Divide and Confer: Aggregating Information without Verification

James Best, Daniel Quigley, Maryam Saeedi, Ali Shourideh*

May 28, 2026

Abstract

We study mechanisms for aggregating information divided across a large population of biased senders. Each sender privately observes an unconditionally independent signal about an unknown state, so no sender's report can be verified against another's. A receiver makes a binary accept/reject decision whose payoffs depend on the state. Even though cross-verification is impossible, we show the receiver can benefit from informational division. We introduce a novel *incentive-compatibility-in-the-large* approach that studies optimal design via the large-population limit. For fixed population size, optimal mechanisms are in general complex. However, we show that in the limit they converge to a simple mechanism that depends only on the payoff from acceptance, and punishes excessive consensus in the direction of the common bias. These surplus burning punishments yield payoffs bounded away from the first best; the resulting inefficiency demonstrates how our concept of informational division is distinct from standard models of information in large populations.

1 Introduction

Decision makers often rely on information that is divided across individuals with whom they have a conflict of interest. A familiar small-scale example is a CEO evaluating a new product:

*Best: Carnegie Mellon University Tepper School of Business (email: jabest@andrew.cmu.edu). Quigley: Exeter College and Department of Economics, University of Oxford (email: daniel.quigley@economics.ox.ac.uk). Saeedi: Carnegie Mellon University Tepper School of Business (email: msaeedi@andrew.cmu.edu). Shourideh: Carnegie Mellon University Tepper School of Business (email: ashourid@andrew.cmu.edu). We thank Nemanja Antic, Ian Ball, Marco Battaglini, Ben Brooks, Raphael Boleslavsky, Gorkem Celik, Archishman Chakraborty, Yeon-Koo Che, Francois Forges, Drew Fudenberg, Marina Halac, Johannes Hörner, Rick Harbaugh, Ian Jewitt, Emir Kamenica, Elliot Lipnowski, Meg Meyer, Dimitri Migrow, Paula Onuchic, Wolfgang Pesendorfer, Anna Sanktjohanser, Vasiliki Skreta, Joel Sobel, Ina Taneva, Alex Wolitzky, Andriy Zapechelnyuk, and audiences at various seminars and conferences, for their feedback.

he must elicit information divided across the marketing manager (demand) and the production manager (costs), who have different incentives from the CEO. This type of elicitation problem is increasingly important in environments where information is dispersed across large populations. For example, governments use polling to assess the effectiveness of policies that depend on local knowledge spread throughout large populations. Similarly, e-commerce and social media platforms often need to aggregate information about consumer products, political news, and opinions from a large number of individuals whose interests differ from those of the platform or society at large.

A standard solution is verification: comparing a source's report to those of others and punishing them collectively in case of disagreement. This strategy is especially effective when agents' information sets overlap to such an extent that each individual's information is largely dispensable.[1] However, in the aforementioned examples, it is not at all clear that individual's information sets do overlap sufficiently to make cross-verification effective. Indeed, when information is divided so each person has some unique piece of indispensable information, it is impossible to verify one person's report against that of another. This raises our main question: how should a decision maker elicit information when information is divided across a large population and cross-verification is impossible?

To answer this question, we study the design of information aggregation rules in a multi-sender cheap talk environment. A receiver faces a binary decision: "accept" or "reject". While the payoff to reject is known, all players' payoffs from accept increase in a stochastic state variable. There is a population of N senders, each observing a private signal about the state, and sharing a common bias toward acceptance. Signals are i.i.d. across senders and the state is the sum of the signals divided by $\sqrt{N}$. This information structure has two critical features for our analysis. First, information is unverifiable: because the signals are unconditionally independent, no sender's report can be verified against another's. Second, it captures informational division: the collective private information (as measured by the variance of the state variable) stays constant in N, while the information held by each sender shrinks as N grows. Our goal is to characterize the transfer-free recommendation mechanism that maximizes the receiver's payoff when the population is arbitrarily large.

