Eliciting Informed Preferences

Modibo Camara Nicole Immorlica Brendan Lucier

November 8, 2024

Preliminary and incomplete. Please do not circulate.

Abstract

If people find it costly to evaluate the options available to them, their choices may not directly reveal their preferences. Yet, it is conceivable that a researcher can still learn about a population's preferences with careful experiment design. We formalize the researcher's problem in a model of robust mechanism design where it is costly for individuals to learn about how much they value a product. We find that it is sometimes possible to identify the population's average values, but that is all. Nothing else can be identified. Finally, we apply our results to social choice and propose a way to combat uninformed voting.

Contents

1 Introduction

People often find it costly to process information about the goods they consume, the services they receive, and the policies that affect them. As a result, their choices may not directly reveal the preferences they would have had if they had processed all the information available to them. To what extent can researchers still learn about a population's preferences through careful experiment design?

Learning about a population's preferences – i.e., their informed preferences – is important for a number of applications. For example, take market research. Suppose that OpenAI is about to launch a product called GPT-5 and wants to forecast demand. In the present, consumers do not know how much value they would get from GPT-5, and find it costly to test the product. In the future, consumers become better informed (as subscribers gain first-hand experience and others rely on second-hand experience). Can we design an experiment that helps OpenAI forecast future demand?

We formalize the researcher's problem as robust mechanism design with information acquisition. There is a sample of agents, and they do not necessarily know their values from some product. The researcher wants to learn about these values, but does not know their distribution, including whether and how they are correlated across agents. Agents can learn about their own and potentially about others' values by acquiring costly signals. For most of our results, we impose little structure on the information acquisition technology, but we always assume that agents can learn their own values at some finite cost.

We use this model to tightly characterize what researchers can learn about the population's preferences. The researcher is able to design any experiment that he wishes, recruit as many agents as he wishes, and spend as much money as he wishes. He *elicits* a given statistic of the population's values if, for every equilibrium, the agents' choices in the experiment identify that statistic.

It turns out that what the researcher can learn is quite limited. The researcher can elicit the population's average value from the product, provided that a minority of agents are initially well-informed (Theorem 2 and Proposition 1). But he cannot elicit any other statistic (Theorem 1), and cannot even elicit the average unless there is a well-informed minority (Proposition 2). Without additional assumptions, many basic statistics – like quantity demanded by an informed population – are not revealed by choices, no matter how carefully-designed those choices are.

We illustrate the value of our positive results in an application to social choice. Specifically, consider the problem of uninformed voting. Empirical evidence suggests many voters are poorly informed (e.g., Delli Carpini and Keeter 1996, Angelucci and Prat 2024) and would vote differently if they were better informed (e.g., Bartels 1996, Fowler and Margolis 2014). Building on our previous results, we design mechanisms that incentivize voters to become better informed. Under some conditions, these mechanisms obtain nearly first-best welfare (Theorems 3 and 4).

We now discuss the model and results in more detail.

Model. We formalize the researcher's problem as robust mechanism design with information acquisition.

The researcher asks a random sample of n agents to participate in a mechanism. There is a product – or more generally, a finite set of alternatives – and each agent i derives value v_i from that product. The distribution of values in the population is unknown to the researcher, and may involve complex patterns of heterogeneity and correlation across agents. The agents send messages to the researcher, and then the researcher makes individualized allocations and transfers.

Initially, agents do not know their own values. Instead, each agent i has some private information and may learn more – both about her own value v_i and perhaps about others' values v_{-i} – by acquiring costly signals. We start with a special case where each agent i decides whether or not to learn her value v_i at some finite cost c_i. We extend most of our results to a general case that allows for the possibility that agents have access to other signals that are unknown to the researcher.

The researcher wants to learn about the distribution of values in the population. More precisely, he wants to learn about demand for the product in the counterfactual where every agent i knew her value v_i. We focus on statistics that can be represented as moments (e.g., average willingness to pay, or quantity demanded at a fixed price). To learn about a statistic, the researcher can recruit as many agents as he likes, spend as much money as he likes, and allocate as many products as he likes.

A statistic is *elicitable* if there exists a mechanism where the observed choices identify that statistic. More precisely, there exists a sequence of mechanisms, indexed by the sample size n, where messages sent in equilibrium pin down the statistic's value in the limit as $n \to \infty$. This must hold for any value distribution and every Bayes-Nash equilibrium, and requires exact (not just partial) identification.

Next, we completely characterize the conditions in which elicitation is possible. By conditions, we mean properties of the statistic, as well as properties of the information acquisition technology (e.g., how the costs c_i are distributed).

Non-Elicitable Statistics. It turns out that most statistics are not elicitable. More precisely, Theorem 1 says that a statistic is not elicitable if it is not equivalent to the population's average value, except in trivial cases where information is always free.

Theorem 1 implies that many natural and economically-relevant statistics are not elicitable, at least not in general. In our motivating example, it means that OpenAI would not be able to forecast the quantity demanded at a given price. To overturn this negative result, the researcher would have to impose stronger assumptions or give up on exact identification, neither of which is needed in classical models of preference elicitation (e.g., Becker et al. 1964; Clarke 1971).

Having established that the population's average value is the only statistic that we can hope to elicit, it is natural to ask whether and when the average is elicitable.

Eliciting the Average. We find that the population's average value is only elicitable sometimes. As a starting point, Theorem 2 says that it is always *weakly elicitable*. By weakly elicitable, we mean that the researcher can identify the statistic if he assumes that agents follow his preferred equilibria.

Building on Theorem 2, we show that the population's average value is elicitable if and only if an *informed minority condition* holds (Propositions 1-2). This condition says that, for every type of agent i, there is a positive probability that their cost of information c_i is zero. In our motivating example, this may hold if there is a minority of "early adopters" who are intrinsically motivated to try out GPT-5.

To elicit the average, the researcher presents agents with an incentivized survey and asks them, before completing the survey, to predict the survey results. We call this mechanism *BDM-with-betting*. It has two stages. In the second stage, agents report their willingness to pay in a standard Becker–DeGroot–Marschak (BDM) mechanism. In the first stage, agents predict the average reported willingness to pay. Prediction accuracy is rewarded according to a proper scoring rule, and the researcher uses the average reported value to estimate the population's average value.

The BDM-with-betting mechanism is relatively simple – even naive – and this begs the question: why does it work? It is simpler than than other mechanisms with comparable features (e.g., Crémer and McLean 1985; Miller et al. 2005). It is naive because it incentivizes agent i to learn about others' willingness to pay, not her own value v_i. Nonetheless, BDM-with-betting works because it incentivizes agents to acquire information up until the point where their reporting errors are uncorrelated. If these errors – the difference between an agent's value and her reported value – are uncorrelated, they vanish in the aggregate. Essentially, betting restores the "wisdom of the crowd", which usually requires strong distributional assumptions.

All in all, our results suggest that what the researcher can learn is quite limited. However, there is a setting where learning the population's average value is all that is needed. That setting is social choice.

Application to Social Choice. We turn to the problem of social choice, where a planner chooses a single alternative on behalf of a population. He forms a committee consisting of n agents sampled from the population (e.g., a citizen's assembly).

We are motivated by the problem of uninformed voting. Existing electoral systems do not give voters much of a reason to become informed about the policies and candidates that appear on their ballot. This is not just a theoretical concern. Empirically, many voters are poorly informed (e.g., Delli Carpini and Keeter 1996, Angelucci and Prat 2023), and making them better-informed could plausibly affect vote margins (e.g., Lau and Redlawsk 1997, Fowler and Margolis 2014).

It turns out that the methods we developed to elicit preferences can be used to design social choice mechanisms that are robust to costly information processing. Theorems 3 and 4 say that the planner can identify the welfare-maximizing alternative, if the informed minority condition holds. The informed minority may consist of voters who are intrinsically interested in policy or civic-minded. This condition is consistent with empirical findings (e.g., Delli Carpini 2000).

In particular, we motivate a simple mechanism – *majority-rule-with-betting* – by restricting attention to binary alternatives and imposing a distributional assumption (Theorem 3). This mechanism is analogous to BDM-with-betting. First, we ask agents to predict the eventual vote margin. Second, they report their preferred alternative. The planner chooses the alternative with majority support.[1] Transfers are only used to reward accurate predictions; there is no vote buying.

There are two key reasons why this mechanism works. The first reason is the

[1] More precisely, the planner chooses this alternative with high probability. See Section 5.3 for details.

same as before: betting restores the “wisdom of the crowd”, assuming truthful voting. The second reason is new. Recall that uninformed voters may not vote truthfully if they believe that how others vote may be informative about their own preferences (e.g., Austen-Smith and Banks 1996). It turns out that betting restores incentives for truth-telling. Intuitively, an uninformed agent i that believes her value v_i is correlated with the vote margin is leaving money on the table. If the betting stakes are large, she prefers to learn her value v_i in order to better predict the vote margin.

Related Literature. We build on three research areas in economics, marketing, computer science, and political science. We briefly (and incompletely) discuss the related literature now, and leave a fuller discussion to Section 6.

First, we contribute to research on preference elicitation (e.g., Becker et al. 1964; Clarke 1971) by relaxing the assumption that individuals know their own preferences. We identify sharp limits on our ability to learn about preferences through observed choices. Moreover, BDM-with-betting follows a line of work on preference and belief elicitation that (implicitly) asks agents to bet on each others' reports. Most existing work uses betting to incentivize information revelation (e.g., Crémer and McLean 1985; Prelec 2004; Miller et al. 2005; Pakzad-Hurson 2022). In contrast, BDM-with-betting only uses betting to incentivize information acquisition.

Second, we contribute to research on information acquisition in games. In particular, we study mechanism design with information acquisition (e.g., Persico 2000; Bergemann and Välimäki 2002). We propose a new model of unstructured information acquisition (also see e.g., Carroll 2019; Denti and Ravid 2023), and find mechanisms that are robust to the details of how agents acquire information.

Third, we contribute to the theory of voting, social choice, and public goods. Historically, this literature has focused on mitigating the negative effects of strategic voting (e.g., Arrow 1950; Groves and Ledyard 1977). We contribute to more recent research on mitigating the negative effects of uninformed voting (e.g., Persico 2004; Feddersen and Sandroni 2006; Gerardi and Yariv 2008; Gershkov and Szentes 2009), largely by relaxing restrictive assumptions (e.g., common values).

Organization. The rest of this paper is organized as follows. Section 2 presents the model. Section 3 shows that statistics other than the average are not elicitable. Section 4 characterizes when the average is elicitable. Section 5 applies our results to social choice. Section 6 discusses the related literature. Section 7 concludes.

The appendices contain additional results and proofs. Appendix A generalizes our results by relaxing or dropping assumptions. Appendix B outlines the proofs of our main results. The Supplemental Appendix contains all omitted proofs.

2 Model

To formalize the problem of preference elicitation, we develop a model of robust mechanism design with information acquisition. This model accommodates the wide range of experiments that a researcher might run.

In the model, there is a researcher and n agents. Agent i receives an alternative

$$x_i \in \mathcal{X} = \{0, 1\}$$

We restrict attention to two alternatives for now, but consider multiple alternatives in Appendix A. Given alternative x_i, agent i receives value

$$v_{ix_i} \in \mathcal{V} = [-\bar{v}, \bar{v}]$$

where we normalize $v_{i0} = 0$ and let v_i refer to v_{i1}. We often refer to alternative $x = 1$ as a product and $x = 0$ as an outside option.

As a running example, suppose the firm OpenAI develops a new product called GPT-5. A consumer's value v_i reflects how much utility she would get from GPT-5. Before the product is launched, the consumer may not know enough about it (e.g., its functionalities, use cases, flaws, etc.) to ascertain her value v_i.

Natually, an uninformed consumer's willingness to pay for GPT-5 may differ from her value v_i. Although OpenAI ultimately cares about willingness to pay, the value v_i is likely to govern demand in the medium- and long-run. Consumers that do not value a product are unlikely to buy it again, or renew their subscription. Even a first-time customer is likely to become better-informed in the years after a product launch, as early adopters share their experiences.

Next, we turn to agents' information about their own and each others' preferences, and the process by which they acquire more information.

2.1 Information Acquisition

We formalize the agent's information and information acquisition in flexible way. In particular, we want the model to accommodate complicated patterns of heterogeneity and correlation across agents, which are difficult to rule out in practice.

Initially, agent i may not know much about her own value v_i. What she does know is captured by her *private type* θ_i. Moreover, she knows that both her value v_i and other agents' values v_{-i} may be correlated with a *hidden state* ω. Finally, she can learn more about her value v_i and the state ω by acquiring costly *signals* s_i. We now define these terms and other model primitives more carefully.

Private Types. Private types $\theta_i \in \Theta$ play two roles in the model. First, agent i's type θ_i reflects her initial information and, in particular, may be correlated with v_i. Second, as we clarify below, different types may have different costs of information.

For example, a consumer's type might indicate whether she used GPT-4 or not. Since types may be correlated with values, the kinds of consumers who used GPT-4 may be more likely to value GPT-5. Since different types may have different costs of information, consumers who used GPT-4 may find it easier to learn about GPT-5.

Hidden State. The hidden state $\omega \in \Omega$ also plays two roles. First, it may affect how agents' values v_i are distributed. Second, it may affect how agents' types θ_i are distributed. Neither the researcher nor the agents know the state initially.

For example, the hidden state might indicate whether GPT-5 is generally useful or not. Since the state may be correlated with values, consumers may value GPT-5 more if it is useful. Since the state may be correlated with types, GPT-5 being useful could be correlated with the fraction of consumers that tried GPT-4.[2]

Joint Distribution. The *joint distribution* F captures the correlation between the hidden state, private types, and values.

We impose a conditional independence assumption. Informally, we assume that any correlation across pairs of agents i and j is generated by the hidden state ω, which acts as a correlating device. Formally, we assume that agent i's characteristics (v_i, θ_i) and agent j's characteristics (v_j, θ_j) are i.i.d. conditional on the state ω.

[2] For instance, one causal story that leads to this correlation is that (i) GPT-4 being useful implies that GPT-5 is useful and (ii) GPT-4 being useful implies that a larger fraction of consumers tried GPT-4.

Given this assumption, it is enough to define F as a distribution over the state ω, type θ_i, and values $v_i \in \mathcal{V}$ for a single agent i. Formally,

$$F \in \mathcal{F} = \Delta(\Omega \times \Theta \times \mathcal{V})$$

The state ω, type profile $\theta = (\theta_1, \ldots, \theta_n)$, and value profile $v = (v_1, \ldots, v_n)$ are generated in two steps. First, the state ω is drawn from marginal distribution F_Ω. Then, pairs (θ_i, v_i) are drawn independently from conditional marginal distribution $F_{\Theta \times \mathcal{V}|\omega}$.

The conditional independence assumption ensures that there is a well-defined sampling distribution $F_{\mathcal{V} \times \Theta|\omega}$ that does not vary with the sample size n.[3]

Signals. To learn more about their own values and the state, agents can acquire signals from a collection $\mathcal{S}$. For agent i, a signal $s_i \in \mathcal{S}$ is a function of the state ω, her type θ_i, and her value v_i. In principle, it could output an arbitrary message, but it is convenient to let the output be a finite sequence of real numbers. More precisely,

$$s_i : \Omega \times \Theta \times \mathcal{V} \to \mathbb{R}^*$$

It is also convenient to think of signals s_i as random variables that map samples $(\omega, \theta_i, v_i) \sim F$ to *signal realizations* $s_i(\omega, \theta_i, v_i)$.

By observing the signal realization $s_i(\omega, \theta_i, v_i)$, agent i may learn directly about the state ω and her own values v_i. She may also learn indirectly about other agents j, since their characteristics (θ_j, v_j) may be correlated with the state ω. Finally, it is worth noting that the same signal s_i can convey different information to different types θ_i. This is because signal realizations can depend on types.

For example, a consumer could learn about their value from GPT-5 by using it. They might think of different use cases (e.g., writing, coding, translation, search, etc.) and evaluate how well it performs in each case. If she learns about the state, she may learn whether GPT-5 is generally useful or not. If GPT-5 is generally useful, then she may infer that other consumers' values are likely to be high. Finally, since the same signal can convey different information to different types, consumers who tried GPT-4 may identify subtle flaws in GPT-5's responses since they know what to look for.

