_{-i})^{(n-m+1)}} q(\theta_i,\theta_{-i})dF\left(\theta_{-i}\right) = Q(\theta_i) \\
M'\left(\theta_i\right) &= \int_{\theta_{-i}|r_H(\theta_i,\theta_{-i})>1-(\theta_i,\theta_{-i})^{(n-m+1)}} \left(1 - r_H(\theta_i,\theta_{-i})\right)\cdot q'(\theta_i,\theta_{-i})dF\left(\theta_{-i}\right) \\
&= \int_{\theta_{-i}|r_H(\theta_i,\theta_{-i})>1-(\theta_i,\theta_{-i})^{(n-m+1)}} \left(1 - r_H(\theta_i,\theta_{-i})\right)\cdot q(\theta_i,\theta_{-i})dF\left(\theta_{-i}\right) \\
&= M\left(\theta_i\right).
\end{aligned}
$$

Denote the expected decrease in the cost of purchasing signals by

$$\Delta = \int_{\theta|r_H(\theta)<1-\theta^{(n-m+1)}} c(q(\theta), r_H(\theta))dF\left(\theta\right) > 0.$$

For every $i \in \{1, \ldots, n\}$ define

$$t_i'(\hat{\theta}) = t_i(\hat{\theta}) - \frac{\Delta}{n}.$$

The new mechanism satisfies incentive compatibility and individual rationality because $Q' = Q$ and $M' = M$, and the transfers decreased by a constant for all types (so that $-M\left(0\right) - T\left(0\right) \geq 0$). By construction, the mechanism is budget-balanced, and since the expected payment of type 0 decreases, then by Equation (24) the expected surplus increases. ■

Proof of Proposition 2

The proof consists of three parts. First, we characterize the three functions, $r_H^*\left(\theta,\lambda\right), r_L^*\left(\theta,\lambda\right)$, and $q^*\left(\theta,\lambda\right)$, that satisfy SP (see Definition 1) and maximize $\hat{\mathcal{L}}$ (see Equation (18)), for any type profile θ and any multiplier λ (for ease of exposition we omit the dependency of $\hat{\mathcal{L}}$ on the type profile and the multiplier in the notation). Second, we show that for any $\lambda \geq 0$, the function $q^*\left(\theta,\lambda\right)$ is increasing in each player's type whereas $r_H^*\left(\theta,\lambda\right)$ is decreasing in each player's type. Hence, the function $Q^*\left(\theta_i,\lambda\right)$ that is induced by $q^*\left(\theta,\lambda\right)$ (according to Equation (14)) is monotone. Third, we show that there exists $\lambda^* \geq 0$ for which the mechanism defined by $r_H^*\left(\theta,\lambda^*\right), r_L^*\left(\theta,\lambda^*\right)$ and $q^*\left(\theta,\lambda^*\right)$ generates zero aggregate virtual surplus. By the Lagrange Sufficiency Theorem (see, e.g., Theorem C.1 in Kelly and Yudovina, 2014), the functions $r_H^*\left(\theta,\lambda^*\right), r_L^*\left(\theta,\lambda^*\right)$ and $q^*\left(\theta,\lambda^*\right)$ define the mechanism that attains the maximal aggregate surplus (Equation (12)) subject to the conditions that: (i) $Q\left(\theta_i\right)$ is monotone and (ii) the aggregate virtual surplus (Equation (13)) is nonnegative.

PART I. Fix a type profile θ and a multiplier $\lambda > 0$ and let $w = w(\theta, \lambda)$. Recall that, as defined in the text, $(\tilde{q}, \tilde{r})$ is the signal that solves (FOCq) and (FOCr), if such a signal exists. By property (P3), if $(\tilde{q}, \tilde{r})$ is interior then it is the unique signal that satisfies the necessary conditions for being a local (or global) maximizer of $\hat{\mathcal{L}}$.

We consider three cases:

Case (i): Suppose that $(\tilde{q}, \tilde{r}_H)$ is not interior, or that (FOCq) and (FOCr) do not have a solution. Then, it is optimal to purchase no signal. This is because the signal that maximizes $\hat{\mathcal{L}}$ must be a "corner solution": either $r_H = 1$, or $q = p/r_H$ (which is equivalent to $r_L = 0$), or $q = 0$ (which is equivalent to acquiring no information). Consider the corner solution $(q, 1)$ for some $q > 0$. This solution does not maximize $\hat{\mathcal{L}}$ because by property (P2) of the cost function, $\frac{1}{n}c_2(q, 1) > 1$ and therefore $\mathcal{L}_2(q, 1\ ;\ w) < 0$. Thus, decreasing r_H increases $\hat{\mathcal{L}}$. Similarly, the corner solution $\left(\frac{p}{r_H}, r_H\right)$ for some $r_H > p$ is also not a maximizer of $\hat{\mathcal{L}}$. This is because by (P2), $\frac{1}{n}c_1(p/r_H, r_H) > 1$, and therefore $\hat{\mathcal{L}}_1(p/r_H, r_H\ ;\ w) < 0$ for any w. Thus, the optimal solution is the corner solution $q = 0$ in which no signal is acquired.