When there is no informational division, i.e., $N = 1$, then the only informative incentive compatible mechanism is the sender-preferred mechanism which effectively delegates the decision to the sender. Since he knows the state and is the only person reporting, his report will maximize the probability of acceptance whenever he wants it, and minimize it otherwise. This has a cost to the receiver: there exists an intermediate region of states—the disagreement region—at which

[1]For instance, Krishna and Morgan (2001) show how to extract information perfectly when two senders have the same information. Gerardi et al. (2009) show a similar result with large populations.

the receiver would prefer reject but the mechanism recommends accept.

Yet, even if information is divided among only two senders, we can use one sender's report to discipline the other. An example in Section 3 illustrates this point: a *simple mechanism*—one whose recommendations depend only on the reported payoff-relevant state—can strictly improve the receiver's payoff. This mechanism does not recommend accept in the disagreement region, but has a reduced probability of acceptance for the higher values of the state. The reduced acceptance at the top acts a "punishment" that deters a sender from over-reporting out of fear that this will cause a rejection when the other sender's report is also high. While this simple mechanism improves the receiver's payoff, the optimal mechanism is more complex: it uses the full profile of reports, beyond the payoff-relevant state, to screen each sender more effectively according to his private information.

The focus of this paper is on aggregating information when it is divided among a large population. As $N \to \infty$, each sender becomes essentially uninformed and the impact of their reports becomes negligible. To characterize the asymptotics of incentive compatibility, we introduce the concept of *incentive compatibility in the large* (ICL), which extends incentive compatibility continuously to $N = \infty$. We say a mechanism is ICL if its induced distribution over outcomes at $N = \infty$ can be approximated arbitrarily well by a sequence of finite-N incentive compatible mechanisms. Our main technical result (Theorem 1) characterizes the set of ICL mechanisms in terms of two finite sets of linear constraints. The first treats each sender as if he were uninformed but still able to have a small impact on outcomes via his report; the second is a monotonicity condition that ensures the probability of acceptance is weakly increasing in reports. Together, these constraints characterize the set of mechanisms that can be approximated by incentive compatible mechanisms for large enough N. Thus, they serve as the limiting analogues of the familiar envelope formulation of incentive compatibility and interim monotonicity constraints.

Theorem 2 applies this result to show that the optimal ICL mechanism is a type of simple mechanism, which we call an *interval mechanism* (depicted in Figure 1). It recommends accept for an intermediate range of state values while rejecting at both low and extremely high values. This interval mechanism ensures incentive compatibility by trading off two scenarios in which a sender's upward lie is pivotal: at the lower threshold, lying induces acceptance when payoffs are low; at the upper threshold, lying induces rejection when payoffs are high.

That optimal mechanisms are simple arises from each sender being essentially uninformed in the limit. In particular, his signal gives him no further information about the reports of the other senders. This is true even conditional on the realized state. As a result, any optimal complex ICL mechanism can be replaced by its simple counterpart—which induces the same expected probability of acceptance conditional on the aggregate state—without harming incentives. By contrast, such a replacement does not preserve incentives when N is small, as our example in

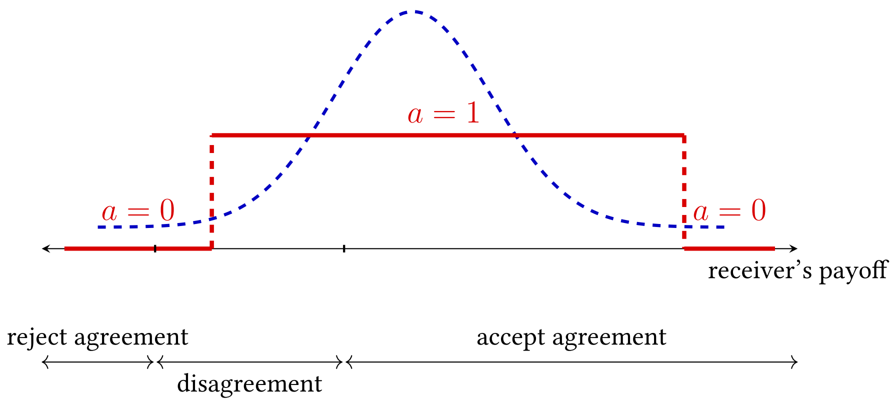

Figure 1: Optimal mechanism in the large economy. In the disagreement region, the preferred action of the senders is accept, and the receiver prefers reject. The dashed blue line represents the distribution of the receiver's payoff from accept at the infinite population limit.