[3] Some readers may wonder why we even need to specify a marginal distribution F_Ω over the state. The reason is that agents' beliefs over the state affects their equilibrium behavior.

Cost Function. A *cost function* $c_i \in \mathcal{C}$ determines the cost of acquiring signal $s_i \in \mathcal{S}$ for agent i of type θ_i. Signals that are unavailable have infinite costs. Formally,

$$c_i : \mathcal{S} \times \Theta \to \mathbb{R}_+ \cup \{\infty\}$$

Each agent i's cost function is drawn independently from the *cost distribution* G. However, the cost of acquiring any given signal s_i may be correlated across agents i. This is because agent i's cost $c_i(s_i, \theta_i)$ of acquiring signal i depends on her type θ_i, which may be correlated with the types θ_{-i} of other agents.[4]

Model Specification. A *model specification* $(\Omega, \Theta, \mathcal{S}, F, G)$ consists of state space Ω, type space Θ, collection of signals $\mathcal{S}$, joint distribution F, and cost distribution G. An *information model* $(\Omega, \Theta, \mathcal{S}, G)$ omits the joint distribution F.

The agents know the model specification, but the researcher may not. The researcher seeks a robust way to elicit preferences.

2.2 Mechanism Design

To elicit preferences, the researcher can commit to a *mechanism*. As usual, agents send messages to the researcher and the researcher uses those messages to determine the alternative x_i and transfers t_i for each agent i.

Definition 1. *Fix a set of message profiles $M = (M_i)_{i=1}^n$. A* mechanism $(\boldsymbol{x}, \boldsymbol{t})$ *consists of:*

1. *Allocation rules that maps message profiles to distributions over alternatives x, i.e.,*

$$\boldsymbol{x}_i : M \to \Delta(\mathcal{X})$$

2. *Transfer rules that map message profiles to distributions over transfers, i.e.,*

$$\boldsymbol{t}_i : M \to \Delta(\mathbb{R})$$

[4] Readers may wonder why we explicitly allow cost functions c_i, c_j to differ across agents. After all, we could always expand the types from θ_i to (θ_i, c_i), and then define a single cost function c that says an agent i of type (θ_i, c_i) has costs given by $c_i(\cdot, \theta_i)$. The reason is that we will eventually assume that the type space is finite (see Section 2.4), in which case this construction would be with loss.

For convenience, let $\mathbf{x}_i(m)$ refer to both the distribution over alternatives as well as the realized alternative. Do the same for transfers $\mathbf{t}_i(m)$.

The agents' strategies must specify both what signals they acquire and what messages they send to the researcher. First, we review the timing of the game:

1. Agent i observes her type θ_i and cost function c_i.
2. Agent i chooses a signal s_i to acquire.
3. Agent i observes her signal realization $s_i(\omega, \theta_i, v_i)$.
4. Agent i sends a message m_i.

Now, we can formally define strategies.

Definition 2. *Agent i's* strategy $(\boldsymbol{s}_i, \boldsymbol{m}_i)$ *pairs a* signal rule *that maps i's type and cost function to a distribution over signals, i.e.,*

$$\boldsymbol{s}_i : \Theta \times \mathcal{C} \to \Delta(\mathcal{S})$$

with a message rule *that maps the signal realization, i's type, and i's cost function to a distribution over messages, i.e.,*

$$\boldsymbol{m}_i : \mathbb{R}^* \times \Theta \times \mathcal{C} \to \Delta(M_i)$$

For convenience, let $\mathbf{s}_i(\theta_i, c_i)$ refer to both the distribution over signals chosen as well as the realized signal s_i. Do the same for messages $\mathbf{m}(s_i(\omega, \theta_i, v_i), \theta_i, c_i)$.

Agent i's utility depends on her values v_i, her alternative x_i, the transfers t_i she receives, and the cost $c_i(s_i, \theta_i)$ for her type θ_i of the signal s_i that she acquires. That is,

$$u_i(v_i, x_i, t_i, s_i, \theta_i, c_i) = v_{ix_i} + t_i - c_i(s_i, \theta_i)$$

In turn, agent i's expected utility from a strategy profile $(\mathbf{s}, \mathbf{m})$ is

$$U_i(\mathbf{s}, \mathbf{m}, \mathbf{x}, \mathbf{t}) = \mathrm{E}[u_i(v_i, \mathbf{x}_i(m), \mathbf{t}_i(m), s_i, \theta_i, c_i)]$$

with signals $s_i = \mathbf{s}_i(\theta_i, c_i)$ and messages $m_i = \mathbf{m}_i(s_i(\omega, \theta_i, v_i), \theta_i, c_i)$. The expectation is taken with respect to the joint distribution F, the cost distribution G, any randomness in the strategy profile $(\mathbf{s}, \mathbf{m})$, and any randomness in the mechanism $(\mathbf{x}, \mathbf{t})$.

With this notation in hand, we can specify our solution concept.

Definition 3. *A strategy profile* $(\boldsymbol{s}, \boldsymbol{m})$ *is a* Bayes-Nash equilibrium *of mechanism* $(\boldsymbol{x}, \boldsymbol{t})$ *if every agent i prefers her strategy* $(\boldsymbol{s}_i, \boldsymbol{m}_i)$ *to every alternative strategy* $\left(\boldsymbol{s}'_i, \boldsymbol{m}'_i\right)$. *That is,*

$$U_i\left(\boldsymbol{s}, \boldsymbol{m}, \boldsymbol{x}, \boldsymbol{t}\right) \geq U_i\left(\left(\boldsymbol{s}'_i, \boldsymbol{s}_{-i}\right), \left(\boldsymbol{m}'_i, \boldsymbol{m}_{-i}\right), \boldsymbol{x}, \boldsymbol{t}\right)$$

Having specified a solution concept, we can formalize the researcher's problem.

2.3 Elicitable Statistics

The researcher wants to learn about a population's values over alternatives.

Specifically, the researcher wants to learn about $F_{\mathcal{V}|\omega}$, the marginal distribution of values conditional on the realized state ω. This is the distribution of interest because it describes realized demand for a product by an informed population. For example, if the realized state ω indicates that GPT-5 is useful, the researcher may learn that demand is high. But he is not trying to learn whether demand would also have been high in the counterfactual state where GPT-5 was not useful.

We focus on learning *statistics* that are moments of the conditional marginal distribution $F_{\mathcal{V}|\omega}$. Two important objects that this leaves out are the values $(v_1, \ldots, v_n)$ of the sampled agents and the distribution $F_{\mathcal{V}|\omega}$ itself. Nonetheless, it follows immediately from Theorem 1 that these two statistics cannot be learned.

Definition 4. *A* σ-statistic *is a moment of the conditional marginal distribution* $F_{\mathcal{V}|\omega}$. *Formally, given a function* $\sigma : \mathcal{V} \to \mathbb{R}$, *the* σ*-statistic is the random variable*

$$\mathrm{E}_F[\sigma(v_i) \mid \omega]$$

Remark 1. To avoid confusion, we rephrase Definition 4. As long as the function σ is well-behaved, the σ-statistic is the probability limit of the sample moment, i.e.,

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{i=1} \sigma(v_i)$$

This is simply the expected value of $\sigma(v_i)$ conditional on the realized state.[5]

[5]In particular, the σ-statistic is not a function $\Omega \to \mathbb{R}$. It does not describe the expected value of $\sigma(v_i)$ for every possible state. It is the expected value of $\sigma(v_i)$ conditional on the realized state.

The researcher's data consists of the messages $m_1, \dots, m_n$ sent by the sample of n agents in equilibrium. Traditionally, a statistic is identified if there exists an estimator whose estimate converges to said statistic in the limit as the sample size n grows. The same is true in our model, except that whether a statistic is identified also depends on the sequence of mechanisms that generates the data.

Definition 5. *Fix an information model* $(\Omega, \Theta, \mathcal{S}, G)$ *and a sequence of mechanisms* $(\boldsymbol{x}^n, \boldsymbol{t}^n)$*. Then:*

1. *A* σ*-statistic is* identified *if there is a sequence of functions* f_n *such that, for all joint distributions* F *and equilibria* $(\boldsymbol{s}^n, \boldsymbol{m}^n)$*,*

$$f_n(m_1, \dots, m_n) \to_p \mathrm{E}[\sigma(v_i) \mid \omega] \tag{1}$$

   with messages $m_i = \boldsymbol{m}_i(s_i(\omega, \theta_i, v_i), \theta_i, c_i)$ *and signals* $s_i = \boldsymbol{s}_i(\theta_i, c_i)$*.*

2. *A* σ*-statistic is* weakly identified *if there is a sequence of functions* f_n *such that, for all joint distributions* F*, condition* (3) *holds for some equilibria* $(\boldsymbol{s}^n, \boldsymbol{m}^n)$*.*

Just as the researcher is free to recruit as many agents as he wishes, he is free to choose any mechanism he wishes. He can *elicit* a statistic if he can design some sequence of mechanisms for which the statistic is identified.

Definition 6. *Fix an information model* $(\Omega, \Theta, \mathcal{S}, G)$*. A* σ*-statistic is* (weakly) elicitable *if it is (weakly) identified for some sequence of mechanisms* $(\boldsymbol{x}^n, \boldsymbol{t}^n)$*.*

This definition allows the researcher's mechanism to depend on the information model. That possibility makes our negative results stronger. However, it may be unreasonable if the researcher does not know much about the information that agents can acquire or their costs of acquiring it. Fortunately, our positive results hold even if the researcher does not have this information (see Remarks 2 and 3).

In practice, the researcher may face challenges that are not captured in this model. He may only have access to a small number of agents, or find it costly to recruit more. He may not be able to design any mechanism he wishes, or face budget constraints. As a first step, we ask what an empowered researcher can achieve when these particular challenges are not present. Even so, we find that what the researcher can achieve is quite restricted (see Theorem 1 and Proposition 2).

Next, we conclude the description of the model by imposing two assumptions.

2.4 Assumptions

We make two simplifying assumptions.

First, whenever possible, we assume that spaces are finite. We maintain this assumption throughout the paper.[6]

Assumption 1. *Restrict attention to information models* $(\Omega, \Theta, \mathcal{S}, G)$ *and statistics* σ *where:*

1. *The state space* Ω *is finite.*
2. *The type space* Θ *is finite and has at least three elements.*[7]
3. *The collection of signals* $\mathcal{S}$ *is finite.*
4. *The cost distribution* G *has finite support.*
5. *The statistic* σ *has a finite discontinuity set.*

Second, we restrict attention to binary signal spaces where each agent i either learns her value v_i or learns nothing. This is not necessary for most of our results, but it is a simple and intuitive special case. Appendix A.1 relaxes this assumption.

To state Assumption 2, we refer to a particular kind of signal.

Definition 7. *A signal* s_i *is* revealing *if agent i learns her values* v_i *after acquiring it. That is, for all states* ω *and type profiles* θ*,*

$$\mathrm{Var}_F[v_i \mid s_i(\omega, \theta, v_i), \theta_i] = 0^{8} \tag{2}$$

A revealing signal for agent i does not necessarily reveal everything. It does not necessarily reveal the values v_{-i} of other agents, the types of other agents θ_{-i}, or the hidden state ω. An agent with access to a revealing signal can precisely learn about her own preferences, but not necessarily more than that.

[6]Although there may be value in extending our results to infinite spaces, it is not clear how it would add to our core economic insights. In particular, when it comes to our negative results, it is unlikely that a researcher could ever empirically distinguish between the finite and infinite cases.

[7]We assume $|\Theta| \geq 3$ because our proof of Theorem 1 requires sufficient heterogeneity among agents. It is not clear whether this is a necessary condition for that result to hold.

[8]It is not necessary to condition the variance on agent i's cost function c_i, even though this is part of the information she has available. This is because c_i is independent of v_i conditional on θ_i.

Assumption 2. *Restrict attention to information models* $(\Omega, \Theta, \mathcal{S}, G)$ *where any given type* θ_i *has only two signals.*[9] *The first signal* $s_i^{\emptyset}$ *is uninformative and costless, i.e.,*

$$s_i^{\emptyset}(\omega, \theta_i, v_i) = 0 \quad \textit{and} \quad c_i(s_i^{\emptyset}, \theta_i) = 0$$

The second signal $s_i^{\theta_i}$ *is revealing. Its cost is finite and drawn from distribution* G^{θ_i}*, i.e.,*

$$c_i(s_i, \theta_i) \sim G^{\theta_i} \quad \textit{where} \quad \operatorname{supp}\left(G^{\theta_i}\right) \subseteq [0, \infty)$$

This binary signal space is essentially the simplest version of our model that captures both the challenges of eliciting preferences and, where possible, the intuition for how to accomplish it. This makes it useful for exposition.

In Appendix A.1, we present our general model of unstructured information acquisition. We replace Assumption 2 with two new assumptions. The first says that the signal space includes (but need not be limited to) a revealing signal. This follows models in robust mechanism design where is a known action available to the agent, but there may also be other unknown actions (e.g., Carroll 2015, Carroll 2019). The second assumption says that the agent can combine multiple signals into one.

All of our positive results (Theorems 2-4) hold in the general model. However, we do not generalize our negative result (Theorem 1). There are two reasons for this. First, it refers to a non-triviality condition that would be more complicated to state for non-binary signal spaces. Second, if anything, negative results that hold in simpler settings tend to be more compelling.

Having described the model, we can now turn to our first result.

3 Non-Elicitable Statistics

It turns out that most statistics are not elicitable, except in trivial cases.

To formalize this result, we begin by identifying non-trivial cases. These are cases where information acquisition is costly with some positive probability.

Definition 8. *Fix an information model* $(\Omega, \Theta, \mathcal{S}, G)$*. The* non-triviality condition *holds if there exists a type* $\theta_i \in \Theta$ *whose cost of the revealing signal is positive with*

[9] More precisely, for each type θ_i and cost function $c_i \in \operatorname{supp}(G)$ there are only two signals s_i that have finite cost $c_i(s_i, \theta_i) < \infty$ with positive probability.

positive probability. That is,

$$\Pr_G\left[c_i\left(s_i^{\theta_i}, \theta_i\right) = 0\right] < 1$$

Our negative result applies for any statistic that is not an *average* – that is, any statistic that is not some affine transformation of the population's average value.

Definition 9. *A σ-statistic is an* average *if $\sigma(v_i) = \alpha v_i + \beta$ for some constants α, β.*

Many natural and economically-relevant statistics are not averages. For example, consider the quantity demanded by an informed population at some fixed price p. This corresponds to the σ-statistic where[10]

$$\sigma(v_i) = \mathbf{1}(v_i \geq p)$$

This statistic is likely to be relevant for firms trying to assess market size or develop pricing strategies for new products. In our running example, it captures medium- and long-run demand as consumers gain first- and second-hand experience with GPT-5. Sadly, Theorem 1 says that the firm cannot elicit this statistic.

Theorem 1. *Fix a σ-statistic that is not an average. If the non-triviality condition holds, then it is not elicitable.*

The strength of Theorem 1 reflects three important features of our model. First, we insist on point identification rather than partial identification. It may be possible to obtain informative bounds on for a given statistic, even if it is not point identified. Second, we insist on identification in every equilibrium. If we only insisted on weak identification, which ignores "bad" equilibria, then it may be possible to elicit more statistics. Third, we insist on identification for all joint distributions F. With stronger distributional assumptions, it may be possible to elicit more statistics.

Next, we provide a high-level intuition for why most statistics are not elicitable. Readers interested in understanding how to convert this intuition into a proof should refer to the proof outline in Appendix B.1.

[10]Technically, the quantity demand is the σ-statistic multiplied by the number of potential customers.

3.1 Intuition for Theorem 1

We provide intuition for Theorem 2 using a simple example. Suppose the type and state spaces are both singletons. Let values v_i be normally-distributed, where

$$v_i \sim N(\mu, \nu) \quad \text{i.i.d.}$$

Finally, let the cost of the revealing signal is $\bar{c}$ for all types, where $\bar{c} \gg \nu$.

This setting – where values v_i are uncorrelated across agents i – is an easy case when it comes to eliciting the average. The researcher can ask each agent to make a report $\hat{v}_i$ equal to their willingness to pay.[11] If they do not acquire information, each agent i reports $\hat{v}_i = \mu$ and the average reported value is μ. If they acquire information, each agent i reports $\hat{v}_i = v_i$ and the average reported value still converges to μ by the law of large numbers. In either case, the average μ is identified.