Case (ii): Suppose that $(\tilde{q}, \tilde{r}_H)$ is interior and SP. Then, $q_H^*(\theta, \lambda) = \tilde{q}$ and $r_H^*(\theta, \lambda) = \tilde{r}_H$. To prove this, it suffices to show that $(\tilde{q}, \tilde{r}_H)$ is better than any corner solution. This is because, by property (P3) of the cost function, there are no other local or global interior maximizers for $\hat{\mathcal{L}}$. The signals in which $r_H = 1$, or $q = p/r_H$ are not optimal because of the argument presented in Case (i) above. To see why acquiring no information $(q = 0)$ is not optimal, consider the function $g(q) \equiv q \cdot \frac{1}{n}c_1(q, \tilde{r}_H) - \frac{1}{n}c(q, \tilde{r}_H)$ that is obtained by plugging (FOCq) into $\hat{\mathcal{L}}$. The fact that $c_{11} > 0$ (by property P1) implies that $g(0) = 0$ and $g'(q) > 0$ for every $q > 0$. Since $(\tilde{q}, \tilde{r}_H)$ is interior, $\tilde{q} > 0$ and consequently $\hat{\mathcal{L}}(\tilde{q}, \tilde{r}_H\ ;\ w) = g(\tilde{q}) > 0$. Namely, the signal $(\tilde{q}, \tilde{r}_H)$ generates a positive value of $\hat{\mathcal{L}}$, which is greater than 0 that is obtained when no information is acquired.

Case (iii): Suppose that $(\tilde{q}, \tilde{r}_H)$ is interior but not SP. Let q_{MSP} be the value that satisfies (FOCq) when r_H is set to its MSP value, i.e., $r_H = 1 - \theta^{(n-m+1)}$. Thus, q_{MSP} satisfies $\mathcal{L}_1\left(q_{MSP}, 1 - \theta^{(n-m+1)}\ ;\ w\right) = 0$. If $q_{MSP} \in (0, p/r_H)$ the optimal signal is $r_H^*(\theta, \lambda) = 1 - \theta^{(n-m+1)}$ and $q^*(\theta, \lambda) = q_{MSP}$. Otherwise, it is optimal to acquire no information, i.e., $q^*(\theta, \lambda) = 0$. To see this, recall again that $(\tilde{q}, \tilde{r}_H)$ is the unique global (and local) maximizer of $\hat{\mathcal{L}}$. Since $(\tilde{q}, \tilde{r}_H)$ is not SP, the maximizer $\hat{\mathcal{L}}$ must be a corner solution: either acquire no information $(q = 0)$, or acquire the signals in which $r_H = 1$ or $q = p/r_H$, or acquire the signal $\left(q_{MSP}, 1 - \theta^{(n-m+1)}\right)$, which is possible only if $q_{MSP} \in (0, p/r_H)$. The argument in Case (i) above implies that a signal with $r_H = 1$ or $q = p/r_H$ does not maximize $\hat{\mathcal{L}}$.[28] An argument similar to the one presented in Case (ii) above, with the only difference that r_H is equal to

[28] That is, unless $\theta^{(n-m+1)} = 0$, in which case $r_H = 1$ is in fact the MSP value of r_H.

$1-\theta^{(n-m+1)}$ instead of $\tilde{r}_H$, implies that if $q_{MSP} > 0$ then $\hat{\mathcal{L}}\left(q_{MSP}, 1-\theta^{(n-m+1)}\ ;\ w\right)$ is positive, which is higher than 0 that is obtained when no information is acquired.

We have thus characterized the functions $r_H^*(\theta,\lambda), r_L^*(\theta,\lambda)$, and $q^*(\theta,\lambda)$ that satisfy SP and maximize $\hat{\mathcal{L}}$ for the given type profile θ and multiplier λ. Notice that property (P3) implies that the signal $(\tilde{q}, \tilde{r}_H)$ is continuous in $\theta_1, \ldots, \theta_n$ and λ,[29] and therefore the functions $r_H^*(\theta,\lambda), r_L^*(\theta,\lambda)$ and $q^*(\theta,\lambda)$ are also continuous in $\theta_1, \ldots, \theta_n$ and λ.

PART II. Fix a player i, a type profile θ_{-i} of all the players other than i, and a multiplier λ. For simplicity, we omit θ_{-i} and λ from the notation and write $q^*(\theta_i)$, $r_H^*(\theta_i)$, $w(\theta_i), c_i^*(\theta_i)$, and $c_{ij}^*(\theta_i)$ instead of $q^*((\theta_i,\theta_{-i}),\lambda)$, $r_H^*((\theta_i,\theta_{-i}),\lambda)$, $w((\theta_i,\theta_{-i}),\lambda)$, $c_i(q^*(\theta_i), r_H^*(\theta_i))$, and $c_{ij}(q^*(\theta_i), r_H^*(\theta_i))$, respectively.