Section 3 illustrates.

To understand why the interval structure is optimal among simple mechanisms, consider a perturbation of the sender-preferred mechanism which introduces some rejection at the bottom of the disagreement region. This has a large benefit for the receiver since it is where she wants to reject most within the disagreement region. Moreover, since senders are nearly indifferent at this boundary, the change minimally affects their incentives to lie. However, this perturbation still violates incentive compatibility. Perhaps surprisingly, the cheapest way to restore incentives is to punish senders by rejecting when the payoff to acceptance is highest for all players. While this punishment is very costly when it occurs, it occurs with extremely low likelihood under truthful reporting. At the same time, it has a very high relative likelihood when a sender lies. Hence, punishing at the top restores incentives for the sender at the smallest ex-ante cost to the receiver.

Nonetheless, these surplus-burning punishments imply that the receiver is strictly bounded away from her first best in the large-economy limit. Beyond the impossibility of crosschecking, this is because our model features informational division. In the rest of the literature on information aggregation—such as Feddersen and Pesendorfer (1997), Pesendorfer and Swinkels (1997), and Gerardi et al. (2009)—information is additive. There, enlarging the population adds information, so even a vanishingly small fraction of the population can eventually identify the payoff-relevant state, and the first best becomes attainable. Indeed, even though crosschecking is impossible in our setting, if information were additive, then conflicts of interest would disappear in the limit and the first best would still be achievable. However, because we divide information, the incentive problem does not disappear, optimal aggregation remains nontrivial at the limit, and we cannot achieve the first best. These considerations suggest that the concept of ICL may be applied to gain new insights by reexamining the broader information-aggregation literature through the lens of informational division, with implications for voting, auctions, markets, and collective action.

The rest of the paper is structured as follows. Section 2 introduces the model. Section 3 presents an illustrative example showing how the receiver can benefit from informational division. Section 4 develops the ICL approach, and Section 5 characterizes the optimal large-economy mechanism. Section 6 extends the analysis to heterogeneous bias. Section 7 provides an extensive discussion of the related literature, and Section 8 concludes. Proofs are in the Appendix.

2 Model

We consider a model of strategic information transmission involving multiple senders (each "he") and a single receiver ("she"). There are N senders, denoted by $i \in \{1, 2, \cdots, N\}$. Each sender privately observes a random signal s_i drawn *independently* from finite set $S = \{t_1, \cdots, t_K\} \subset \mathbb{R}$, where t_k is increasing in k, and the probability that signal $s_i = t_k$ is f_k. We denote the ordered vector of possible signal realizations by $\mathbf{t} = (t_1, \cdots, t_K) \in \mathbb{R}^K$, the probability distribution by $\mathbf{f} = (f_1, \cdots, f_K)$, and normalize the mean of s_i to zero: $\mathbb{E}[s_i] = \mathbf{f} \cdot \mathbf{t} = 0$. We denote the profile of senders' signals by $\mathbf{s} = (s_1, s_2, ..., s_N)$. Note, as signals are drawn independently, s_i contains no information about any other element of s. Hence, one sender's signal cannot be used to evaluate the truthfulness of another sender's report.[2]