Interestingly, this setting is a hard case when it comes to eliciting σ-statistics other than the average. There are two reasons for this.

1. *In this example, most statistics are hard to identify if agents are not informed.*

   To see this, consider a σ-statistic that is not an average. Typically, the σ-statistic will be sensitive to variance ν. If agents do not acquire information, their willingness to pay depends only on μ, and does not pin down the σ-statistic.

2. *In this example, it is difficult to incentive agents to acquire information.*

   There are two broad reasons for agent i to acquire information. The first reason is to decide whether she is willing to "buy" the product at a given price. This is not sufficient because the information cost $\bar{c}$ far exceeds the agent's uncertainty over her value v_i. The second reason is to better predict the messages m_{-i} sent by other agents. This is not sufficient because agent i's value v_i is independent of any private information that the other agents might have.

One way to circumvent these difficulties is by having agents directly report the variance ν – or more generally, the joint distribution F – and penalize them if any of their reports disagree. This does not require agents to acquire information because F (and therefore, ν) is common knowledge. The problem with this mechanism is that it has many equilibria, most of which are non-truthful.

[11] It is easy to incentivize truth-telling using the Becker–DeGroot–Marschak (BDM) mechanism.

It is possible to imagine many other mechanisms, including ones that take advantage of knowledge of the cost distribution G. For example, let the cost of the revealing signal be zero with 50% probability. Now, it is easy to identify the variance ν. First, ask agents to report their willingness to pay. Second, estimate the mean μ. Finally, estimate the variance ν by excluding all reports $\hat{v}_i = \mu$ from the data.

To prove Theorem 1, we must rule out all possible mechanisms, for all non-trivial cost distributions, and all possible σ-statistics that are not an average. Naturally, this is more involved, and so we leave the proof outline to Appendix B.1. For now, we turn to the one kind of statistic that Theorem 1 did not rule out: the average.

4 Eliciting the Average

We find that the researcher can elicit the population's average value, provided that there is an *informed minority*. This stands in sharp contrast to essentially every other statistic, which the researcher cannot elicit in any non-trivial cases.

First, we establish that the researcher can always weakly elicit the average.

Theorem 2. *Let the σ-statistic be an average. Then it is weakly elicitable.*

The weakness of Theorem 2 is that it only guarantees weak elicitation. In particular, for some joint distributions F, there are equilibria where the σ-statistic is not identified. The partial model specifications for which this is not true are precisely those that satisfy the informed minority condition.

Definition 10. *Fix an information model* $(\Omega, \Theta, \mathcal{S}, G)$*. The* informed minority condition *holds if, for every type $\theta_i \in \Theta$, there is a positive probability that the revealing signal $s_i^{\theta_i}$ costs nothing. That is,*

$$\Pr_G\left[c_i\left(\theta_i, s_i^{\theta_i}\right) = 0\right] > 0$$

For example, the informed minority condition may hold if 20% of a population is already familiar with the product being studied. The researcher cannot distinguish – at least not directly – the 20% that is informed from the 80% that is uninformed. A naive mechanism that failed to account for the costs of information acquisition would collect data that mostly represents the preferences of uninformed individuals.

Although each type of agent must have a positive probability of being familiar with the product, the informed minority need not be representative of the population. For example, suppose there are red and blue agents. The informed minority condition still holds if 30% of red agents and 10% of blue agents are familiar with the product.

The informed minority condition controls whether the average is elicitable. We formalize this in two propositions. Proposition 1 is almost a corollary of Theorem 2.

Proposition 1. *Let the σ-statistic be an average. Then it is elicitable if the informed minority condition holds.*

Proposition 2. *Suppose the non-triviality condition holds and let the σ-statistic be an average. The σ-statistic is not elicitable unless the informed minority condition holds.*

Remark 2. Theorem 2 and Proposition 1 remain true even if the researcher does not know the information model, and Proposition 2 holds even if the researcher does know the information model.[12]

Whether Propositions 1-2 are best seen as positive or negative results depends whether the informed minority condition is plausible in a given application. When it comes to market research, the informed minority may consist of "early adopters" who are intrinsically-motivated to learn about products that are new and unfamiliar. At the same time, this intrinsic motivation may not apply to products that are both unfamiliar and do not seem especially exciting.

The rest of this section provides intuition for these results. In Section 4.1, we describe the mechanism that we use in the proofs of Theorem 2 and Proposition 1. In Section 4.2, we explain at a high level why the mechanism works. Readers interested in understanding these results at a deeper level should refer to the examples in Appendix B.2 and the proof outline in Appendix B.3.

4.1 Mechanism

At a high-level, our mechanism presents agents with an incentivized survey and asks them, before completing the survey, to predict the survey results.

[12] Formally, there is a sequence of mechanisms that, for all information models, weakly identifies the average (Theorem 2). The same sequence of mechanisms will, for all information models satisfying the informed minority condition, identify the average (Proposition 1).

More precisely, we build on the *Becker–DeGroot–Marschak (BDM) mechanism*. This mechanism is widely-used in behavioral experiments to elicit willingness to pay. It asks each agent to report their values $\hat{v}_i$, draws a random price p, and sells the product to the agent if and only if the reported value exceeds the price.

We also rely on *proper scoring rules*. These are rules that can be used to incentivize an agent to report their beliefs about some random variable.[13] Here, the random variable is the average reported value in the BDM mechanism, among agents other than i. More precisely, agent i's beliefs are

$$b_i \in \mathcal{B} = \Delta(\mathcal{V})$$

and the average reported value among agents other than i are

$$\tilde{v}_i = \frac{1}{n-1} \sum_{j \neq i} \hat{v}_j$$

We specialize the definition of proper scoring rules to our setting.

Definition 11. *A* scoring rule *for agent i maps her reported belief $\hat{b}_i \in \mathcal{B}$ and the average reported values $\tilde{v}_i \in \mathcal{V}^m$ to a numerical score, i.e.,*

$$\mathrm{SR} : \mathcal{B} \times \mathcal{V} \to \mathbb{R}$$

It is proper *if she maximizes the expected score by reporting her beliefs truthfully, i.e.,*

$$\forall b_i \in \mathcal{B}, \quad b_i \in \arg\max_{\hat{b}_i} \mathrm{E}_b\left[\mathrm{SR}\left(\hat{b}_i, \tilde{v}_i\right)\right]$$

We can now introduce the *BDM-with-betting mechanism*. It involves two stages. The second stage is simply the BDM mechanism. The first stage asks agents to predict the average of the reported values in the second stage. Agents are paid more if their predictions turn out to be more accurate.

Definition 12. *The* BDM-with-betting mechanism $(\boldsymbol{x}, \boldsymbol{t})$ *is parameterized by a proper*

[13] Proper scoring rules tend to give higher scores to beliefs that assign higher probability to the observed value. There are many known proper scoring rules and they are easy to construct (e.g., McCarthy 1956). We mostly rely on the *continuous ranked probability score* and the *quadratic scoring rule*.

scoring rule SR *and scaling parameter* λ*. Each agent i sends a message*

$$m_i = (\hat{v}_i, \hat{b}_i) \in \mathcal{V} \times \mathcal{B}$$

that consists of a reported value $\hat{v}_i$ *and a reported belief* $\hat{b}_i$*. She receives the product if her reported value exceeds the random price p, i.e.,*

$$\boldsymbol{x}_i\left(\hat{v}, \hat{b}\right) = \mathbf{1}(\hat{v}_i \geq p) \quad \textit{where} \quad p \sim \textsc{Uniform}[-\bar{v}, \bar{v}]$$

She is charged if she receives the product and earns a bonus from the scoring rule, i.e.,

$$\boldsymbol{t}_i\left(\hat{v}, \hat{b}\right) = \lambda \cdot \mathrm{SR}\left(\hat{b}_i, \tilde{v}_i\right) - p \cdot \mathbf{1}(\hat{v}_i \geq p)$$

We argue that BDM-with-betting is simple compared to other mechanisms that elicit preferences by having agents bet on each others' reports (e.g., Crémer and McLean 1988, Miller et al. 2005). There are two ways in which it is simple.

1. *Estimation is straightforward.*

   The researcher's estimate is the average reported value, i.e.,

   $$f_n(m_1, \ldots, m_n) = \frac{1}{n}\sum_{i=1}^{n} \hat{v}_i$$

   In particular, the researcher does not try to estimate the population's average value from their reported beliefs.

2. *Incentives for truth-telling are straightforward.*

   The second stage (BDM) incentivizes agents to truthfully report their willingness to pay given any information they acquired. The first stage (betting) plays no role when it comes to incentivizing truthfulness. In particular, the researcher does not use bets made in the first stage to detect or disincentive lies in the second stage.

In contrast, other mechanisms that elicit preferences by having agents bet on each others' reports tend to use those bets in more complicated ways.[14]

[14]Full surplus extraction mechanisms use bets to make inferences about agents' willingness to pay (e.g., Crémer and McLean 1988). Peer prediction mechanisms have agents report signal realizations,

What is interesting about BDM-with-betting is not the mechanism itself. It is that this mechanism can incentivize agents to acquire precisely the kind of information needed to identify the population's average value. Next, we explain why.

4.2 Intuition for Theorem 2

The BDM-with-betting mechanism may seem naive. It does not incentivize agents to learn their values v_i, at least not in general. Instead, it incentivizes agents to learn about others' reported values $\hat{v}_{-i}$. Why is that enough to elicit the population's average value? Why is it not enough to elicit any other statistic?

We provide intuition using a simple example. Suppose the type space is a singleton and the state space is $\Omega = \{-1, 1\}$. For coefficient $\beta \in \mathbb{R}$, agents' values are

$$v_i = \beta\omega + \epsilon_i \quad \text{where} \quad \epsilon_i \sim N(0, 1) \text{ i.i.d.}$$

Finally, agents that acquire the revealing signal learn their values v_i and nothing else. There are two cases to consider.

Uncorrelated Values. Let values v_i be uncorrelated across agents (i.e., $\beta = 0$). Recall from Section 3.1 that this is the hard case for eliciting statistics other than the average, because it is generally impossible to incentivize agents to acquire the revealing signal. However, the average has a special property: it is not necessary for agents to acquire information when their values are uncorrelated. To see this, compare the researcher's estimate if the agents acquire information, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n} v_i = \frac{1}{n}\sum_{i=1}^{n} \epsilon_i \to_p 0$$

to the researcher's estimate if the agents do not acquire information, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[v_i] = 0$$

which are treated as bets on others' reported signal realizations (e.g., Miller et al. 2005). In both cases, the designer uses betting as a way to incentivize truth-telling. In our setting, incentivizing truth-telling is straightforward, and betting is a way to incentivize information acquisition.

Either way, the researcher's estimate is correct in the limit.

More generally, when values are uncorrelated, the researcher can take advantage of the *wisdom of the crowd*. It is not necessary that agents learn their values perfectly. What is important is that the errors $\hat{v}_i - v_i$ they make are uncorrelated across agents i. This property is what sets the average apart from other statistics.

Correlated Values. Let values v_i be correlated across agents (say, $\beta = 1$). If the betting stakes λ are sufficiently high, then agent i will acquire the revealing signal if it helps her predict the average reported value $\tilde{v}_i$. Suppose that agent i expects all other agents $j \neq i$ to acquire their revealing signals and learn their values v_j. Then she expects the average reported value $\tilde{v}_i$ to be

$$\tilde{v}_i = \frac{1}{n-1} \sum_{j \neq i} v_j \to_p \omega$$

Ideally, agent i would learn directly about the state ω, but that is not feasible in this example. But she can learn about her value v_i, which is correlated with the state ω. Therefore, betting incentivizes her to acquire the revealing signal.

There are two equilibria in this example, provided that the betting stakes λ are sufficiently large. In the informed equilibrium, all agents acquire the revealing signal and report their values $\hat{v}_i = v_i$. The researcher's estimate is consistent, i.e.,

$$\frac{1}{n} \sum_{i=1}^{n} \hat{v}_i = \frac{1}{n} \sum_{i=1}^{n} v_i = \omega$$

In the uninformed equilibrium, all agents acquire the uninformative signal and report their expected values $\hat{v}_i = 0$. Now, the researcher's estimate is not consistent, i.e.,

$$\frac{1}{n} \sum_{i=1}^{n} \hat{v}_i = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[v_i] = 0$$

There is no wisdom of the crowd in the uninformed equilibrium, since errors $\hat{v}_i - v_i$ are correlated across agents i.

This multiplicity does not always arise (e.g., Example 1 in Appendix B.2) but it is unavoidable in general (Proposition 2). However, if the informed minority condition holds, we can rule out the uninformed equilibrium. To see this, suppose that 20% of agents acquire the revealing signal at no cost. Then the remaining 80% of agents i want to acquire the revealing signal in order to better predict the vote share $\tilde{v}_i$, i.e.,

$$\tilde{v}_i = \frac{1}{n-1} \sum_{j \neq i} v_j \rightarrow_p \frac{1}{5} \cdot \omega + \frac{4}{5} \cdot 0 = \frac{\omega}{5}$$

Consistent with Proposition 1, the informed minority condition ensures that only the informed equilibrium survives.

Generalizing the Example. Naturally, this simple example bypasses the difficulties that a general proof of Theorem 2 must overcome.

In general, there is no clean separation between "uncorrelated values" and "correlated values". Whether values are correlated depends on what information agents begin with and the properties of the signals they acquire. This is especially true for non-binary signal spaces (see Appendix A.1), where agents may deviate from the revealing signal in many ways. For example, they may prefer to learn directly about the state ω, or only acquire partial information about their values v_i.

In particular, it is possible for values to be uncorrelated when agents acquire the revealing signal, but correlated when agents do not (e.g., Example 3 in Appendix B.2). This makes the best response map unstable – since each agent acquires information only if other agents do not – and implies that pure-strategy equilibria may not exist. Rather than attempting to construct mixed-strategy equilibria in strategy spaces of arbitrary size and dimension, we take a non-constructive approach to the proof.

Despite these difficulties, we find that BDM-with-betting can always sustain a sequence of equilibria where a wisdom-of-the-crowd effect holds in the limit. That is, betting incentivizes agents to acquire information until the point where errors $\hat{v}_i - v_i$ are (approximately) uncorrelated. We outline the proof in Appendix B.3.

At this point, we have fully characterized the combinations of statistics and information models where elicitation is possible. These results are largely negative, but not entirely. In particular, there is an important class of applications where it suffices to elicit the population's average value: social choice.

5 Application to Social Choice

We now turn to the problem of social choice, where a planner must choose a single alternative on behalf of a population. Specifically, we seek social choice mechanisms that work well even when processing information is costly.

It turns out that the methods we developed to elicit preferences can be used to design social choice mechanisms that are robust to costly information processing. This new setting raises new theoretical challenges and hints at new applications. Potential applications include:

1. *Democratic elections and referendums, especially at the local level.*

   Empirical studies suggest that many voters are poorly-informed (e.g., Delli Carpini and Keeter 1996) and may vote differently if they were better informed (e.g., Bartels 1996). If voters are not processing the information available about a candidate or ballot measure, they could easily make misguided decisions.

2. *Corporate governance.*

   Shareholders routinely elect board members and vote on issues that affect their interests (e.g., whether to approve a merger). Although large shareholders have strong financial incentives to stay well-informed, smaller shareholders and retail investors may not (e.g., Kastiel and Nili 2016, Brav et al. 2023).

3. *Collective decision-making in organizations.*

   For example, a committee may make decisions on behalf of the faculty at a university. Alternatively, a workforce may hold votes on whether to unionize, who to elect as a union official, etc. As in the previous applications, individuals may have little incentive to do their due diligence before voting.

Before presenting our result (Theorem 3), we describe a model of social choice (Section 5.1) and the challenges that good mechanisms must overcome (Section 5.2). Then we propose a mechanism (Section 5.3) and explain why it works (Section 5.4).

5.1 Model

We convert our model of preference elicitation into a model of social choice.

The researcher is now a planner. The n agents are now voters participating in a committee (or citizen's assembly, mini-public, jury, etc.) that take actions on behalf of a larger population. Alternatives $x \in \mathcal{X}$ indicate whether a measure passes ($x = 1$) or not ($x = 0$). The mechanism must assign the same alternative x to all agents.