To show that $q^*(\theta_i)$ is increasing in θ_i while $r_H^*(\theta_i)$ is decreasing in θ_i we prove three lemmas that, taken together with the results of Part I, imply the desired monotonicity result. Lemma 4 asserts that if a signal is acquired when player i's type is θ_i, then a signal is also acquired when player i's type is higher than θ_i. This implies that there exists some cutoff type θ_i' such that if $\theta_i < \theta_i'$ then no signal is acquired, i.e., $q^*(\theta_i) = 0$ and $r_H^*(\theta_i) = 1$,[30] and if $\theta_i > \theta_i'$ then *some* information is acquired (i.e., $q^*(\theta_i) > 0$).

Lemma 5 shows that if an interval $(a,b) \subset [\theta_i', 1-p]$ is such that the optimal signal is interior when $\theta_i \in (a,b)$, then $q^*(\theta_i)$ is increasing in θ_i, and $r_H^*(\theta_i)$ is decreasing in θ_i, as θ_i varies within (a,b). Lemma 6 shows that this monotonicity of $q^*(\theta_i)$ and $r_H^*(\theta_i)$ holds also when θ_i varies within an interval $(a,b) \subset [\theta_i', 1-p]$, if the optimal signal for every $\theta_i \in (a,b)$ is such that r_H attains the MSP value $1-(\theta_i,\theta_{-i})^{(n-m+1)}$ and q is determined according to Equation (FOCq).

Lemma 4 *Suppose that $q^*(\theta_i) > 0$ for some θ_i. Then $q^*(\theta_i) > 0$ for all $\theta_i' > \theta_i$.*

Proof. The fact that $q^*(\theta_i) > 0$ implies that acquiring a signal is beneficial for the type profile (θ_i, θ_{-i}), which means that $\hat{\mathcal{L}}(q^*(\theta_i), r_H^*(\theta_i)\ ;\ w(\theta_i)) \geq 0$. The fact that $(q^*(\theta_i), r_H^*(\theta_i))$ is optimal for the type profile (θ_i,θ_{-i}) implies that $r_H^*(\theta_i)$ satisfies SP, i.e. $r_H^*(\theta_i) \geq 1 - (\theta_i,\theta_{-i})^{(n-m+1)}$.

Consider a type θ_i' of player i such that $\theta_i' > \theta_i$. Since the signal $(q_H^*(\theta_i), r_H^*(\theta_i))$ is SP for the type profile (θ_i,θ_{-i}), it follows that it is also SP for the type profile (θ_i',θ_{-i}), because $(\theta_i',\theta_{-i})^{(n-m+1)} \geq (\theta_i,\theta_{-i})^{(n-m+1)}$. Thus,

$$\hat{\mathcal{L}}(q^*(\theta_i'), r_H^*(\theta_i')\ ;\ w(\theta_i')) \geq \hat{\mathcal{L}}(q^*(\theta_i), r_H^*(\theta_i)\ ;\ w(\theta_i')) > \hat{\mathcal{L}}(q^*(\theta_i), r_H^*(\theta_i)\ ;\ w(\theta_i)) \geq 0.$$

[29] By Berge's theorem, the set-valued map from w to maximizers of $\hat{\mathcal{L}}(q, r_H; w)$ is upper-hemicontinuous. However, since by (P3), for every w the function $\hat{\mathcal{L}}(\cdot,\cdot;w)$ has a unique maximizer, i.e., the set-valued map is a singleton for each w, it follows that $\tilde{r}_H, \tilde{q}$ are continuous in w, which is itself continuous in $\theta_1,\ldots,\theta_n$ and λ.

[30] When $q = 0$ no information is acquired and the value of r_H does not matter. Hence, there's no loss in assuming that in this case $r_H = 1$.

The first inequality is because the signal $q^*(\theta_i'), r_H^*(\theta_i')$ generates the highest value of $\hat{\mathcal{L}}$ among all the signals that satisfy SP. The second inequality follows from the facts that, all else equal, $\hat{\mathcal{L}}$ is increasing in w, and $w(\theta_i') > w(\theta_i)$ because the virtual value is increasing in each player's type. Finally, $\hat{\mathcal{L}}(q^*(\theta_i'), r_H^*(\theta_i')\ ;\ w(\theta_i')) > 0$ implies that $q^*(\theta_i') > 0$. ■

Lemma 5 *Suppose that $r_H^*(\theta_i) = \tilde{r}_H(\theta_i)$ and $q^*(\theta_i) = \tilde{q}(\theta_i)$ for all θ_i in some interval (a, b). If $\theta_i, \theta_i' \in (a, b)$ such that $\theta_i' > \theta_i$, then $q^*(\theta_i') > q^*(\theta_i)$ and $r_H^*(\theta_i') < r_H^*(\theta_i)$.*

Proof. By definition, for any $\theta_i \in (a, b)$ the (interior) signal $(q^*(\theta_i), r_H^*(\theta_i))$ satisfies

$$r_H^*(\theta_i) - 1 + w(\theta_i) = \frac{1}{n} c_1^*(\theta_i), \quad \text{and} \tag{26}$$

$$q^*(\theta_i) = \frac{1}{n} c_2^*(\theta_i). \tag{27}$$

In addition, since $q^*(\theta_i), r_H^*(\theta_i)$ maximizes $\hat{\mathcal{L}}(\cdot, \cdot\ ;\ w(\theta_i))$, it follows that the determinant of the Hessian matrix that is associated with $\hat{\mathcal{L}}$ is nonnegative when evaluated at $q^*(\theta_i), r_H^*(\theta_i)$. That is,

$$\left(-\frac{1}{n} c_{11}^*(\theta_i)\right) \cdot \left(-\frac{1}{n} c_{22}^*(\theta_i)\right) - \left(1 - \frac{1}{n} c_{12}^*(\theta_i)\right)^2 \geq 0. \tag{28}$$