The receiver chooses an action $a \in \{0, 1\}$. We refer to $a = 0$ as *reject*, and $a = 1$ as *accept*. The payoffs of the receiver and the senders are, respectively

$$u_R(a, \omega) = a(\omega + r)$$
$$u_S(a, \omega) = a(\omega + b)$$

where ω is the payoff-relevant *state*, given by

$$\omega = \sum_{i=1}^{N} \frac{s_i}{\sqrt{N}}, \tag{1}$$

and $b > 0$ and $r < b$ are parameters determining the senders' and the receiver's preference toward accept, respectively. The flexibility to vary preferences over a via b and r, allows us to treat $\mathbb{E}[s_i] = 0$ as a normalization without loss of generality. The assumption $b > r$ captures the senders' "bias" toward accept. In other words, a conflict of interest exists between the senders and the receiver for realizations of the state ω between $-b$ and $-r$, where the senders prefer accept while the receiver prefers reject. To emphasize the relationship between ω and s, we sometimes

[2] This structure of private information also satisfies the statistical properties of privacy developed in Strack and Yang (2024) and He et al. (2026). Hence, privacy of information and verifiability are intimately related.

use $\omega(\mathbf{s})$ instead of ω; we refer to the set $\{\mathbf{s} : \omega(\mathbf{s}) \in (-b, -r)\}$ as the *disagreement region.*

Our definition of ω implies that the effect of each signal s_i on the payoff-relevant state shrinks at rate $\sqrt{N}$. Hence, $s_i/\sqrt{N}$ is the analogue of sender i's private type in the mechanism design literature: we say *type* k to refer to $t_k/\sqrt{N}$. We model payoffs in this way to ensure that the variance of the aggregate private information is constant in N, while the private information held by each individual sender shrinks. For large N, this approach allows us to think of increasing N as dividing the same information among more agents. Indeed, as $N \to \infty$, we may interpret the model as one in which each sender observes a disjoint increment of a Brownian motion whose terminal value determines the payoff from accept.

We study the design of transfer-free recommendation mechanisms that maximize the receiver's payoff. By the revelation principle of Myerson (1982), it is without loss of generality to focus on direct recommendation mechanisms (henceforth, mechanism): these collect the senders' reported signals, $\tilde{\mathbf{s}} = (\tilde{s}_1, \tilde{s}_2 \cdots, \tilde{s}_N)$, and recommend an action to the receiver. Formally, a mechanism is a mapping $\sigma : S^N \to [0,1]$, where $\sigma(\mathbf{s}) = \Pr(a = 1|\mathbf{s})$ is the conditional probability of recommending $a = 1$.[3]

The revelation principle requires that the mechanism σ satisfy Bayesian incentive compatibility for the senders and obedience for the receiver:

1. *Incentive compatibility*: For each sender $i \in \{1, \cdots, N\}$, and for all $t_k, t_l \in S$,

$$\mathbb{E}\left[\sigma(\mathbf{s})(\omega(\mathbf{s}) + b) \,|s_i = t_k, \tilde{s}_i = t_k\right] \geq \mathbb{E}\left[\sigma(\tilde{\mathbf{s}})(\omega(\mathbf{s}) + b) \,|s_i = t_k, \tilde{s}_i = t_l\right]. \tag{IC}$$

2. *Obedience*:

$$\mathbb{E}\left[\sigma(\mathbf{s})(\omega(\mathbf{s}) + r)\right] \geq 0 \geq \mathbb{E}\left[(1 - \sigma(\mathbf{s}))(\omega(\mathbf{s}) + r)\right]. \tag{Ob}$$

Inequality (IC) requires that each sender prefer to truthfully report his own signal to the mediator, given that the other senders do the same and the receiver obeys the mediator's recommendation. The obedience constraint (Ob) states that the receiver should be willing to follow the recommendation of the mechanism, given that the senders are truthful.

Since we are interested in mechanisms that maximize the receiver's payoff and the receiver can take only two actions, it is sufficient to focus on incentive compatibility and ignore obedience. To see this, note that if a mechanism σ violated either of the inequalities in (Ob), then it would be dominated by the (trivially obedient) uninformative mechanism which always recommends the receiver's ex ante optimal action $a = \mathbf{1}(r \geq 0)$. We thus have the following lemma:[4]

[3] This approach contrasts with a literature on dynamic multi-sender cheap talk that assesses specific dynamic communication protocols in which communication may not be truthful. For example, see Aumann and Hart (2003), Krishna and Morgan (2004), Ambrus et al. (2013), Golosov et al. (2014), Migrow (2021), and Antic et al. (2025).