Assumption 3. *Restrict attention to mechanisms* $(\boldsymbol{x}, \boldsymbol{t})$ *where, for all message profiles m,*

$$\boldsymbol{x}_1(m) = \boldsymbol{x}_2(m) = ... = \boldsymbol{x}_n(m)$$

In a slight abuse of notation, let $\mathbf{x}(m)$ *refer to the one alternative chosen for all agents.*

We maintain the assumption that the planner can recruit as many agents as he likes and spend as much money as he likes. This strikes us a reasonable starting point for large populations, like the residents of a large municipality or the employees of a major corporation. Since the population is large, even a large sample of agents will only be a small fraction of the population, and the cost of transfers to n agents will tend to be small relative to the potential gains for everyone else.

The planner seeks a sequence of mechanisms that *converges to efficiency*. We measure efficiency by comparing the expected *welfare* of a given mechanism with *first-best welfare*. Welfare is the population's average value from an alternative x, i.e.,

$$\mathrm{E}[x \cdot v_i \mid \omega]^{15}$$

In turn, expected welfare given mechanism $(\mathbf{x}, \mathbf{m})$ and strategy profile $(\mathbf{s}, \mathbf{m})$ is

$$W(\mathbf{s}, \mathbf{m}, \mathbf{x}, \mathbf{t}) = \mathrm{E}[\mathbf{x}(m) \cdot v_i \mid \omega]$$

with signals $s_i = \mathbf{s}_i(\theta_i, c_i)$ and messages $m_i = \mathbf{m}_i(s_i(\omega, \theta_i, v_i), \theta_i, c_i)$. Finally, first-best welfare is the highest expected welfare that the planner could achieve, if he knew the joint distribution F and the hidden state ω. Formally, it is

$$\mathrm{OPT} = \mathrm{E}_F\left[\max_x \mathrm{E}_F[x \cdot v_i \mid \omega]\right]$$

We ignore the cost of information acquisition in all of these definitions, but this is essentially without loss if the population is large. After all, only the committee

[15] Welfare naturally varies with the state ω. To see this, let state ω represent the quality of a proposed measure. Presumably, welfare from passing that measure is high if and only if the quality is high.

members will be asked to acquire information. Their information costs will tend to be small relative to the potential gains for the broader population.

Definition 13. *Fix a model specification* $(\Omega, \Theta, \mathcal{S}, G, F)$. *A mechanism* $(\boldsymbol{x}, \boldsymbol{t})$ *is* α-efficient *if expected welfare is within* α *of first-best for every equilibrium* $(\boldsymbol{s}, \boldsymbol{m})$, *i.e.,*

$$W(\boldsymbol{s}, \boldsymbol{m}, \boldsymbol{x}, \boldsymbol{t}) \geq \text{OPT} - \alpha \tag{3}$$

Let $\bar{\alpha}(\boldsymbol{x}, \boldsymbol{t})$ *be the smallest constant* α *that satisfies inequality* (3).

Definition 14. *Fix an information model* $(\Omega, \Theta, \mathcal{S}, G)$. *A sequence* $(\boldsymbol{x}^n, \boldsymbol{t}^n)$ *of mechanisms* converges to efficiency *if, for every joint distribution F,*

$$\lim_{n \to \infty} \bar{\alpha}(\boldsymbol{x}^n, \boldsymbol{t}^n) = 0$$

Finally, we make a strong distributional assumption that allows us to focus on more realistic mechanisms. We drop this assumption in Appendix A.2.

Definition 15. *A model specification* $(\Omega, \Theta, \mathcal{S}, G, F)$ *is* symmetric *if the population's median willingness to pay is always equal to the population's average willingness to pay. Here, agent i's willingness to pay* w_i *is her expected value* v_i *given her information, i.e.,*

$$w_i = \mathrm{E}_F[v_i \mid s_i(\omega, \theta_i, v_i), \theta_i]$$

Formally, symmetry is the requirement that, for any state ω *and signal profile s,*

$$\mathrm{med}_F(w_i \mid \omega) = \mathrm{E}_F[w_i \mid \omega]$$

For example, suppose that values, types, and the state are generated by a multivariate normal distribution. The signals that agents acquire may affect their beliefs about these variables in potentially-complicated ways, but their posteriors remain multivariate normal. Then symmetry would be satisfied.

Assumption 4. *Restrict attention to symmetric model specifications.*

The benefit of Assumption 4 is that we can build on the *majority-rule mechanism*, rather than more complex mechanisms like quadratic voting or VCG. The majority rule asks each agent i to report her preferred alternative $\hat{x}_i \in \{0, 1\}$, and then chooses the alternative x that the majority of agents report. Assumption 4 ensures that, if voters knew their values v_i, then majority rule would converge to efficiency.

We drop this assumption in Appendix A.2. Our efficiency results do generalize (see Theorem 4), but the mechanism changes. Specifically, we build on the VCG mechanism rather than the majority-rule mechanism. The downside to this is that VCG mechanism has several practical limitations (Rothkopf 2007).

Having finished describing the model, we can now turn to the planner's problem.

5.2 Challenges

When trying to solve the planner's problem, we encounter challenges that do not arise in models of social choice where agents know their own values.

To explain these challenges, we use the majority rule mechanism as a foil. Majority rule does not converge to efficiency when information acquisition is costly. This reflects three basic problems.

1. *Majority rule only rewards information acquisition in the event that an agent is* pivotal. *This is is unlikely (e.g., Mulligan and Hunter 2003).*

   Agent i is pivotal if changing her report $\hat{x}_i$ would change the alternative selected by the mechanism. If she is not pivotal, then information has no instrumental value. The probability that an agent is pivotal tends to vanish quickly as the sample size n grows, so there is little expected gain from acquiring information. This is typically true even if n is relatively small.

2. *Majority rule does not account for the positive externalities of information acquisition, where one agent's informed decision can benefit everyone else.*

   For example, suppose that agents have common values and that agent i knows she will be pivotal. Acquiring information that identifies the optimal alternative would not only improve her well-being, but also the well-being of everyone else. Since agent i does not internalize this benefit to others, she may acquire less information than the planner would like.

3. *When agents are imperfectly informed, majority rule does not incentivize them to truthfully report the alternative they prefer given the information they have. This can distort the outcome of the vote (e.g., Austen-Smith and Banks 1996).*

> Since agent i's report only matters if she is pivotal, it is optimal for her to condition her beliefs on being pivotal before making a report. For example, suppose that agent i believes a ballot measure is likely to be both good for her and very popular among the other agents. If she is truthful, she would vote in favor of the measure. However, if she conditions on her being pivotal, then she focuses on a contingency where the measure was not as popular as expected. This may lead her to revise her beliefs about her own value v_i downwards.

We stress that these problems are not unique to majority rule. They apply to many other mechanisms (e.g., quadratic voting, VCG mechanism, etc.) that do not account for voters' costs of processing information.

Of these challenges, the third one is the most daunting. The first two deal with incentives to acquire information, which we already studied in the preference elicitation context. The third challenge deals with incentives to be truthful when voters have not acquired all available information. We know from Theorem 1 that it is not possible to incentivize voters to acquire all available information. Moreover, the way that we incentivized truthfulness in Theorem 2 relied on the ability to allocate different alternatives to different agents. This is not possible in social choice.

Fortunately, it turns out that the kind of information that voters need to acquire in order to be truthful is closely related to the kind of information that is relevant to eliciting the population's average value. We know how to incentivize agents to acquire the latter. Essentially the same mechanism incentivizes them to acquire the former. Next, we describe this mechanism and when it converges to efficiency.

5.3 Mechanism

We introduce the *majority-rule-with-betting mechanism*. Majority-rule-with-betting has voters cast their votes according to the majority rule and asks them, before voting, to predict the results. Theorem 3 asks when it converges to efficiency.

To be more precise, majority-rule-with-betting has voters cast their votes according to the majority rule with high probability. For technical reasons, there is also an ϵ probability that a randomly-chosen agent dictates the outcome.[16]

[16] This ensures that voters are pivotal with positive probability and therefore avoids situations where voters must condition on zero probability events when casting their votes.

To define this mechanism, we need to update our notation from Section 4.1. Let the *vote share* from agent i's perspective be

$$\tilde{n}_i = \frac{1}{n} \sum_{j \neq i} \hat{x}_j$$

Let agent i's belief over the vote share be

$$b_i \in \mathcal{B} = \Delta([0, 1])$$

A scoring rule maps a reported belief $\hat{b}_i \in \mathcal{B}$ and a vote share $\tilde{n}_i \in \mathcal{V}^m$ to a score, i.e.,

$$\mathrm{SR} : \mathcal{B} \times [0, 1] \to \mathbb{R}$$

Definition 16. *The* majority-rule-with-betting mechanism $(\boldsymbol{x}, \boldsymbol{t})$ *is parameterized by a proper scoring rule* SR*, scaling parameter* λ*, and probability* ϵ*. Agent i sends a message*

$$m_i = \left(\hat{x}_i, \hat{b}_i\right) \in \{0, 1\} \times \mathcal{B}$$

that consists of a reported alternative $\hat{x}_i$ *and a reported belief* $\hat{b}_i$*. With probability* $1 - \epsilon$*, the planner selects the alternative that the majority reports, i.e.,*

$$x(\hat{x}, \hat{b}) = \mathbf{1}\left(\frac{1}{n} \sum_{i=1}^{n} \hat{x}_i \geq \frac{1}{2}\right)$$

With probability ϵ*, the planner randomly chooses a voter* $i \sim \textsc{Uniform}(1, \ldots, n)$ *and selects their reported alternative, i.e.,*

$$x(\hat{x}, \hat{b}) = \hat{x}_i$$

Finally, each agent i is paid for her prediction according to the scoring rule, i.e.,

$$\boldsymbol{t}_i\left(\hat{v}, \hat{b}\right) = \lambda \cdot \mathrm{SR}\left(\hat{b}_i, \tilde{n}_i\right)$$

Although we make no claims about the practical viability of this mechanism, it is still worth emphasizing that the mechanism avoids two ethical red lines. First, it does not allow for vote buying, since transfers are used only to reward accurate predictions.

Second, it counts all votes equally regardless of voters' predictions.

Majority-rule-with-betting may not appear that different from the status quo in many democracies, where voters are free to participate in political betting markets. But there are differences. First, we do not ask voters to bet against each other. Doing so would give rise to the Grossman and Stiglitz (1980) paradox and undermine the typical voter's incentive to acquire information. Second, we rely on *sortition* – that is, random sampling of the population. Voters are members of a citizen's assembly (or a committee, etc.) tasked with making a decision on behalf of the population.[17]

Having described the mechanism, we can ask when it converges to efficiency

Theorem 3. *If the informed minority condition holds, there is a sequence of majority-rule-with-betting mechanisms that converge to efficiency.*

Remark 3. As in Theorem 2, (i) the informed minority condition is not necessary if the planner is willing to assume best-case equilibrium selection and (ii) the planner does not need to know the partial model specification (see Remark 2).

In practice, the informed minority may consist of people who are intrinsically interested in politics or motivated by a sense of civic duty. Empirically, "a substantial percentage [of Americans] is very informed, an equally large percentage is very poorly informed, and the plurality of citizens fall somewhere in between" (Delli Carpini 2000). This is supported by a number of studies that find substantial heterogeneity in voter informedness (e.g., Palfrey and Poole 1987; Delli Carpini and Keeter 1996; Pew Research Center 2007; Prior and Lupia 2008; Angelucci and Prat 2024).

We conclude our excursion into social choice with a high-level intuition for Theorem 3. In particular, we explain why majority-rule-with-betting resolves the three challenges we outlined in Section 5.2. Readers interested in understanding this result more deeply should refer to the proof outline in Appendix B.4.

5.4 Intuition for Theorem 3

To understand why majority-rule-with-betting converges to efficiency, recall the three challenges in Section 5.2. The first two deal with incentives to acquire information, while the third challenge deals with incentives to be truthful.

[17] Sortition has been used in various forms by democracies dating back to ancient Athens. In the United States, for example, sortition is used to form jury pools.

Incentives to Acquire Information. Majority-rule-with-betting addresses the first two challenges in the same way that BDM-with-betting elicits the average.

Assuming voters are truthful, betting sustains equilibria where a "wisdom of the crowd" effect applies. Any uncertainty voter i has about her value v_i is uncorrelated with the uncertainty other voters j have about their value v_j. The fact that voters make errors in how they vote does not affect the average as long as those errors are not correlated. This ensures that efficient equilibria exist where individual voting errors cancel out; the informed minority condition rules out other equilibria.

Of course, symmetry (Assumption 4) plays an important role here. It ensures that majority rule will choose the alternative that maximizes the population's average willingness to pay. In turn, the previous paragraph ensures that the population's average willingness to pay matches the population's average value.

The biggest challenge in translating the proof of Theorem 2 is that what voters are betting on now (the vote share) is much coarser than what agents were betting on previously (the average reported value). This means that voters are no longer incentivized to acquire certain kinds of information. Fortunately, we show that these kinds of information are not essential, and that efficient equilibria still exist.

Incentives to be Truthful. The third challenge is the most daunting. On the one hand, we cannot incentivize voters to acquire all available information (Theorem 1). On the other hand, imperfectly-informed voters may not be truthful.[18]

Fortunately, it turns out that the kind of information that voters need to acquire in order to be truthful is closely related to the kind of information that is relevant to eliciting the population's average value. This means that a convenient side effect of betting on the vote share is that it restores truth-telling incentives.

To see this, recall why majority rule may not incentivize truth-telling on its own. Agent i knows that her vote $\hat{x}_i$ is pivotal only if the vote share $\tilde{v}_i$ is exactly 50%. Therefore, it is in her best interest to vote in favor of the measure whenever her expected value v_i conditional on being pivotal is positive, i.e.,

$$\mathrm{E}[v_i \mid s_i(\omega, \theta_i, v_i), \theta_i, \tilde{v}_i = 50\%] \geq 0 \tag{4}$$

[18] This is true even if voters are very close to informed, as long as the chance of being pivotal is small.

A truthful agent would not condition on pivotality and vote in favor whenever

$$\mathrm{E}[v_i \mid s_i(\omega, \theta_i, v_i), \theta_i] \geq 0 \tag{5}$$

What determines whether agent i will be truthful? It is whether agent i perceives the vote share $\tilde{v}_i$ to be correlated with her value v_i even after conditioning on her information. If not, then conditioning on pivotality does not affect agent i's beliefs about her value. As a result, conditions (4) and (5) will be the same.

The reason why high-stakes betting incentivizes agent i to acquire information until she is truthful is because it incentivizes her to acquire any information that is correlated with the vote share $\tilde{v}_i$. On the one hand, if v_i is correlated with the vote share $\tilde{v}_i$, then agent i benefits from acquiring her revealing signal. Once she acquires the revealing signal, she prefers to vote truthfully. On the other hand, if v_i is not correlated with the vote share $\tilde{v}_i$, then agent i also prefers to vote truthfully.

Formalizing this argument requires some care, for two reasons. First, since the probability that agents are pivotal may be quite small, their beliefs may be extremely sensitive to the vote share $\tilde{v}_i$. One implication of this is that we can only incentivize agents to be nearly truthful with high probability. Second, when agent i has access to many signals (see Appendix A.1), she may not respond to correlation between her value v_i and the vote share $\tilde{v}_i$ by acquiring the revealing signal. Instead, she may acquire another signal that still leaves her uncertain about her value v_i.

We leave a detailed proof outline to Appendix B.4. Aside from the generalized model and results in Appendix A, we have now presented our main findings.

6 Related Literature

7 Conclusion

Revealed preference plays a foundational role in economics, by linking observed choices with unobserved preferences. This link is disrupted when people find it costly to process information about the goods they consume, the services they receive, and the policies that affect them. We ask whether it is possible to repair that link. That is, we ask whether observed choices, in any possible mechanism and from arbitrarily many individuals, can identify the distribution of preferences in a population.

It is not possible to fully restore the link between choices and preferences when processing information is costly (Theorem 1). It is possible to partially restore the link (Theorem 2), and that is enough for some applications (Theorems 3-4). For other applications, it may be possible to fully restore the link under stronger distributional assumptions and sensible equilibrium refinements. It may also be possible to obtain informative bounds on statistics that are relevant to those applications. But it is clear that choices cannot reveal preferences at the same level of generality and precision as classical theory suggests (e.g., Becker et al. 1964; Clarke 1971).