The fact that (26) and (27) hold simultaneously for all $\theta_i \in (a, b)$ implies that the derivatives with respect to θ_i on both sides of each equation must also be the same. Thus,

$$\frac{dr_H^*(\theta_i)}{d\theta_i} + \frac{dw(\theta_i)}{d\theta_i} = \frac{dq^*(\theta_i)}{d\theta_i} \cdot \frac{1}{n} c_{11}^*(\theta_i) + \frac{dr_H^*(\theta_i)}{d\theta_i} \cdot \frac{1}{n} c_{12}^*(\theta_i), \text{ and}$$

$$\frac{dq^*(\theta_i)}{d\theta_i} = \frac{dq^*(\theta_i)}{d\theta_i} \cdot \frac{1}{n} c_{21}^*(\theta_i) + \frac{dr_H^*(\theta_i)}{d\theta_i} \cdot \frac{1}{n} c_{22}^*(\theta_i).$$

Since c is twice continuously differentiable, $c_{12}^*(\theta_i) = c_{21}^*(\theta_i)$. Solving for $dr_H^*(\theta_i)/d\theta_i$ and $dq^*(\theta_i)/d\theta_i$ we obtain

$$\frac{dq^*(\theta_i)}{d\theta_i} = \frac{\frac{1}{n} c_{22}^*(\theta_i) \cdot \frac{dw(\theta_i)}{d\theta_i}}{\frac{1}{n} c_{11}^*(\theta_i) \cdot \frac{1}{n} c_{22}^*(\theta_i) - \left(1 - \frac{1}{n} c_{12}^*(\theta_i)\right)^2} \tag{29}$$

$$\frac{dr_H^*(\theta_i)}{d\theta_i} = \frac{\left(1 - \frac{1}{n} c_{12}^*(\theta_i)\right) \cdot \frac{dw(\theta_i)}{d\theta_i}}{\frac{1}{n} c_{11}^*(\theta_i) \cdot \frac{1}{n} c_{22}^*(\theta_i) - \left(1 - \frac{1}{n} c_{12}^*(\theta_i)\right)^2} \tag{30}$$

The numerator of (29) is positive because $dw(\theta_i)/d\theta_i > 0$ and $c_{22}^*(\theta_i) > 0$ (by property P1). The denominator of (29) is positive by Equation (28). Thus, $dq^*(\theta_i)/d\theta_i > 0$ and therefore $q^*(\theta_i') > q^*(\theta_i)$.

To show that $r_H^*(\theta_i') < r_H^*(\theta_i)$ it suffices to show that $\frac{1}{n}c_{12}^*(\theta_i) > 1$. This is because, by Equations (28) and (30) and since $dw(\theta_i)/d\theta_i > 0$, the sign of $\frac{dr_H^*(\theta_i)}{d\theta_i}$ is the same as the sign of $1 - \frac{1}{n}c_{12}^*(\theta_i)$. To see why $\frac{1}{n}c_{12}^*(\theta_i) > 1$, hold $r_H^*(\theta_i)$ fixed and define the functions $g(q) \equiv \frac{1}{n}c_2(q, r_H^*(\theta_i))$ and $h(q) \equiv q$. Since g is increasing and convex in q (because of P1), while h is increasing and linear in q, it follows that g and h intersect for at most two values of q, where the slope of g is greater (less) than the slope of h at the higher (lower) intersection point. Since $g(0) = h(0)$ and $g(q^*(\theta_i)) = h(q^*(\theta_i))$, as implied by Equation (27), and since $q^*(\theta_i) > 0$, we deduce that $g'(q^*(\theta_i)) > h'(q^*(\theta_i))$, i.e. $\frac{1}{n}c_{12}^*(\theta_i) > 1$. ■

Lemma 6 *Suppose that* $r_H^*(\theta_i, \theta_{-i}) = 1 - (\theta_i, \theta_{-i})^{(n-m+1)}$ *and* $q^*(\theta_i, \theta_{-i}) > 0$ *is determined according to (FOCq) for all* θ_i *in some interval* (a, b)*. If* $\theta_i, \theta_i' \in (a, b)$ *and* $\theta_i' > \theta_i$*, then* $q^*(\theta_i') \geq q^*(\theta_i)$ *and* $r_H^*(\theta_i') \leq r_H^*(\theta_i)$*.*

Proof. The fact that $r_H^*(\theta_i') \leq r_H^*(\theta_i)$ follows immediately from the observation that $\theta_i' > \theta_i$ implies that $(\theta_i', \theta_{-i})^{(n-m+1)} \geq (\theta_i, \theta_{-i})^{(n-m+1)}$. Therefore $dr_H^*(\theta_i)/d\theta_i < 0$. It remains to show that $q^*(\theta_i') \geq q^*(\theta_i)$.