[4] The formal proof of Lemma 1 is omitted for brevity and is available on request. When there are more than two

Lemma 1. *If σ achieves the highest payoff for the receiver among all mechanisms that satisfy incentive compatibility (IC), then σ satisfies obedience (Ob).*

Lemma 1 implies that the problem of finding the receiver's best mechanism boils down to the following:

$$\max_{\sigma:S^N\to[0,1]} \mathbb{E}\left[\left(\omega\left(\mathbf{s}\right)+r\right)\sigma\left(\mathbf{s}\right)\right] \tag{P}$$

subject to incentive compatibility (IC).

Remark 1. In light of Lemma 1, the problem of finding the best mediated mechanism has an alternative interpretation. It is equivalent to a design problem in which the receiver can commit to her actions as a function of the senders' reports. Hence, the analysis extends naturally to other environments such as voting.

2.1 Simple Mechanisms

We call a mechanism *simple* if it only depends on the reported state $\tilde{\omega}$ rather than on the entire signal profile $\tilde{\mathrm{s}}$. With a small abuse of notation, we denote these mechanisms by $\sigma(\tilde{\omega})$. The following three benchmark mechanisms are all simple. First is the *receiver-preferred* mechanism, $\sigma^R\left(\tilde{\omega}\right)=\mathbf{1}\left[\tilde{\omega}+r\geq 0\right]$. This attempts to give the receiver her first best. However, it is incentive compatible only in the uninteresting case where the disagreement region is empty. Otherwise, facing σ^R, a sender with signal $s_i=t_k$ prefers to report $\tilde{s}_i=t_{k+1}$ as this lie will only induce $a=1$ for some marginal state $\omega>-b$, i.e., when he strictly prefers acceptance. Second is the *sender-preferred* mechanism, $\sigma^S\left(\tilde{\omega}\right)=\mathbf{1}\left[\tilde{\omega}+b>0\right]$, which is trivially incentive compatible as it gives the senders their preferred outcome, but is not always obedient. Third is the *uninformative* mechanism, $\sigma^U\left(\tilde{\omega}\right)=\mathbf{1}\left[r>0\right]$, which trivially satisfies incentive compatibility and obedience.

Whenever there is no informational division, i.e., when $N=1$, all mechanisms must be simple since $s_1=\omega$. Moreover, one of these three simple benchmarks solves the problem (P). In the uninteresting case where the disagreement region is empty, σ^R obviously solves (P). Outside this case, incentive compatibility severely limits what can be done. Condition (IC) reduces to

$$\sigma\left(\omega\right)\left(\omega+b\right)\geq\sigma\left(\tilde{\omega}\right)\left(\omega+b\right),\forall\omega,\tilde{\omega}\in S.$$

As the sender strictly prefers $a=1$ if and only if $\omega>-b$, he will report whatever maximizes (minimizes) σ whenever $\omega>-b$ ($\omega\leq-b$). Thus, $\sigma\left(\omega\right)$ must be constant on the sets $(-\infty,-b]$ and $(-b,\infty)$, and non-decreasing in ω. If $\mathbb{E}[\omega+r\mid\omega>-b]\geq 0>\mathbb{E}[\omega+r\mid\omega\leq-b]$, then the receiver clearly prefers σ^S among all such mechanisms. Otherwise, she must prefer σ^U. That

actions, dropping obedience may not be without loss. See Whitmeyer (2024) for a counterexample. A similar result holds in Ball (2024). We thank Ian Ball for raising this point.

is, the optimal mechanism is σ^S if and only if the information that $\omega \geq -b$ is valuable for the receiver.[5]

The reason we cannot do better with one sender is because he knows the state and is the only person reporting. This implies that to reduce the probability of accept in the disagreement region, we also have to pay the cost of equally reducing the probability of accept when the receiver agrees, i.e., wherever $\omega \geq -r$. However, when $N > 1$, each sender has fewer deviations available and less information; thus, there may be scope to use informational division to improve incentives.[6] In the next section, we demonstrate that such improvements can be achieved by simple mechanisms, though the restriction to such simple mechanisms can itself be costly.