Readers may object to our emphasis on informed preferences (which are difficult to elicit) rather than uninformed preferences (which are easier to elicit). After all, as long as people are uninformed, it is their uninformed preferences that govern their behavior. But revealed preference is not only used to predict behavior, it is also used to measure welfare. When structural models are used to guide policy decisions, or theoretical models are used to motivate particular markets, it would be a mistake to equate welfare with people's uninformed preferences. Even when our goal is to predict behavior, informed preferences matter for many applications.[19]

Naturally, our results have weaknesses that may limit their practical relevance. For example, the researcher in our model can recruit an unlimited number of individuals and spend an unlimited amount of money. This strengthens our negative results, but weakens our positive results. In addition, we rely on Bayes-Nash equilibrium, a solution concept that is quite strong and often leads to unrealistic conclusions. Finally, our results are entirely theoretical. We do not provide empirical evidence that mechanisms like BDM-with-betting work in practice.

Nonetheless, we believe that further work that addresses these weaknesses could have broad practical relevance. In particular, there are many institutions where costly information processing can lead to market failures (e.g., elections), and many institutions that exist in part to correct such market failures (e.g., reputation and recommender systems). Understanding and addressing these market failures may ultimately require methods for eliciting informed preferences.

[19] For example, consider a firm estimating customer lifetime value. If repeat customers become better informed about a product over time, their lifetime demand depends on their informed preferences.

References

Angelucci, C., & Prat, A. (2023). *Is journalistic truth dead? measuring how informed voters are about political news*.

Angelucci, C., & Prat, A. (2024). Is journalistic truth dead? measuring how informed voters are about political news. *American Economic Review*, *114*(4), 887–925.

Arrow, K. J. (1950). A difficulty in the concept of social welfare. *Journal of Political Economy*, *58*(4), 328–346.

Austen-Smith, D., & Banks, J. S. (1996). Information aggregation, rationality, and the condorcet jury theorem. *The American Political Science Review*, *90*(1), 34–45.

Bartels, L. M. (1996). Uninformed votes: Information effects in presidential elections. *American Journal of Political Science*, *40*(1), 194–230.

Becker, G. M., Degroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, *9*(3), 226–232.

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

Brav, A., Malenko, A., & Malenko, N. (2023). Corporate governance implications of the growth in indexing.

Carroll, G. (2015). Robustness and linear contracts. *American Economic Review*, *105*(2), 536–63.

Carroll, G. (2019). Robust incentives for information acquisition. *Journal of Economic Theory*, *181*, 382–420.

Clarke, E. H. (1971). Multipart pricing of public goods. *Public Choice*, *11*, 17–33.

Crémer, J., & McLean, R. P. (1985). Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent. *Econometrica*, *53*(2), 345–361.

Crémer, J., & McLean, R. P. (1988). Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions. *Econometrica*, *56*(6), 1247–1257.

Delli Carpini, M. X. (2000). In search of the informed citizen: What americans know about politics and why it matters. *The Communication Review*, *4*(1), 129–164.

Delli Carpini, M. X., & Keeter, S. (1996). *What americans know about politics and why it matters*. Yale University Press.

Denti, T., & Ravid, D. (2023). *Robust predictions in games with rational inattention*.

Feddersen, T., & Sandroni, A. (2006). Ethical voters and costly information acquisition. *Quarterly Journal of Political Science*, *1*(3), 287–311.

Fowler, A., & Margolis, M. (2014). The political consequences of uninformed voters. *Electoral Studies*, *34*, 100–110.

Gerardi, D., & Yariv, L. (2008). Information acquisition in committees. *Games and Economic Behavior*, *62*(2), 436–459.

Gershkov, A., & Szentes, B. (2009). Optimal voting schemes with costly information acquisition. *Journal of Economic Theory*, *144*(1), 36–68.

Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American Economic Review*, *70*(3), 393–408.

Groves, T., & Ledyard, J. (1977). Optimal allocation of public goods: A solution to the "free rider" problem. *Econometrica*, *45*(4), 783–809.

Jensen, J. L. W. V. (1905). Om konvekse funktioner og uligheder imellem middelvaerdier. *Nyt tidsskrift for matematik*, *16*, 49–68.

Kastiel, K., & Nili, Y. (2016). In search of the absent shareholders: A new solution to retail investors' apathy. *Delaware Journal of Corporate Law*, *41*, 55.

Lau, R. R., & Redlawsk, D. P. (1997). Voting correctly. *The American Political Science Review*, *91*(3), 585–598.

McCarthy, J. (1956). Measures of the value of information. *Proceedings of the National Academy of Sciences*, *42*(9), 654–655.

Miller, N., Resnick, P., & Zeckhauser, R. (2005). Eliciting informative feedback: The peer-prediction method. *Management Science*, *51*(9), 1359–1373.

Mulligan, C. B., & Hunter, C. G. (2003). The empirical frequency of a pivotal vote. *Public Choice*, *116*(1/2), 31–54.

Pakzad-Hurson, B. (2022). Crowdsourcing and optimal market design.

Palfrey, T. R., & Poole, K. T. (1987). The relationship between information, ideology, and voting behavior. *American Journal of Political Science*, *31*(3), 511–530.

Persico, N. (2000). Information acquisition in auctions. *Econometrica*, *68*(1), 135–148.

Persico, N. (2004). Committee Design with Endogenous Information. *The Review of Economic Studies*, *71*(1), 165–191.

Pew Research Center. (2007). *Public knowledge of current affairs little changed by news and information revolution* (tech. rep.). Washington, D.C.

Prelec, D. (2004). A bayesian truth serum for subjective data. *Science*, *306*(5695), 462–466.

Prior, M., & Lupia, A. (2008). Money, time, and political knowledge: Distinguishing quick recall and political learning skills. *American Journal of Political Science*, *52*(1), 169–183.

Rothkopf, M. H. (2007). Thirteen reasons why the vickrey-clarke-groves process is not practical. *Oper. Res.*, *55*(2), 191–197.

A Generalized Results

We generalize the model in three respects. First, we allow for more complex forms of information acquisition by relaxing Assumption 2. Second, we show that the symmetry condition is not necessary for efficient social choice by dropping Assumption 4. Third, we allow for multiple alternatives, i.e.,

$$\mathcal{X} = \{0, \dots, m\}$$

where $v_{ix} \in [0, \bar{v}]$ is agent i's value from alternative x. We no longer require $v_{0x} = 0$.

A.1 Relaxing Assumption 2

Assumption 2 restricted attention to information models where each type of agent had access to two signals: a revealing signal and an uninformative signal. We drop this assumption and replace it with two assumptions that are jointly much weaker.

First, we assume that agents can combine signals. If an agent acquires one signal, that does not preclude them from acquiring another one. It also does not increase the cost of acquiring that other signal. This ensures that incentivizing the agents to learn about one aspect of their preferences does not immediately prevent them from learning anything else.

Assumption 5. *Restrict attention to information models $(\Omega, \Theta, \mathcal{S}, G)$ with the following structure. The set of signals $\mathcal{S}$ consists of base signals and combined signals. For every set of base signals $\left\{s_i^1, \dots, s_i^\ell\right\} \subseteq \mathcal{S}$, there is a combined signal $s_i \in \mathcal{S}$ where*

$$\forall \omega \in \Omega, \theta_i \in \Theta, \quad s_i\left(\omega, \theta_i\right) = \left(s_i^1(\omega, \theta_i), \dots, s_i^\ell(\omega, \theta_i)\right) \tag{6}$$

The combined signal costs no more than the combined cost of its components, i.e.,

$$\forall \theta_i \in \Theta, \quad c_i\left(s_i, \theta_i\right) \leq \sum_{j=1}^{\ell} c_i\left(s_i^j, \theta_i\right) \tag{7}$$

Second, we assume that each agent i can learn her own values $v_i = (v_{i1}, \dots, v_{im})$ at a finite cost. This corresponds to a revealing signal, as in Definition 7. The only difference is that condition (2) needs to be generalized to

$$\forall x \in \mathcal{X}, \quad \mathrm{Var}_F[v_{ix} \mid s_i(\omega, \theta_i, v_i), \theta_i] = 0$$

since we now allow for more than two alternatives.

Assumption 6. *Restrict attention to information models* $(\Omega, \Theta, \mathcal{S}, G)$ *where each type* θ_i *has access to a revealing signal* $s_i^{\theta_i} \in \mathcal{S}$ *that has finite cost, i.e.,*

$$c_i\left(s_i^{\theta_i}, \theta_i\right) < 0 \tag{8}$$

The fact that agents have no remaining uncertainty about their preferences is not essential. What is essential is this: if another agent j can learn something about agent i's values v_i, then agent i can learn it as well.[20]

In the context of social choice, Assumption 6 seems especially suitable when agents acquire information by processing a common pool of existing knowledge. Sources of existing knowledge might include newspaper articles and editorials, speeches and press releases, and expert analyses that help agents understand the implications of some policy or candidate for their own well-being. This assumption may be less suitable in settings where a group is expected to generate new knowledge or where some individuals have access to inside information.[21]

All of our positive results continue to hold verbatim when the model is generalized in this way. For clarity, we restate them now.

Theorem 2 (Generalized)**.** *Make Assumptions 1, 5, and 6. Let the σ-statistic be an average. Then it is weakly elicitable.*

Proposition 1 (Generalized)**.** *Make Assumptions 1, 5, and 6. Let the σ-statistic be an average. Then it is elicitable if the informed minority condition holds.*

Theorem 3 (Generalized)**.** *Make Assumptions 1, 3, 4, 5, and 6. If the informed minority condition holds, there is a sequence of majority-rule-with-betting mechanisms that converge to efficiency.*

In Appendices B and C, we only refer to the generalized versions of these results.

[20] If this were not the case, the incentive problem would be quite different. We would need to encourage the agents to reveal information to each other, rather than to encourage them to learn for themselves.

[21] There are other institutions that seem better suited for acquiring and disseminating information of this sort, like academic research, opposition research, or investigative journalism.

A.2 Removing Assumption 4

Assumption 4 restricts attention to model specifications that satisfied the symmetry condition. Theorem 4 shows that this assumption is not entirely necessary.

Theorem 4. *Make Assumptions 1, 3, 5, and 6. If the informed minority condition holds, there is a sequence of mechanisms that converge to efficiency.*

This generalizes Theorem 3 by dropping Assumption 4 and allowing for any number $m \geq 2$ of alternatives (although the rate of convergence slows quickly as m grows). The cost of generalization is that we cannot use majority-rule-with-betting. There are two reasons for this. First, without the symmetry condition, majority rule would not be efficient even if agents knew their values in advance. Second, majority rule is only well-defined when there are $m = 2$ alternatives.

To prove Theorem 4, we introduce the *VCG-with-betting mechanism*. This replaces the majority rule with the VCG mechanism, and has agents bet on the average reported value rather than the vote share. For additional details on the mechanism and the proof of Theorem 4, see the Supplemental Appendix.

B Proof Outlines

We outline the proofs of our main results. We encourage the reader to review the generalized model in Appendix A before proceeding.

B.1 Proof of Theorem 1

We construct two joint distributions F and F' that the planner must be able to distinguish if the σ-statistic is identified. There are four steps to this construction.

1. First, fix a σ-statistic that is elicitable, and let $(\mathbf{x}^n, \mathbf{t}^n)$ be a sequence of mechanisms that identifies it. Suppose for contradiction that the σ-statistic is not an average. That is, σ is non-affine. For every $\epsilon > 0$, there must exist $a, b \in \mathcal{V}$ where
$$\sigma\left(\frac{a+b}{2}\right) \neq \frac{\sigma(a)+\sigma(b)}{2} \tag{9}$$

and $a \leq b \leq a + \epsilon$.[22]

2. Second, fix a non-trivial information model $(\Omega, \Theta, \mathcal{S}, G)$. Let $\tilde{c} > 0$ be the smallest cost where there exists a type θ_i – say, $\theta_i = 1$ – and probability $p > 0$ such that
$$\Pr\left[c_i\left(s_i^{\theta_i}\right) \geq \tilde{c}\right] = p$$
This exists by non-triviality and by Assumption 1, which ensures that the cost distribution G has finite support. Since $|\Theta| \geq 3$ by Assumption 1, label two other types so that $\{1, 2, 3\} \subseteq \Theta$. Similarly, choose a state $\omega \in \Omega$ and label it $\omega = 1$.

3. Third, construct a joint distribution F as follows. With probability one, the state is $\omega = 1$ and the type is $\theta_i = 1$. Let values be drawn according to
$$v_i \sim a + (b - a) \cdot \textsc{Bernoulli}(0.5)$$
Note that the expected value v_i and σ-statistic are
$$\mathrm{E}_F[v_i] = \frac{a + b}{2} \quad \text{and} \quad \mathrm{E}_F[\sigma(v_i)] = \frac{\sigma(a) + \sigma(b)}{2}$$

4. Fourth, construct another joint distribution F' as follows. With probability one, the state is $\omega = 1$. Let types be drawn according to
$$\Pr_{F'}[\theta_i = 1] = p \quad \text{and} \quad \Pr[\theta_i = 2] = \Pr[\theta_i = 3] = \frac{1 - p}{2}$$

[22] To prove this, there are two cases. First, suppose that this condition fails and σ is continuous. Consider a small interval $[c, d] \subseteq \mathcal{V}$, where "small" means that $d - c \leq \epsilon$. By definition, since we cannot find constants $a, b \in [c, d]$ such that inequality (9) holds, σ is both midpoint convex and midpoint concave on the interval $[c, d]$. By Jensen (1905), it follows that σ is both convex and concave – in other words, affine – on the interval $[c, d]$. If σ is affine on every small subinterval $[c, d]$, it must (i) be differentiable and (ii) have a constant derivative on the entire domain $\mathcal{V}$. Therefore, it is affine.

Second, suppose that σ is discontinuous at $d \in \mathcal{V}$. By Assumption 1, σ has a finite discontinuity set. So, there are intervals $[d', d)$ and $(d, d'']$ over which σ is continuous. Let $a \in [d', d]$ and $b \in [d, d'']$. As $d', d'' \to d$, either (i) $\lim \sigma(d') \neq \sigma(d)$, (ii) $\lim \sigma(d'') \neq \sigma(d)$, or both. If (i), let $a = d'$ and $b = d$. If (ii), let $b = d''$ and $a = d$. As $d', d'' \to d$, eventually $b - a \leq \epsilon$ and inequality 9 holds.

Let values be drawn according to

$$v_i = \begin{cases} \frac{a+b}{2} & \theta_i = 1 \\ a & \theta_i = 2 \\ b & \theta_i = 3 \end{cases}$$

Note that the expected value v_i and σ-statistic are

$$\mathrm{E}_{F'}[v_i] = \frac{a+b}{2} \quad \text{and} \quad \mathrm{E}_{F'}[\sigma(v_i)] = (1-p) \cdot \frac{\sigma(a)+\sigma(b)}{2} + p \cdot \sigma\left(\frac{a+b}{2}\right)$$

Critically, the σ-statistic takes on a different value when the joint distribution is F relative to when it is F'. It follows that a planner who can identify the σ-statistic can also distinguish between joint distributions F and F'. Accordingly, our goal is to show that the planner cannot distinguish between joint distributions F and F'.

Lemma 1. *Set $\epsilon > 0$ sufficiently small. For any mechanism $(\boldsymbol{x}, \boldsymbol{t})$, there exists a pair of strategy profiles $(\boldsymbol{s}, \boldsymbol{m})$ and $(\boldsymbol{s}', \boldsymbol{m}')$ where:*

1. *Given model specification $(\Omega, \Theta, \mathcal{S}, G, F)$, $(\boldsymbol{s}, \boldsymbol{m})$ is an equilibrium.*

2. *Given model specification $(\Omega, \Theta, \mathcal{S}, G, F')$, $(\boldsymbol{s}', \boldsymbol{m}')$ is an equilibrium.*

3. *Consider message profiles m and m', defined as*

$$\begin{aligned} m_i &= \boldsymbol{m}_i\left(s_i(\omega, \theta_i, v_i), \theta_i, c_i\right) \quad \textit{where} \quad s_i = \boldsymbol{s}_i\left(\theta_i, c_i\right) \\ m_i' &= \boldsymbol{m}_i'\left(s_i'(\omega, \theta_i, v_i), \theta_i, c_i\right) \quad \textit{where} \quad s_i' = \boldsymbol{s}_i'\left(\theta_i, c_i\right) \end{aligned}$$

The distribution of m given joint distribution F is the same as the distribution of m' given joint distribution F'.