Given $r_H^*(\theta_i)$, the value of $q^*(\theta_i)$ is determined according to (FOCq):

$$r_H^*(\theta_i) - 1 + w(\theta_i) = \frac{1}{n}c_1^*(\theta_i).$$

Since the equality holds for any $\theta \in (a, b)$, the derivatives with respect to θ_i on both sides of the equation must also be the equal. Therefore:

$$\frac{dr_H^*(\theta_i)}{d\theta_i} + \frac{dw(\theta_i)}{d\theta_i} = \frac{dq^*(\theta_i)}{d\theta_i} \cdot \frac{1}{n}c_{11}^*(\underline{\theta}_i) + \frac{dr_H^*(\theta_i)}{d\theta_i} \cdot \frac{1}{n}c_{12}^*(\theta_i)$$

Solving for $dq^*(\theta_i)/d\theta_i$, we obtain

$$\frac{dq^*(\theta_i)}{d\theta_i} = \frac{\left(\frac{1}{n}c_{12}^*(\theta_i) - 1\right) \cdot \left(-\frac{dr_H^*(\theta_i)}{d\theta_i}\right) + \frac{dw(\theta_i)}{d\theta_i}}{\frac{1}{n}c_{11}^*(\theta_i)}. \tag{31}$$

The fact that $r_H^*(\underline{\theta}_i) = 1 - (\theta_i, \theta_{-i})^{(n-m+1)}$ means that the SP constraint is binding. Hence, $\hat{\mathcal{L}}_2(q^*(\underline{\theta}_i), r_H^*(\underline{\theta}_i)\ ;\ w(\underline{\theta}_i)) \geq 0$, or equivalently $q^*(\theta_i) \geq \frac{1}{n}c_2^*(\theta_i)$. Hence, an argument that is similar to the one presented in the last paragraph of the proof of Lemma 5 leads to the conclusion that

$$\frac{1}{n}c_{21}^*(\theta_i) > 1.$$

Since, in addition, $dw(\theta_i)/d\theta_i > 0$ (because virtual values are increasing in types), $c_{11}^*(\theta_i) > 0$ (by property P1) and $dr_H^*(\theta_i)/d\theta_i < 0$ it follows that Equation (31) implies that $dq^*(\theta_i)/d\theta_i > 0$ and therefore $q^*(\theta_i') > q^*(\theta_i)$. ■

Part III. For any two signal functions $q : \Theta^n \to [0,1]$ and $r_H : \Theta^n \to [0,1]$, denote the aggregate surplus they induce (Equation (12)) by $W(q, r_H)$, and the aggregate virtual surplus they induce (Equation (13)) by $V(q, r_H)$. Notice that q and r_H in this part of the proof denote functions rather than constants. Using this notation, the Lagrangian associated with our maximization problem (defined in Lemma 2) is

$$\mathcal{L}(q, r, \lambda) = \frac{1}{1+\lambda} W(q, r) + \frac{\lambda}{1+\lambda} V(q, r) \tag{32}$$

$$= \int_\theta \left[q(\theta) \cdot [r_H(\theta) + w(\theta, \lambda) - 1] - \frac{1}{n} c(q(\theta), r_H(\theta)) \right] dF^n(\theta) \tag{33}$$

where (33) is identical to (16) and given here for ease of reference. For any λ, denote by $q^{*,\lambda}$ and $r_H^{*,\lambda}$ the signal functions that were characterized in Part I of the proof and correspond to this value of λ (that is, $q^{*,\lambda}(\theta) = q^*(\theta, \lambda)$ and $r_H^{*,\lambda}(\theta) = r_H^*(\theta, \lambda)$, for any type profile θ).

If $V(q^{*,0}, r_H^{*,0}) \geq 0$ then the proof is complete. In this case, the optimal signal functions are $q^{*,0}$ $r_H^{*,0}$. As we show in Parts I and II, these functions maximize the aggregate surplus W *pointwise* (that is, for any type profile θ) and induce a monotone function Q. Since $V(q^{*,0}, r_H^{*,0}) \geq 0$ they also generate a nonnegative aggregate virtual surplus.

Suppose alternatively that $V(q^{*,0}, r_H^{*,0}) < 0$. Then, there exists $\lambda^* > 0$ for which $V(q^{*,\lambda^*}, r_H^{*,\lambda^*}) = 0$. To see why, notice first that if $V(q^{*,0}, r_H^{*,0}) < 0$, then it must be the case that a signal is purchased for a set of type profiles with a non-zero measure. Notice also that for any signal (q, r_H), the value of $\hat{\mathcal{L}}$ (defined in Equation (18)) is increasing in each player's type (via w). Taken together, these facts imply that when $\lambda = 0$, the maximum of $\hat{\mathcal{L}}$ is strictly positive for all type profiles in a neighborhood of $\theta^{\max} = (1-p, \ldots, 1-p)$. Suppose now that $\lambda \to \infty$. Because for any type profile θ that is close to $\theta^{\max}$, the value of $w(\theta, \infty)$ is close to the value of $w(\theta, 0)$ (which is close to $1-p$), and because c is continuous, then when $\lambda \to \infty$ the maximum of $\hat{\mathcal{L}}$ is strictly positive (and bounded away from zero) for all type profiles in the neighborhood of $\theta^{\max}$. Thus, when $\lambda \to \infty$, Equation (33) implies that $\mathcal{L}(q^{*,\lambda}, r_H^{*,\lambda}, \lambda) > 0$ and Equation (32) implies that $V\left(q^{*,\lambda}, r_H^{*,\lambda}\right) > 0$. Finally, notice that $V(q^{*,\lambda}, r_H^{*,\lambda})$ is continuous in λ. This is because, for every θ: (i) the optimal signal $q^{*,\lambda}(\theta), r_H^{*,\lambda}(\theta)$ is continuous in λ,[31] and (ii) V is continuous in $q(\theta)$ and $r(\theta)$. Since we started by assuming that $V(q^{*,0}, r_H^{*,0}) < 0$, the continuity of $V(q^{*,\lambda}, r_H^{*,\lambda})$ in λ implies that there exists $\lambda^* > 0$ for which $V(q^{*,\lambda^*}, r_H^{*,\lambda^*}) = 0$. By the Lagrangian sufficiency theorem, the optimal signal is then characterized by the functions q^{*,λ^*} and r_H^{*,λ^*}.

For completeness of the argument we now explain why the Lagrangian sufficiency theorem,

[31] To see this, notice that for any θ, the values of $q^*(\theta)$ and $r^*(\theta)$ depend on λ only thorough $w(\theta, \lambda)$, and as the analysis in Part I suggests they are continuous in w.

adapted to our problem's setting, implies that $q^{*,\lambda^*}, r_H^{*,\lambda^*}$ are the optimal signal functions. To see this, notice that since $V(q^{*,\lambda^*}, r_H^{*,\lambda^*}) = 0$ it follows that $\mathcal{L}(q^{*,\lambda^*}, r_H^{*,\lambda^*}, \lambda^*) = \frac{1}{1+\lambda^*} W(q^{*,\lambda^*}, r_H^{*,\lambda^*})$. Since q^{*,λ^*} and r_H^{*,λ^*} maximize the integrand of $\mathcal{L}(\cdot,\cdot,\lambda^*)$ *pointwise* (that is, for every type profile θ), $\mathcal{L}(q^{*,\lambda^*}, r_H^{*,\lambda^*}, \lambda^*) \geq \mathcal{L}(q, r_H, \lambda^*)$ for any two signal functions q and r_H. Since our problem's constraint requires the signal functions q and r_H to generate a nonnegative aggregate virtual surplus (i.e., $V(q, r_H) \geq 0$), it follows that $\mathcal{L}(q, r_H, \lambda^*) = \frac{1}{1+\lambda^*} W(q, r_H) + \frac{\lambda^*}{1+\lambda^*} V(q, r_H) \geq \frac{1}{1+\lambda^*} W(q, r_H)$. Taken together, these observations imply that $W(q^{*,\lambda^*}, r_H^{*,\lambda^*}) \geq W(q, r_H)$ for any two signal functions q, r_H that satisfy $V(q, r_H) \geq 0$.■

Proof of Proposition 3

By Corollary 2, there exists an optimal auxiliary mechanism in which truth-telling is a dominant strategy. Consider an actual mechanism with the same functions q^*, r_H^* and r_L^* and the same transfer rules (the only difference between the auxiliary and actual mechanisms is that in the latter the players are not bound by their report in the ensuing voting game). We will show that truth-telling is a dominant strategy also in the actual mechanism.

Assume, by contradiction, that truth-telling is not a dominant strategy in the actual mechanism. This means that there is a type θ_i of player i that prefers to report some $\theta_i' \neq \theta_i$ in the actual mechanism, but not in the auxiliary mechanism, when the other players report some θ_{-i} (which may not coincide with their true types).

It cannot be that $q^*(\theta_i', \theta_{-i}) = 0$ (to simplify the notation we omit throughout this proof the dependence of q^*, r^* and r_L^* on the value of λ^*). To see why, note that when no information is acquired (i.e., $q^*(\theta_i', \theta_{-i}) = 0$) player i prefers the action $a = 0$ in the voting game that follows the actual mechanism. But this is precisely the action that the auxiliary mechanism chooses when $q^*(\theta_i', \theta_{-i}) = 0$. Since player i does not want to deviate and report θ_i' in the auxiliary mechanism, he has no incentive to do so in the actual mechanism.