Lemma 1 ensures that, with suitable equilibrium selection, the data $(m_1, \ldots, m_n)$ has the same distribution regardless of whether the joint distribution is F or F'. In turn, the estimate $f_n(m_1, \ldots, m_n)$ in Definition 5 is identically-distributed in both cases. This leads to a contradiction, as follows. Identification requires that

$$f_n(m_1, \ldots, m_n) \to_p \mathrm{E}_F[\sigma(v_i) \mid \omega] \quad \text{and} \quad f_n(m_1, \ldots, m_n) \to_p \mathrm{E}_{F'}[\sigma(v_i) \mid \omega] \tag{10}$$

However, by construction, the σ-statistic takes on a different value for distributions F and F'. That is,

$$\mathrm{E}_F[\sigma(v_i) \mid \omega] \neq \mathrm{E}_{F'}[\sigma(v_i) \mid \omega] \tag{11}$$

Together, conditions (10) and (11) imply that the estimate $f_n(m_1, \dots, m_n)$ converges to two different quantities. This is a contradiction.

B.2 Special Cases of Theorem 2

We work through Theorem 2 in three examples, before outlining the general proof in Section B.3. These examples are intended to build intuition for how and when BDM-with-betting elicits (or weakly elicits) the average.

In these examples, we maintain three conventions. First, as in Section 2, we assume there are two alternatives: an outside option with value zero and a product with value v_i to agent i. Second, we temporarily drop the requirement that values v_i must lie within a bounded interval $[0, \bar{v}]$. Third, we always focus on the limit as both the sample size and the betting stakes grow (i.e., $n \to \infty$ and $\lambda_n \to \infty$).

B.2.1 Elicitation without an Informed Minority

The first example illustrates that there are natural model specifications, which violate the informed minority condition, where BDM-with-betting has a unique equilibrium. This stands in contrast to the example in Section 4.2, which had multiple equilibria.

Example 1. *Fix a model specification where the state represents product quality, i.e.,*

$$\omega \sim \textsc{Bernoulli}\,(0.5)$$

Each agent i's value v_i is positively correlated with the state, i.e.,

$$v_i \sim N(\omega, 1)$$

Each agent i has a binary type that represents a noisy assessment of the state, i.e.,

$$\theta_i \sim \textsc{Bernoulli}\left(\frac{1+\omega}{3}\right)$$

Each agent i has access to a revealing signal s_i^R that costs $\bar{c}$, where

$$s^R(\omega, \theta_i, v_i) = v_i$$

and an uninformative signal s_i^U that costs nothing, where

$$s_i^U(\omega, \theta_i, v_i) = 0$$

The unique equilibrium that survives asymptotically sets

$$\forall i = 1, \dots, n: \quad \mathbf{s}_i(\theta_i, c_i) = s_i \quad \text{and} \quad \mathbf{m}_i(s_i(\omega, \theta_i, v_i), \theta_i, c_i) = v_i$$

That is, all agents i acquire the revealing signal and report their values v_i.

This equilibrium is unique because acquiring the revealing signal always helps agents predict the average reported value. To see this, consider two extreme cases. First, if agents $j \neq i$ acquire the informative signal, the average reported value is

$$\tilde{v}_i = \frac{1}{n-1} \sum_{j \neq i} \hat{v}_j = \frac{1}{n-1} \sum_{j \neq i} v_j \to_p \omega$$

Second, if agents $j \neq i$ acquire the uninformative signal, the average reported value is

$$\tilde{v}_i = \frac{1}{n-1} \sum_{j \neq i} \hat{v}_j = \frac{1}{n-1} \sum_{j \neq i} \mathrm{E}\left[v_j \mid \theta_j\right] \to_p \frac{4+\omega}{9}$$

In both cases, the more that agent i learns about the state ω, the more accurately she is able to predict the average reported value $\tilde{v}_i$. Moreover, she can learn more about the state ω by acquiring her revealing signal; this reveals her value v_i, which is correlated with ω. As long as the scaling parameter λ_n is sufficiently large compared to the cost $\bar{c}$, agent i is better off acquiring the revealing signal.

In this example, the average reported value identifies the population's average value because every agent i learns her value v_i. Next, we consider an example where the same holds even though no agent learns her value v_i perfectly.

B.2.2 Elicitation with Partial Information Acquisition

The second example illustrates why BDM-with-betting remains effective if agents have access to multiple signals. In particular, we consider what happens if agents react to betting incentives by learning about others' values rather than their own.

Example 2. *The model specification is the same as in Example 1, except that agents have access to two additional signals. First, there is a product quality signal s_i^Q that costs $\bar{c}/2$ and reveals the state ω, i.e.,*

$$s_i^Q(\omega, \theta_i, v_i) = \omega$$

Second, to satisfy Assumption 5, there is a combined signal s_i^C that costs $3\bar{c}/2$, i.e.,

$$s_i^C(\omega, \theta_i, v_i) = (\omega, v_i)$$

The unique equilibrium that survives asymptotically sets

$$\forall i = 1, \dots, n: \quad \mathbf{s}_i(\theta_i) = s_i^Q \quad \text{and} \quad \mathbf{m}_i(s_i(\omega, \theta_i, v_i), \theta_i) = \omega$$

Here, agents i do not learn their own values v_i. They prefer the quality signal s_i^Q to the revealing signal s_i^R and the combined signal s_i^C. All three signals are equally good at predicting the vote margin, but the quality signal costs the least. Nonetheless, the average reported value equals the population's average value, since

$$\frac{1}{n}\sum_{j=1}^{n} \hat{v}_j = \frac{1}{n}\sum_{j=1}^{n} \mathrm{E}\left[v_j \mid \theta_j, \omega\right] = \omega = \mathrm{E}[v_i \mid \omega]$$

Examples 1 and 2 both have a symmetric pure-strategy equilibrium. Next, we consider an example where that is not the case.

B.2.3 Elicitation with Mixed-Strategy Equilibria

The third example illustrates how unstable best response dynamics can lead to mixed-strategy equilibria. The best response function is unstable in the sense that, as other agents j acquire more information, agent i wants to acquire *less* information. This is contrary to the intuition from the example in Section 4.2, where the incentive to acquire information is strongest when other agents also acquire information.

Example 3. *The model specification is the same as in Example 1, with two changes. First, agent i's value v_i is no longer correlated with the state, i.e.,*

$$v_{i1} \sim \textsc{Uniform}[-2, 2+2\gamma]$$

where $\gamma \in \mathbb{R}$ is a constant. Second, agent i's type θ_i provides information about both her value v_i and the state ω, i.e.,

$$\theta_i = \mathbf{1}(v_i > 1 - 2\omega)$$

Here, the state ω no longer represents quality. Instead, it controls whether the agents' private signals are biased in favor of the product ($\omega = 1$) or not ($\omega = 0$). Each agent i either receives "good news" ($\theta_i = 1$), or "bad news" ($\theta_i = 0$).

In any symmetric pure strategy profile, each agent i prefers to become informed if and only if other agents are *not* informed. To see this, consider two cases.

1. If agents $j \neq i$ acquire the revealing signal s_j^R, the average reported value is

$$\tilde{v}_i = \frac{1}{n-1}\sum_{j\neq i}\hat{v}_j = \frac{1}{n-1}\sum_{j\neq i} v_j \to_p \gamma$$

   Since the parameter γ is common knowledge, agent i does not need to acquire the revealing signal s_i^R in order to predict the vote margin. She prefers to deviate to the uninformative signal s_i^U, which costs less.

2. If agents $j \neq i$ acquire the uninformative signal s_j^U, the average reported value is

$$\tilde{v}_i = \frac{1}{n-1}\sum_{j\neq i}\hat{v}_j = \frac{1}{n-1}\sum_{j\neq i} \mathrm{E}\left[v_j \mid \theta_j\right] \to_p \begin{cases} \frac{\gamma^2+2\gamma-2}{\gamma+4} & \omega = 0 \\ \frac{\gamma^2+6\gamma+2}{\gamma+4} & \omega = 1 \end{cases}$$

   When λ_n is sufficiently large, agent i prefers to acquire the revealing signal, which is correlated with ω and therefore helps her predict the vote margin.

However, there is an equilibrium in mixed strategies. In that equilibrium, agents acquire the informative signal with high probability and the uninformative signal with low probability. Intuitively, the minority of uninformed agents is large enough to affect the average reported value (but just barely). The remaining agents are willing to acquire information to predict what this uninformed minority will report (but again, just barely). The size of the uninformed minority needed to incentivize information acquisition is decreasing as the betting stakes grow ($\lambda_n \to \infty$). Asymptotically, the uninformed minority vanishes as a fraction of the sample size, and the average reported value converges to the population's average value.

In this example, characterizing the mixed-strategy equilibrium is straightforward because there are only two signals, and they are common to all agents. When many signals are available, strategies are complicated and high-dimensional, which makes constructing mixed-strategy equilibria quite difficult. We overcome this difficulty with a non-constructive approach to proving Theorem 2.

B.3 Proof of Theorem 2

We outline the proof of Theorem 2. After describing the sequence of mechanisms in more detail, we introduce key notation and sufficient conditions for eliciting the average. Then we verify those sufficient conditions in Lemmas 2-4.

Let each mechanism $(\mathbf{x}^n, \mathbf{t}^n)$ be a BDM-with-betting mechanism where the proper scoring rule is a simple extension of the continuous ranked probability score. More precisely, the scoring rule is

$$\mathrm{SR}\,(b_i, \tilde{v}_i) = \bar{v}m - \sum_{x=1}^{m} \int_{[0,\bar{v}]} \left(\Pr_{b_i}[\tilde{v}_{ix} \leq z] - \mathbf{1}(\tilde{v}_{ix} \leq z)\right)^2 dz$$

Note that we apply the continuous ranked probability score separately, for each alternative x, to the reported marginal distribution over the average reported value $\tilde{v}_{ix}$. The overall score is the sum of the scores associated with each alternative x.

B.3.1 Notation and Sufficient Conditions

We begin by introducing key notation and finding sufficient conditions for eliciting the population's average value.

Let agent i's *error* be the difference between her true and reported values, i.e.,

$$\epsilon_i = v_i - \hat{v}_i$$

Since the BDM mechanism incentivizes truth-telling, her reported value is

$$\hat{v}_i = \mathrm{E}[v_i \mid s_i(\omega, \theta_i, v_i), \theta_i]$$

The *average error* is the difference between the *average value* and the actual average reported value. That is,

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} \epsilon_i}_{\text{average error}} = \underbrace{\frac{1}{n}\sum_{i=1}^{n} v_i}_{\text{average value}} - \underbrace{\frac{1}{n}\sum_{i=1}^{n} \hat{v}_i}_{\text{actual average reported value}}$$

where the average value is what the average reported value would have been if each participant i knew her own values v_i. If the average error is small, the alternative will be nearly optimal.

In the limit as the sample size n grows, the average error, average value, and actual average reported value converge in probability:

$$\frac{1}{n}\sum_{i=1}^{n} \epsilon_i \to_p \mathrm{E}[\epsilon_i \mid \omega] \qquad \frac{1}{n}\sum_{i=1}^{n} v_i \to_p \mathrm{E}[v_i \mid \omega] \qquad \frac{1}{n}\sum_{i=1}^{n} \hat{v}_i \to_p \mathrm{E}[\hat{v}_i \mid \omega] \tag{12}$$

These averages converge in probability to expectations $\mathrm{E}[\cdot \mid \omega]$ conditional on the realized state ω. This follows from the fact that the variables $\epsilon_i, v_i, \hat{v}_i$ are conditionally i.i.d., which allows us to invoke the law of large numbers after conditioning on ω.

For convenience, we often refer to the probability limits (12) as the average error, average value, and actual average reported value, respectively.

To show that the alternative is nearly optimal, it is sufficient to show that the average error vanishes, i.e.,

$$\mathrm{E}[\epsilon_i \mid \omega] \to_p 0 \tag{13}$$

which ensures that the actual average reported value converges to the average value. But, as we saw in Example 2, it is not critical that the individual errors vanish. As long as the errors ϵ_i are uncorrelated across participants i, the average error will vanish. This is why BDM-with-betting works even though it does not directly incentivize participants to learn about their own values: it really only needs to incentivize participants i, j to acquire information up to the point where their errors ϵ_i, ϵ_j are uncorrelated.

To show that the average error vanishes (13), it suffices to show that

$$\mathrm{Cov}[\epsilon_i, \mathrm{E}[\hat{v}_i \mid \omega]] \to 0 \quad \text{and} \quad \mathrm{Cov}\left[\epsilon_i, \mathrm{E}\left[v_j \mid \omega\right]\right] \to 0 \tag{14}$$

Intuitively, if the error ϵ_i is neither correlated with the actual average reported value $\mathrm{E}[\hat{v}_i \mid \omega]$, nor with the average value $\mathrm{E}\left[v_j \mid \omega\right]$, then it cannot be correlated with the average error

$$\mathrm{E}\left[\epsilon_j \mid \omega\right] = \mathrm{E}\left[v_j \mid \omega\right] - \mathrm{E}[\hat{v}_i \mid \omega]$$

If ϵ_i is not correlated with the average error, then it must not be correlated with the errors ϵ_j of other participants j. As we just discussed, this ensures that the average error vanishes.

B.3.2 Verifying Sufficient Conditions

We verify the sufficient conditions for eliciting the average, in three lemmas.

First, we show that agent i's error ϵ_i cannot be too correlated with the average reported value $\mathrm{E}\left[\hat{v}_j \mid \omega\right]$.

Lemma 2. *In any equilibrium of the mechanism $(\boldsymbol{x}_n, \boldsymbol{t}_n)$, the cross-covariance matrix between the error ϵ_i and the conditional expected report $\mathrm{E}\left[\hat{v}_j \mid \omega\right]$ has bounded entries. That is,*

$$\left\|\mathrm{Cov}\left[\epsilon_i, \mathrm{E}\left[\hat{v}_j \mid \omega\right]\right]\right\| \leq \sqrt{\frac{\bar{v}^3 \bar{c}}{\lambda_n}}$$

where $\|\cdot\|$ is the sup-norm.

Lemma 2 follows from the fact that participant i wants to acquire any signal that helps her better predict the average reported value $\mathrm{E}\left[\hat{v}_j \mid \omega\right]$. Suppose that, for the sake of contradiction, her error ϵ_i is correlated with the average reported value. Then, if there were a signal that revealed her error ϵ_i, she would want to acquire it. However, this signal exists. By Assumptions 1 and 2, i can combine her revealing signal (which reveals v_i) with the signals that she has already acquired (which determine $\hat{v}_i$). This reveals her error $\epsilon_i = v_i - \hat{v}_i$ and costs at most $\bar{c}$ more than the signals she already acquired. This combination is a profitable deviation whenever the scaling parameter λ_n is large relative to the cost $\bar{c}$. That, in turn, contradicts the premise that ϵ_i is correlated with the average reported value in equilibrium.

Next, we must show that ϵ_i is not too correlated with the average value $\mathrm{E}\left[v_j \mid \omega\right]$. However, this is not true in every equilibrium. We must show that there exists an equilibrium in which ϵ_i is not too correlated with the average value.

We take an indirect approach. For every model specification, we define an *auxiliary model specification* where, by construction, ϵ_i is not correlated with the average value. Using Lemma 2, we show that the average error vanishes in every equilibrium of the auxiliary model specification (Lemma 3). Then we show, when n is large, every equilibrium of the auxiliary model specification is also an equilibrium of the original model specification (Lemma 4).

Essentially, the auxiliary model specification forces agents to only acquire signals that are maximally predictive of the conditional expectation $\mathrm{E}\left[v_j \mid \omega\right]$.