Suppose that $q^*(\theta_i', \theta_{-i}) > 0$. When the posterior belief $r_L^*(\theta_i', \theta_{-i})$ is realized, the auxiliary mechanism votes for $a = 0$ on player i's behalf. But since $r_L^*(\theta_i', \theta_{-i}) < p$ this is also the action that player i prefers in the voting game that follows the actual mechanism. Suppose then that the posterior $r_H^*(\theta_i', \theta_{-i})$ is realized. Recall that since signals that are purchased in the optimal auxiliary mechanism are SP, $r_H^*(\theta_i', \theta_{-i}) \geq 1 - (\theta_i', \theta_{-i})^{(n-m+1)} \geq p$. If for such a posterior, player i votes for $a = 1$ in the second-stage game following the actual mechanism, then again his action coincides with the action that the auxiliary mechanism chooses for him. Therefore, for i to have a profitable deviation in the actual mechanism but not in the auxiliary mechanism, it must be the case that after $r_H^*(\theta_i', \theta_{-i}) > 1 - (\theta_i', \theta_{-i})^{(n-m+1)}$ player i prefers to vote for $a = 0$. This means that player i of type θ_i strictly gains by increasing the chances of the default action. He may further increase his net utility if $m(\theta_i', \theta_{-i}) + t_i(\theta_i', \theta_{-i}) < m(\theta_i, \theta_{-i}) + t_i(\theta_i, \theta_{-i})$. Since by the monotonicity of q^* we have $q^*(0, \theta_{-i}) \leq q^*(\theta_i, \theta_{-i})$, and since $m(0, \theta_{-i}) + t_i(0, \theta_{-i}) \leq m(\theta_i, \theta_{-i}) + t_i(\theta_i, \theta_{-i})$ (which immediately follows from the fact that type 0 does not want to report θ_i in the auxiliary mechanism), it must be the case that the most profitable deviation is to report $\theta_i' = 0$.

If $q^*(0, \theta_{-i}) < q^*(\theta_i, \theta_{-i})$ or $m(0, \theta_{-i}) + t_i(0, \theta_{-i}) < m(\theta_i, \theta_{-i}) + t_i(\theta_i, \theta_{-i})$ then player i has a profitable deviation already in the auxiliary mechanism by reporting that his type is 0. This contradicts truth-telling being a dominant strategy. Otherwise, player i is indifferent between reporting the truth and his most profitable deviation in the actual mechanism, contradicting our initial assumption that player i has a profitable deviation in the actual mechanism. We have therefore established that truth-telling is a dominant strategy in the actual mechanism.

Finally, note that in the optimal auxiliary mechanism in which truth-telling is a dominant strategy the budget balancedness constraint is satisfied only ex ante. Therefore, the budget of the actual mechanism defined above is also balanced only ex ante. However, since truth-telling is a dominant strategy in the actual mechanism, it is also a Bayesian Nash equilibrium. Thus, by Borgers (2015, p.47), we can modify the transfers to satisfy ex-post budget balancedness without affecting the interim expected transfers, and hence truth-telling remains a Bayesian Nash equilibrium. Furthermore, the individual rationality of the auxiliary mechanism also carries over to the actual mechanism. Thus, the resulting actual mechanism satisfies incentive compatibility, individual rationality, and budget balancedness ex-post. Since, as we explained in the main text, the expected surplus that is achievable by the optimal actual mechanism is bounded above by the expected surplus that is achievable by the optimal auxiliary mechanism, it follows that the actual mechanism we defined above is the optimal one. ■

Proof of Proposition 4

We prove each of the three cases separately. Specifically, we show that the optimal mechanism results in no signal acquisition for some positive measure of type profiles. And, when κ is sufficiently small, then there is also a positive measure of type profiles for which the optimal mechanism results in acquiring both interior and MSP signals.

Case (i). Consider any type profile θ such that $\theta_i < \varepsilon$ for some small $\varepsilon > 0$ and for all $i \in \{1, \ldots, n\}$. The SP constraint implies that, if a signal is acquired for this type profile, then $r_H \geq 1 - \theta^{(n-m+1)} > 1 - \varepsilon$. To check whether acquiring such a signal is better than staying uninformed, we evaluate the derivative of $\hat{\mathcal{L}}$ (Equation (18)) with respect to q, for values of

$r_H \geq 1-\varepsilon$ and of $q>0$

$$\hat{\mathcal{L}}_1\left(q, r_H\ ;\ w\left(\theta, \lambda^*\right)\right)=r_H-\left(1-w\left(\theta, \lambda^*\right)\right)-\frac{1}{n} c_1\left(q, r_H\right)<1-(1-\varepsilon)-\frac{1}{n} c_1\left(q, r_H\right) \tag{34}$$

$$<\varepsilon-\frac{1}{n} c_1(0,1-\varepsilon), \tag{35}$$

where the first inequality is because $r_H<1$ and $w\left(\theta, \lambda^*\right)<\varepsilon$, and the second inequality is because $c_{11}>0$ and $c_{12}>0$. Since $c_1(0,1)>0$,[32] and because c is continuously differentiable, it follows that for ε sufficiently small, $\varepsilon<\frac{1}{n} c_1(0,1-\varepsilon)=\frac{\kappa}{n} \hat{c}_1(0,1-\varepsilon)$. Thus, for small values of ε, $\hat{\mathcal{L}}_1\left(q, r_H\ ;\ w\left(\theta, \lambda^*\right)\right)<0$ for all of $r_H \geq 1-\varepsilon$ and $q>0$, implying that decreasing q increases the value of $\hat{\mathcal{L}}$. Therefore, acquiring no signal is better than acquiring any informative signal for the type profile θ.

Case (ii). Consider the type profile θ for which $\theta_1=\ldots=\theta_n=(1-p)$. Notice that $w\left(\theta, \lambda^*\right)=1-p$, regardless of λ^*. Notice also that since $1-\theta^{(n-m+1)}=p$, the SP constraint has no bite for the type profile θ, because any signal with $r_H \geq p$ is supermajority persuasive. Therefore, the optimal signal for θ, i.e., $q^*(\theta)$ and $r_H^*(\theta)$, is the signal that maximizes

$$\hat{\mathcal{L}}\left(q, r_H\ ;\ (1-p)\right)=q\left(r_H-p\right)-\frac{1}{n} c\left(q, r_H\right)=q\left(r_H-p\right)-\frac{\kappa}{n} \hat{c}\left(q, r_H\right). \tag{36}$$

When κ is sufficiently small, the maximum of (36) is strictly positive and is attained at $q^*>0$ and $r_H^*>p$. Therefore, for the type profile θ, acquiring a signal is better than staying uninformed.

Consider now any type profile θ' that is "close to" θ in the sense that $\theta_i'>(1-p)-\varepsilon$ for some small $\varepsilon>0$ and for all $i \in\{1, \ldots, n\}$. Since the maximizers of $\hat{\mathcal{L}}$ are continuous in w (because of P3; see also the discussion in footnote 29), and because w is continuous in each of the components of the type profile, it follows that $q_H^*\left(\theta'\right)$ is close to $q_H^*(\theta)$, and $r_H^*\left(\theta'\right)$ is close to $r_H^*(\theta)$. This implies that for the type profile θ', acquiring a signal is better than staying uninformed. Also, the optimal signal for θ' is SP but not MSP. This is because, when ε is sufficiently small, $r_H^*\left(\theta'\right)>p+\varepsilon>1-\left(\theta'\right)^{(n-m+1)}$.

Case (iii). Consider the type profile θ for which $\theta_1=\cdots=\theta_{n-1}=0$ and $\theta_n=1-p$. There exists κ^* such that for all $\kappa<\kappa^*$: (i) the multiplier λ^*, defined in Proposition 2, is close to zero such that $w\left(\theta, \lambda^*\right)>\frac{(1-p)}{2 n}$, and (ii) $\kappa<\frac{(1-p)}{4 \cdot \hat{c}_1(0,1)}$.[33] Suppose that $\kappa<\kappa^*$. Since $m>1$, the SP constraint implies that, if a signal is acquired for the type profile θ, then it must be that

[32] To see this, notice that $c_1(0, p)=0$ and $c_{12}>0$. In our leading specification, where the cost function is proportional to the mutual information between the signal realization and the state, we have that $c_1(0,1)=-\kappa \ln (p)$.

[33] Note that $\hat{c}_1(0,1)$ is finite because $c_1(0,1)$ is finite.

$r_H^*(\theta) = 1$. To check whether acquiring a signal with $r_H = 1$ is better than staying uninformed, we evaluate the derivative of $\hat{\mathcal{L}}$ with respect to q, at $r_H = 1$ and $q = 0$:

$$\hat{\mathcal{L}}_1(0, 1\ ;\ w(\theta, \lambda^*)) = 1 - (1 - w(\theta, \lambda^*)) - \frac{\kappa}{n}\hat{c}_1(0,1) > \frac{(1-p)}{2n} - \frac{\kappa}{n}\hat{c}_1(0,1) > \frac{(1-p)}{4n} > 0,$$

where the first inequality is because $w(\theta, \lambda^*) > \frac{(1-p)}{2n}$ and the second inequality is because $\kappa < \frac{(1-p)}{4 \cdot \hat{c}_1(0,1)}$. Thus, when $r_H^*(\theta) = 1$, the value of $q^*(\theta)$ is strictly positive, and acquiring a signal with $r_H = 1$ is better than staying uninformed. The fact that $c_2(q, 1) = \infty$ for all q implies that the derivative of $\hat{\mathcal{L}}$ with respect to r_H, evaluated at $r_H = 1$ and $q = q^*(\theta)$, is negative, i.e., $q^*(\theta) - \frac{1}{n}c_2(q^*(\theta), 1) < 0$. Therefore, the signal $q^*(\theta), r_H^*(\theta)$ is not interior and the MSP constraint is strictly binding.

Consider now any type profile θ' that is "close to" θ in the sense that $\theta'_i < \varepsilon$ and $\theta'_n > (1-p) - \varepsilon$ for some small $\varepsilon > 0$ and for all $i \in \{1, \ldots, n-1\}$. As before, the fact that the maximizers of $\hat{\mathcal{L}}$ are continuous in w, and w is continuous in each of the components of θ, implies that the interior maximizer of $\hat{\mathcal{L}}$ violates SP, and that in the optimal signal MSP is binding, i.e., $r_H^*(\theta') = 1 - (\theta')^{(n-m+1)}$. Thus, $r_H^*(\theta')$ is close to $r_H^*(\theta')$, and $q^*(\theta')$ is close to $q^*(\theta)$, and therefore acquiring an MSP signal is better than staying uninformed. ■