Definition 17. *For any given model specification* $(\Omega, \Theta, \mathcal{S}, G, F)$*, we define an* auxiliary model specification $(\Omega, \Theta, \mathcal{S}, F, \tilde{c})$*. There are three steps to this construction.*

1. *For any given type* $\theta_i \in \Theta$*, let* $\bar{s}_i^{\theta_i}$ *be the signal that combines all other signals* $s_i \in \mathcal{S}$ *that have finite cost* $c\left(s_i, \theta_i\right) < \infty$.[23] *This exists by Assumption 1.*

2. *For any given type* $\theta_i \in \Theta$*, we can evaluate a given signal* s_i *by how well it predicts the average value* $\mathrm{E}\left[v_j \mid \omega\right]$*. Formally, let* predictiveness *be the maximum expected score when type* θ_i *acquires signal* s_i*, i.e.,*

$$\rho(\theta_i, s_i) = \mathrm{E}\left[\max_{b_i} \mathrm{E}[\mathrm{SR}\left(b_i, \mathrm{E}[v_i \mid \omega]\right) \mid \theta_i, s_i(\omega, \theta_i, v_i)] \mid \theta_i\right] \tag{15}$$

3. *For any given type* $\theta_i \in \Theta_i$*, let every signal* s_i *that is less predictive than the combined signal* $\bar{s}_i^{\theta_i}$ *have infinite cost according to the auxiliary cost function* $\tilde{c}$*. Let all other signals cost the same as in the original cost function* c*. More precisely,*

$$\tilde{c}(\theta_i, s_i) = \begin{cases} \infty & \rho\left(\theta_i, s_i\right) < \rho\left(\theta_i, \bar{s}_i^{\theta_i}\right) \\ c(\theta_i, s_i) & \text{otherwise} \end{cases} \tag{16}$$

In Lemma 3, we show that the average error vanishes in every equilibrium of the auxiliary model specification.

[23] If the signal s_i is a combined signal as defined in Assumption 1, let $\bar{s}_i^{\theta_i}$ include all base signals that s_i combines.

Lemma 3. *Fix the auxiliary model specification. For any sequence of equilibria* $(\boldsymbol{s}^n, \boldsymbol{m}^n)$ *of the mechanisms* $(\boldsymbol{x}^n, \boldsymbol{t}^n)$*, the average reported value converges in probability to the population's average value.*

The proof of Lemma 3 follows the logic of condition (14). Lemma 2 already that the error ϵ_i cannot be too correlated with the average reported value $\mathrm{E}[\hat{v}_j \mid \omega]$. All that remains, by condition (14), is to show that the error ϵ_i cannot be correlated with the average value $\mathrm{E}[v_j \mid \omega]$. This is because the only signals s_i with finite cost in the auxiliary model specification are those that are maximally predictive of the average value. Suppose, for the sake of contradiction, that ϵ_i is correlated with the average value. Then there exists another signal – which combines s_i with the revealing signal, and reveals ϵ_i – that predicts the average value better than s_i. This contradicts the premise that s_i is maximally predictive.

At this point, we have shown that the average error vanishes for all auxiliary model specifications. Lemma 4 extends this result to general model specifications.

Lemma 4. *There exists a constant* N *such that, for any sample size* $n \geq N$*, every equilibrium of the auxiliary model specification is also an equilibrium of the original model specification.*

To prove Lemma 4, it is enough to show that, in the candidate equilibrium of the original model specification, no participant i will deviate to a signal that is unavailable in the auxiliary model specification. That is, we want to show that participant i wants to acquire only signals that are maximally predictive of the average value $\mathrm{E}[v_j \mid \omega]$. We know, in the limit, that she will acquire only signals that are maximally predictive of the average reported value $\mathrm{E}[\hat{v}_j \mid \omega]$. Moreover, by Lemma 3, in any equilibrium of the auxiliary model specification, the average reported value converges to the average value (as the average error vanishes). Therefore, she will only acquire signals that are maximally predictive of the average value.

This completes the proof of Theorem 1. The proofs of Lemmas 2-4 are somewhat involved, so we leave them to the Supplemental Appendix.

B.4 Proof of Theorem 3

C Supplemental Appendix

C.1 Proof of Theorem 4

We outline the proof of Theorem 4. The argument is similar to the proof of Theorem 3 (see Section B.4). The two differences are that we need a new mechanism and that we generalize the argument to $m \geq 2$ alternatives. This does not affect the intuition behind the result, but does introduce some technical complications. Increasing the number of alternatives also dramatically slows the rate of convergence.

C.1.1 VCG-with-Betting Mechanism

We define a new mechanism called *VCG-with-betting*. This mechanism is analogous to BDM-with-betting and majority-rule-with-betting. The difference is that, in the second stage, agents report their willingness to pay to the VCG mechanism. As with majority-rule-with-betting, we also need to add some noise to the mechanism to avoid issues that arise when agents condition on low-probability events.

The VCG-with-betting mechanism has four tuning parameters. First, there is proper scoring rule SR, which evaluates the accuracy of agent i's reported beliefs $\hat{b}_i$ over the average reported value $\tilde{v}_i$. Second, there is the scaling parameter λ that controls how large the betting stakes are. Third, there is a probability ϵ that some agent i is chosen as dictator. Fourth, there is a randomized *bias term* $v_0 \in \mathcal{V}$. This biases the mechanism towards alternatives x where v_{0x} is large. Let the average reported value $\tilde{v}_i$ include the bias term, i.e.,

$$\tilde{v}_i = \frac{1}{n+1}\left(v_0 + \sum_{j \neq i}^{n} \hat{v}_j\right)$$

The agents do not know the realization of the bias term.

Definition 18. *The* VCG-with-betting mechanism $(\boldsymbol{x}, \boldsymbol{t})$ *is parameterized by a proper scoring rule* SR*, scaling parameter λ, probability ϵ, and randomized bias term v_0. Each agent i sends a message*

$$m_i = \left(\hat{v}_i, \hat{b}_i\right) \in \mathcal{V} \times \mathcal{B}$$

that consists of a reported value $\hat{v}_i$ and a reported belief $\hat{b}_i$. There are two cases.

1. *With probability $1 - \epsilon$, the planner selects the alternative that maximizes the average reported value plus the bias term, i.e.,*

$$\boldsymbol{x}\left(\hat{v}, \hat{b}\right) \in \arg\max_{x \in \mathcal{X}} \left(v_{0x} + \sum_{i=1}^{n} \hat{v}_{ix} \right)$$

Each agent i is paid according to the VCG mechanism's transfer rule and the proper scoring rule, i.e.,

$$\boldsymbol{t}_i\left(\hat{v}, \hat{b}\right) = v_{0\boldsymbol{x}\left(\hat{v}, \hat{b}\right)} + \sum_{j \neq i} \hat{v}_{j\boldsymbol{x}\left(\hat{v}, \hat{b}\right)} - \max_{x \in X} \left(v_{0x} + \sum_{j \neq i} \hat{v}_{jx} \right) + \lambda \cdot \mathrm{SR}\left(\hat{b}_i, \tilde{v}_i\right)$$

2. *With probability ϵ, the planner randomly chooses an agent $i \sim$ UNIFORM$(1, \ldots, n)$. This agent is paid according to a BDM mechanism (where the bias term v_0 determines the randomized price) and the proper scoring rule, i.e.,*

$$\boldsymbol{t}_i\left(\hat{v}, \hat{b}\right) = v_{0\boldsymbol{x}\left(\hat{v}, \hat{b}\right)} - \max_{x \in X} v_{0x} + \lambda \cdot \mathrm{SR}\left(\hat{b}_i, \tilde{v}_i\right)$$

The other agents $j \neq i$ are paid nothing.

C.1.2 Sequence of VCG-with-Betting Mechanisms

Theorem 4 refers to a sequence of mechanisms $\left(\mathbf{x}^n, \mathbf{t}^n\right)$. More precisely, this is a sequence of VCG-with-betting mechanisms, with scaling parameter λ_n, probability ϵ_n, and bias term v_0 drawn according to a distribution that depends on n.

First, we specify the scoring rule SR. Suppose that the agent i's reports beliefs in the form of a probability density function $\hat{b}_i : [0, \bar{v}]^m \to \mathbb{R}_+$.[24] Let SR be the sum of the quadratic scoring rule and the continuous ranked probability score used in the

[24] This is without loss. The noise terms that we will add to the VCG-with-betting mechanism ensure that the distribution of $\tilde{v}_i$ is absolutely continuous and, therefore, admits a probability density function.

proof of Theorem 2. That is,

$$\mathrm{SR}\left(b_i, \tilde{v}_i\right) = \underbrace{2\hat{b}_i(\tilde{v}_i) - \int_{\mathbb{R}} \hat{b}_i^2(z)dz}_{\text{quadratic scoring rule}} + \underbrace{\bar{v}m - \sum_{x=1}^{m} \int_{[0,\bar{v}]} \left(\Pr_{b_i}[\tilde{v}_{ix} \leq z] - \mathbf{1}(\tilde{v}_{ix} \leq z)\right)^2 dz}_{\text{continuous ranked probability score from Theorem 2}}$$

This scoring rule is proper since the sum of two proper scoring rules is proper. The reason to keep the continuous ranked probability score is because it makes it easier to adapt parts of the proof of Theorem 2 when proving Theorem 4.

Second, we specify the distribution of the bias term. We allow the bias term to be correlated with the ϵ_n probability event where some agent i is chosen as quasi-dictator. If an agent is chosen as quasi-dictator, then v_0 is uniformly distributed, i.e.,

$$v_{0x} \sim \textsc{Uniform}\left([0, \bar{v}]\right) \quad \text{i.i.d. across } x \in \mathcal{X}$$

Otherwise, v_{0x} follows a Laplace distribution with scaling parameter $\beta_n > 0$, i.e.,

$$v_{0x} \sim \textsc{Laplace}\left(0, \beta_n\right) \quad \text{i.i.d. across } x \in \mathcal{X}$$

where v_{0x} is drawn independently of $v_{0x'}$ for any two outcomes x, x'. The uniform distribution ensures that agents have strict incentives to report truthfully conditional on being quasi-dictator. The Laplace distribution ensures that agent i's expected value conditional on the vote margin $\tilde{v}_i$ is a smooth function of $\tilde{v}_i$.

C.1.3 Convergence to Efficiency

We begin by formalizing what it means for VCG-with-betting to be nearly truthful. Essentially, each agent's reported values should be probably approximately equal to her true expected values.

Definition 19. *The VCG-with-betting mechanism* $(\boldsymbol{x}^n, \boldsymbol{t}^n)$ *is* (ϕ_1, ϕ_2)-truthful *if the following holds for every model specification and equilibrium* $(\boldsymbol{s}, \hat{\boldsymbol{v}}, \hat{\boldsymbol{b}})$*. Let agent i acquire signal* $s_i = \boldsymbol{s}_i(\theta_i, c_i)$ *and report values* $\hat{v}_i = \hat{\boldsymbol{v}}_i\left(s_i(\omega, \theta_i, v_i), \theta_i, c_i\right)$*. Then*

$$\forall \theta_i \in \Theta, \quad \Pr[\|\hat{v}_i - \mathrm{E}[v_i \mid s_i(\omega, \theta_i, v_i), \theta_i]\| > \phi_1] \leq \phi_2$$

where $\|\cdot\|$ *is the sup-norm.*

Next, we verify that the VCG-with-betting mechanism is nearly-truthful. The intuition is analogous to the intuition for majority-rule-with-betting in Section 5.4.

Lemma 5. *Let parameters $\lambda_n \to \infty$, $\epsilon_n \to 0$, and $\beta_n \to 0$ at appropriate rates. The VCG-with-betting mechanism $(\boldsymbol{x}^n, \boldsymbol{t}^n)$ is (ϕ_{1n}, ϕ_{2n})-truthful where*

$$\lim_{n\to\infty} \phi_{1n} = 0 \quad \textit{and} \quad \lim_{n\to\infty} \phi_{2n} = 0$$

This is the key step in proving Theorem 4. The rest of the proof follows from a generalized version of Proposition 1, which holds even if agents report nearly truthfully rather than exactly truthfully (see Section C.2). By construction, VCG-with-betting nearly maximizes the average reported value with high probability. Proposition 1 ensures that the average reported value converges to the population's average value, in every equilibrium. Therefore, VCG-with-betting nearly maximizes the population's average value with high probability, in every equilibrium.

C.2 Proof of Proposition 1

C.3 Proof of Proposition 2

C.4 Proof of Lemma 1

We first discuss strategy profile $(\mathbf{s}, \mathbf{m})$, and then turn to strategy profile $(\mathbf{s}', \mathbf{m}')$.

Let $(\mathbf{s}, \mathbf{m})$ be some equilibrium when the model specification is $(\Omega, \Theta, \mathcal{S}, G, F)$. Define messages and signals in equilibrium as follows:

$$m_i = \mathbf{m}_i\left(s_i(\omega, \theta_i, v_i), \theta_i, c_i\right) \quad \text{and} \quad s_i = \mathbf{s}_i\left(\theta_i, c_i\right)$$

I claim that the vector $(v_i, \theta_i, c_i, s_i, m_i)$ is independent across agents i. This follows from three observations. First, the cost functions c_i are independent across agents i.[25] Second, the values v_i and types θ_i are independent across agents i after conditioning on the state ω. Third, the state is $\omega = 1$ with probability one, so values v_i and types θ_i are unconditionally independent across agents i.

[25]Recall that, in general, the realized costs $c_i(\theta_i, s_i)$ are type-dependent and therefore may be correlated across agents. However, the cost functions c_i themselves are independent.

Consider the distribution of acquired signals s_i. We claim that agent i acquires the revealing signal with probability $1-p$. We begin by making some observations. The revealing signal is only useful to agent i insofar as it affects her choice of which message m_i to send. With probability $1-p$, the cost of the revealing signal is zero and we can assume, without loss, that agent i acquires it. With probability p, the cost of the revealing signal is $\tilde{c} > 0$ and the agent will only acquire it if she benefits from the information that it provides. In this case, the type and state are fixed with probability one, so the information is her value v_i. However, this information does not affect her beliefs about other agents' messages, since v_i and m_{-i} are independent.

Suppose that agent i acquires the revealing signal when her cost of doing so is $\tilde{c}$. We want to find a profitable deviation. Consider her expected utility

$$U_i(\mathbf{s}, \mathbf{m}, \mathbf{x}, \mathbf{t}) = \mathrm{E}[u_i(v_i, \mathbf{x}_i(m), \mathbf{t}_i(m), s_i, \theta_i, c_i)]$$

Recall that her value $v_i \in \{a, b\}$. Suppose agent i sends message m_i^a if she learns $v_i = a$ and m_i^b if she learns $v_i = b$. Then

$$\begin{aligned} &\mathrm{E}\big[u_i(a, \mathbf{x}_i(m_i^b, m_{-i}, \mathbf{t}_i(m_i^b, m_{-i}), s_i, \theta_i, c_i)\big] \\ &\geq \mathrm{E}\big[u_i(b, \mathbf{x}_i(m_i^b, m_{-i}, \mathbf{t}_i(m_i^b, m_{-i}), s_i, \theta_i, c_i)\big] - \epsilon && (17) \\ &\geq \mathrm{E}\big[u_i(b, \mathbf{x}_i(m_i^a, m_{-i}, \mathbf{t}_i(m_i^a, m_{-i}), s_i, \theta_i, c_i)\big] - \epsilon && (18) \\ &\geq \mathrm{E}\big[u_i(a, \mathbf{x}_i(m_i^a, m_{-i}, \mathbf{t}_i(m_i^a, m_{-i}), s_i, \theta_i, c_i)\big] - 2\epsilon && (19) \end{aligned}$$

where inequalities (17) and (19) hold because the distance between a and b is at most ϵ, and inequality (18) holds because agent i prefers to send message m_i^b when $v_i = b$. Note that we do not need to condition the expectation on the agent's private information.[26] Now, consider the deviation where agent i does not acquire the revealing signal and always sends message m_i^b. Then she reduces her costs by $\bar{c}$ and, by inequalities (17)-(19), reduces her gains by at most 2ϵ. This is a profitable deviation as long as $\epsilon < \bar{c}/2$.

Next, we construct a strategy profile $(\mathbf{s}', \mathbf{m}')$ and show that it is an equilibrium when the model specification is $(\Omega, \Theta, \mathcal{S}, G, F')$. Let agent i acquire the uninformative signal $s_i^{\emptyset}$ regardless of her type θ_i or cost function c_i. Let her send messages as follows.

1. If type $\theta_i = 1$, randomly draw a cost function $\hat{c}_i \sim G$ where type $\theta_i = 1$'s cost of

[26] This is because v_i is independent of m_{-i}, v_i is independent of c_i, and θ_i is fixed with probability one.

the revealing signal is $\hat{c}_i^1 > 0$. Then send the message $\mathbf{m}_i(0, 1, \hat{c}_i)$.

2. If type $\theta_i = 2$, randomly draw a cost function $\hat{c}_i \sim G$ where type $\theta_i = 1$'s cost of the revealing signal is $\hat{c}_i^1 = 0$. Then send the message $\mathbf{m}_i(a, 1, \hat{c}_i)$.

3. If type $\theta_i = 2$, randomly draw a cost function $\hat{c}_i \sim G$ where type $\theta_i = 1$'s cost of the revealing signal is $\hat{c}_i^1 = 0$. Then send the message $\mathbf{m}_i(b, 1, \hat{c}_i)$.

That is, the three types of agents that arise in joint distribution F' mimic pseudo-types of agents that arise in joint distribution F. The first pseudo-type is the agent whose revealing signal is costly. The second and third pseudo-types are agents whose revealing signal costs nothing and whose value is $v_i = a$ and $v_i = b$, respectively.

To see that $(\mathbf{s}', \mathbf{m}')$ is an equilibrium, first consider information acquisition. When the joint distribution is F', agent v_i knows her value perfectly, and she knows the state $\omega = 1$ with probability one. Therefore, there is no informational benefit to acquiring the revealing signal. The agent will always weakly prefer to acquire the uninformative signal, as prescribed.

We must also show that agent i does not want to deviate from message rule $\mathbf{m}_i'$. We rely on the following claim.

Claim 1. *The distribution of message profiles m' given joint distribution F', where*

$$m_i' = \boldsymbol{m}_i'\left(s_i'(\omega, \theta_i, v_i), \theta_i, c_i\right) \text{ and } \quad s_i' = \boldsymbol{s}_i'(\theta_i, c_i)$$

is the same as the distribution of message profiles m given joint distribution F.

Suppose that agent i of type θ_i sends message $\tilde{m}_i$. Having acquired the uninformative signal, her expected utility is

$$\mathrm{E}_{F'}\left[v_i \cdot \mathbf{x}_i(\tilde{m}_i, m_{-i}') + \mathbf{t}_i(\tilde{m}_i, m_{-i}') \mid \theta_i\right]$$

There are three cases to consider, corresponding to the three possible types θ_i.

1. Suppose $\theta_i = 1$. Then agent i's expected utility is

$$\begin{aligned}
&\mathrm{E}_{F'}\left[\frac{a+b}{2} \cdot \mathbf{x}_i(\tilde{m}_i, m_{-i}') + \mathbf{t}_i(\tilde{m}_i, m_{-i}') \mid \theta_i = 1\right] \\
&= \mathrm{E}_{F'}\left[\frac{a+b}{2} \cdot \mathbf{x}_i(\tilde{m}_i, m_{-i}') + \mathbf{t}_i(\tilde{m}_i, m_{-i}')\right] \qquad \text{(since } \theta_i \perp m_{-i}')
\end{aligned}$$

$$
\begin{aligned}
&= \mathrm{E}_F\left[\frac{a+b}{2} \cdot \mathbf{x}_i(\tilde{m}_i, m_{-i}) + \mathbf{t}_i(\tilde{m}_i, m_{-i})\right] && \text{(by Claim 1)} \\
&= \mathrm{E}_F[\mathrm{E}_F[v_i] \cdot \mathbf{x}_i(\tilde{m}_i, m_{-i}) + \mathbf{t}_i(\tilde{m}_i, m_{-i})] && \text{(by inspection of } F\text{)} \\
&= \mathrm{E}_F[v_i \cdot \mathbf{x}_i(\tilde{m}_i, m_{-i}) + \mathbf{t}_i(\tilde{m}_i, m_{-i})] && \text{(by LIE)}
\end{aligned}
$$

This objective is identical to the expected utility of the first pseudo-type that arises in joint distribution F. Since $(\mathbf{s}, \mathbf{m})$ is an equilibrium when the joint distribution is F, the first pseudo-type must (almost surely) choose a message $\tilde{m}_i$ that maximizes this objective. It follows that agent i must also choose messages $\tilde{m}_i$ that maximize this objective, since she mimics the first pseudo-type.

2. Suppose $\theta_i = 2$. Then agent i's expected utility is

$$
\begin{aligned}
&\mathrm{E}_{F'}\left[a \cdot \mathbf{x}_i(\tilde{m}_i, m'_{-i}) + \mathbf{t}_i(\tilde{m}_i, m'_{-i}) \mid \theta_i = 1\right] \\
&\quad = \mathrm{E}_{F'}\left[a \cdot \mathbf{x}_i(\tilde{m}_i, m'_{-i}) + \mathbf{t}_i(\tilde{m}_i, m'_{-i})\right] && \text{(since } \theta_i \perp m'_{-i}\text{)} \\
&\quad = \mathrm{E}_F[a \cdot \mathbf{x}_i(\tilde{m}_i, m_{-i}) + \mathbf{t}_i(\tilde{m}_i, m_{-i})] && \text{(by Claim 1)} \\
&\quad = \mathrm{E}_F[v_i \cdot \mathbf{x}_i(\tilde{m}_i, m_{-i}) + \mathbf{t}_i(\tilde{m}_i, m_{-i}) \mid v_i = a]
\end{aligned}
$$

This objective is identical to the expected utility of the second pseudo-type that arises in joint distribution F. She conditions on $v_i = a$ because she acquires the revealing signal and v_i is the signal realization. Since agent i mimics the second-pseudo type, we use the same argument as in the previous case to conclude that agent i is optimizing.

3. Suppose $\theta_i = 3$. This argument is identical to the previous case, except that we replace $v_i = a$ with $v_i = b$ and "second pseudo-type" with "third pseudo-type".

It follows that message m'_i maximizes agent i's expected utility.

Having shown that $(\mathbf{s}', \mathbf{m}')$ is an equilibrium when the model specification is $(\Omega, \Theta, \mathcal{S}, G, F')$, the proof is almost complete. The final part of Lemma 1 states that the distribution of message profile m given joint distribution F is the same as the distribution of m' given joint distribution F'. This is precisely Claim 1.

C.4.1 Proof of Claim 1

We begin by showing that the messages m_i and m'_i have identical marginal distributions. This follows from three observations.

1. Suppose the joint distribution is F. With probability p, agent i has a positive cost $c_i^1 > 0$ for the revealing signal. Her message is

$$m_i = \mathbf{m}_i(s_i^{\emptyset}(\omega, \theta_i, v_i), \theta_i, c_i) = \mathbf{m}_i(0, 1, c_i)$$

   where the second equality follows from the fact that $s_i^{\emptyset}(\omega, \theta_i, v_i) = 0$ and $\theta_i = 1$.

   Suppose the joint distribution is F'. With probability p, agent i's type is $\theta_i = 1$. Her message is $m_i' = \mathbf{m}(0, 1, \hat{c}_i)$. The is the same as above, except that c_i is replaced with $\hat{c}_i$ (which is identically distributed given that $c_i^1 > 0$).

2. Suppose the joint distribution is F. With probability $(1-p)/2$, agent i has value $v_i = a$ and zero cost $c_i^1 = 0$ for the revealing signal. Her message is

$$m_i = \mathbf{m}_i(s_i^{\theta_i}(\omega, \theta_i, v_i), \theta_i, c_i) = \mathbf{m}_i(a, 1, c_i)$$

   where the second equality follows from the fact that $s_i^{\emptyset}(\omega, \theta_i, v_i) = a$ and $\theta_i = 1$.

   Suppose the joint distribution is F'. With probability $(1-p)/2$, agent i's type is $\theta_i = 2$. Her message is $m_i' = \mathbf{m}(a, 1, \hat{c}_i)$. The is the same as above, except that c_i is replaced with $\hat{c}_i$ (which is identically distributed given that $c_i^1 = 0$).

3. Repeat the last argument, replacing $v_i = a$ with $v_i = b$ and $\theta_i = 2$ with $\theta_i = 3$.

Furthermore, the message m_i' is independent of m_{-i}', by the same argument that showed m_i is independent of m_{-i}. Therefore, the distribution of message profiles m' given F' is the same as the distribution of message profiles m given F.

C.5 Proof of Lemma 2

C.6 Proof of Lemma 3

C.7 Proof of Lemma 4

This proof has four parts. First, consider any signal s_i' whose cost in the auxiliary model specification was different from its cost in the original model specification. Equivalently, consider any signal s_i' where

$$\rho\left(\theta_i, \bar{s}_i^{\theta_i}\right) - \rho\left(\theta_i, s_i\right) > 0$$

Since $\Theta \times \mathcal{S}$ is a finite set, we can define a gap $\delta > 0$ where

$$\delta := \max_{(\theta_i, s_i) \text{ s.t. } \rho(\theta_i, s_i) < \rho\left(\theta_i, \bar{s}_i^{\theta_i}\right)} \rho\left(\theta_i, \bar{s}_i^{\theta_i}\right) - \rho\left(\theta_i, s_i\right) \tag{20}$$

This will be useful later in the proof.

Second, we consider a sequence of moments and show that it converges. For any random variable Z, define a moment that depends on additional parameter $z \in \mathbb{R}$ as follows:

$$f(Z, z) = \mathrm{E}\left[\left(\Pr_i[Z \le z] - \mathbf{1}(Z \le z)\right)^2 \mid \theta_i\right]$$

Let Z_n be a sequence of random variables where $Z_n \to_p Z$ conditional on θ_i. To find $\lim_{n\to\infty} f(Z_n, \cdot)$, we first consider the limits of two random variables that appear in the definition of f.

1. The random variable $\Pr_i[Z_n \le z]$ converges in probability to $\Pr_i[Z \le z]$, conditional on θ_i and for almost all z. To see this, fix any constant $t > 0$ and define the probability

   $$p_n^+(\theta_i) := \Pr[\Pr_i[Z \le z] - \Pr_i[Z_n \le z] \ge t \mid \theta_i]$$

   I begin by showing that $p_n^+(\theta_i) \to 0$. Observe that

   $$\begin{aligned} \Pr[Z \le z \mid \theta_i] - \Pr[Z_n \le z \mid \theta_i] &= \mathrm{E}[\Pr_i[Z \le z] - \Pr_i[Z_n \le z]] && \text{(LIE)} \\ &\ge p_n^+(\theta_i) \cdot t + (1 - p_n^+(\theta_i)) \cdot O(t) && \text{(LIE)} \end{aligned}$$

   However, since $Z_n \to_p Z$, we know that $Z_n \to_d Z$. By the Portmanteau theorem,

   $$\Pr[Z_n \le z \mid \theta_i] \to \Pr[Z \le z \mid \theta_i] \quad \text{for almost all } z \in \mathbb{R}$$ [27]

   Combining this with the previous inequality implies

   $$p_n^+(\theta_i) \cdot t + (1 - p_n^+(\theta_i)) \to 0$$

   Since t does not depend on n, this can only be true when $p_n^+(\theta_i) \to 0$. By

[27] The distribution of Z can have at most finitely many atoms. As long as z is not one of those atoms, the set of points$\{z' \in \mathbb{R} \mid z' \le z\}$ is a continuity set for that distribution, and we can apply the Portmanteau theorem.

symmetry, we can apply this argument again to argue that

$$p_n^-(\theta_i) := \Pr[\Pr_i[Z_n \le z] - \Pr_i[Z \le z] \ge t \mid \theta_i] \to 0$$

By the definition of convergence in probability, the fact that $p_n^+(\theta_i) \to 0$ and $p_n^-(\theta_i) \to 0$ implies that $\Pr_i[Z_n \le z] \to_p \Pr_i[Z \le z]$ conditional on θ_i

2. The random variable $\mathbf{1}(Z_n \le z)$ converges in probability to $\mathbf{1}(Z \le z)$, conditional on θ_i and for almost all z. This follows from the continuous mapping theorem. The distribution of Z can have at most finitely many atoms and the indicator function has only one discontinuity point, at z. Provided that z is not one of those atoms, the probability that Z matches a discontinuity point is zero. Therefore, we can invoke the continuous mapping theorem for all but finitely many values of z.

It follows from these two results and the continuous mapping theorem that

$$(\Pr_i[Z_n \le z] - \mathbf{1}(Z_n \le z))^2 \to_p (\Pr_i[Z \le z] - \mathbf{1}(Z \le z))^2 \quad \text{conditional on } \theta_i$$

for almost all $z \in \mathbb{R}$. Since convergence in probability implies convergence in distribution, we can apply the Portmanteau theorem to show that

$$f(Z_n, \cdot) \to f(Z, \cdot) \quad \text{a.e.}$$

Set $Z_n = \tilde{v}_{ix,n}$ and $Z = \mathrm{E}\left[v_{jx} \mid \omega\right]$, where $Z_n \to_p Z$ conditional on θ_i by Claim **??**. Then

$$f(\tilde{v}_{ix,n}, \cdot) \to f(\mathrm{E}\left[v_{jx} \mid \omega\right], \cdot) \quad \text{a.e.}$$

We will use this fact momentarily, when applying the bounded convergence theorem.

Third, we characterize the limiting behavior of $\tilde{\rho}_n(\theta_i, s_i)$ as $n \to \infty$.

$$\begin{aligned}
\lim_{n\to\infty} \lambda_n^{-1}\tilde{\rho}_n(\theta_i, s_i) &= \bar{v}m - \sum_{x=1}^{m} \lim_{n\to\infty} \int_0^{\bar{v}} f(\tilde{v}_{ix,n}, z)dz && \text{(definition of } \tilde{\rho}_n) \\
&= \bar{v}m - \sum_{x=1}^{m} \int_0^{\bar{v}} f(\mathrm{E}\left[v_{jx} \mid \omega\right], z)dz && \text{(bounded convergence theorem)} \\
&= \rho(\theta_i, s_i) && \text{(definition of } \rho)
\end{aligned}$$

It follows that there exists a threshold $N(\theta_i, s_i)$ such that, for any $n \geq N(\theta_i, s_i)$,

$$\left|\rho(\theta_i, s_i) - \lambda_n^{-1}\tilde{\rho}_n(\theta_i, s_i)\right| < \frac{\delta}{4} \tag{21}$$

Let $N = \max_{(\theta_i, s_i)\in\Theta\times\mathcal{S}} N(\theta_i, s_i)$. This quantity exists since $\Theta \times \mathcal{S}$ is a finite set.

Fourth and finally, we show that the agent i of type θ_i does not have a profitable deviation when n is large. Assume $n \geq N$. Let s_i be the signal that she acquires in the equilibrium of the auxiliary model specification. We only need to consider deviations to signals s_i' that were not available in the auxiliary model specification, which implies

$$\rho\left(\theta_i, \bar{s}_i^{\theta_i}\right) - \rho\left(\theta_i, s_i'\right) \geq \delta \tag{22}$$

where $\delta > 0$ was defined in equation (20). Since s_i was chosen in equilibrium of the auxiliary model specification, it must have finite cost. By the definition of $\tilde{c}$ in (16), this implies

$$\rho\left(\theta_i, \bar{s}_i^{\theta_i}\right) = \rho\left(\theta_i, s_i\right) \tag{23}$$

$$\begin{aligned}
\tilde{\rho}_n\left(\theta_i, s_i\right) - \tilde{\rho}_n\left(\theta_i, s_i'\right) &> \lambda_n \rho\left(\theta_i, s_i\right) - \rho\left(\theta_i, s_i'\right) - \frac{\delta\lambda_n}{2} && \text{(equation (21))}\\
&= \lambda_n \rho\left(\theta_i, \bar{s}_i^{\theta_i}\right) - \rho\left(\theta_i, s_i'\right) - \frac{\delta\lambda_n}{2} && \text{(equation (23))}\\
&\geq \frac{\delta\lambda_n}{2} && \text{(equation (22))}
\end{aligned}$$

This deviation is not profitable when the loss in expected transfers exceeds the cost. We just characterized the loss in expected transfers, and the difference in costs is at most $\tilde{\tilde{c}}$. Therefore, a sufficient condition for the deviation to not be profitable is

$$\frac{\delta\lambda_n}{2} \geq \tilde{\tilde{c}}$$

Since $\lambda_n \to \infty$, there is some threshold N' such that this condition holds for all $n \geq N'$. Then the agent has no profitable deviations whenever $n \geq \max\{N, N'\}$.

C.8 Proof of Lemma ??

C.9 Proof of Lemma ??

C.10 Proof of Lemma ??

C.11 Proof of Lemma 5