3 An Illustrative Example

This section presents an example to develop intuition for a) how we can improve the receiver's payoff with a simple mechanism when information is divided across multiple strategic senders; and b) why the optimal mechanism may be complex, i.e., depend on the whole vector s beyond $\omega(\mathbf{s})$. We provide supporting calculations in Online Appendix B.

Let $S = \left\{-\sqrt{2}, 0, \sqrt{2}\right\}$ where $\mathbf{f}$ is uniform on S and $N = 2$. Hence, each sender has type $s_i/\sqrt{2}$, which we denote by $l = -1$ (low), $m = 0$ (medium), and $h = 1$ (high), and the state is

$$\omega(\mathbf{s}) = \frac{s_1 + s_2}{\sqrt{2}} \in \{-2, -1, 0, 1, 2\}.$$

Finally, we let $r = 0$ and $b = 3$ so that $u_R(a, \omega) = a\omega$ and $u_S(a, \omega) = a(\omega + 3)$.

In this example, the sender preferred mechanism σ^S is uninformative, since $\omega + b > 0$ for all ω. Hence, if a *single* sender observed both s_1 and s_2 (equivalently, observed ω perfectly), then the only truthful and obedient mechanism would be $\sigma^S = \sigma^U$ with expected payoff of 0 for the receiver. However, when the information is divided between two senders, the following truthful and obedient simple mechanism (also illustrated in Figure 2a) increases receiver payoffs:

$$\sigma^\dagger(\omega) = \begin{cases} 0 & \text{if } \omega \in \{-2, -1\} \\ 1 & \text{if } \omega \in \{0\} \\ \frac{1}{3} & \text{if } \omega \in \{1, 2\} \end{cases}. \tag{2}$$

[5] While this is the case in our environment, there is a literature on single sender cheap talk (Crawford and Sobel, 1982) showing conditions under which a mediator can improve payoffs. See for instance, Goltsman et al. (2009), Salamanca (2021), Corrao and Dai (2023), and Best and Quigley (2024).

[6] In the online appendix we also provide results on the optimality of the sender preferred allocation with a rich (continuous) signal space for $N > 1$. If the bias is not too large and the senders are not too many, then it is optimal.

Obviously, this is a strict improvement on no information, giving a receiver payoff of $\frac{4}{27} > 0$. This mechanism takes advantage of the fact that when action $a = 1$ is most costly to the receiver, i.e., when $\omega \in \{-2, -1\}$, it is also least desirable to the sender. As a result, a surplus-burning reduction in the probability of $a = 1$ at the top of the distribution—where the sender values acceptance the most—acts as a sufficiently large punishment to deter upwards lies.

To see this, notice that an upward lie by a sender (l reporting m, or m reporting h) shifts the reported state from ω to $\omega + 1$. Hence, sender i's lie is *pivotal* only when $\sigma^\dagger(\omega) \neq \sigma^\dagger(\omega + 1)$. In this example, this occurs in two states, $\omega \in \{-1, 0\}$. When the true state is $\omega = -1$ the lie benefits the sender by increasing the probability of acceptance; but when $\omega = 0$, the lie decreases its probability. The expected benefit of the lie is $\Pr(\omega = -1 \mid s_i)[\sigma^\dagger(0) - \sigma^\dagger(-1)](-1 + b) = 2/3$, which is offset by the expected cost $\Pr(\omega = 0 \mid s_i)[\sigma^\dagger(1) - \sigma^\dagger(0)](0 + b) = -2/3$, so the mechanism is incentive compatible.[7]

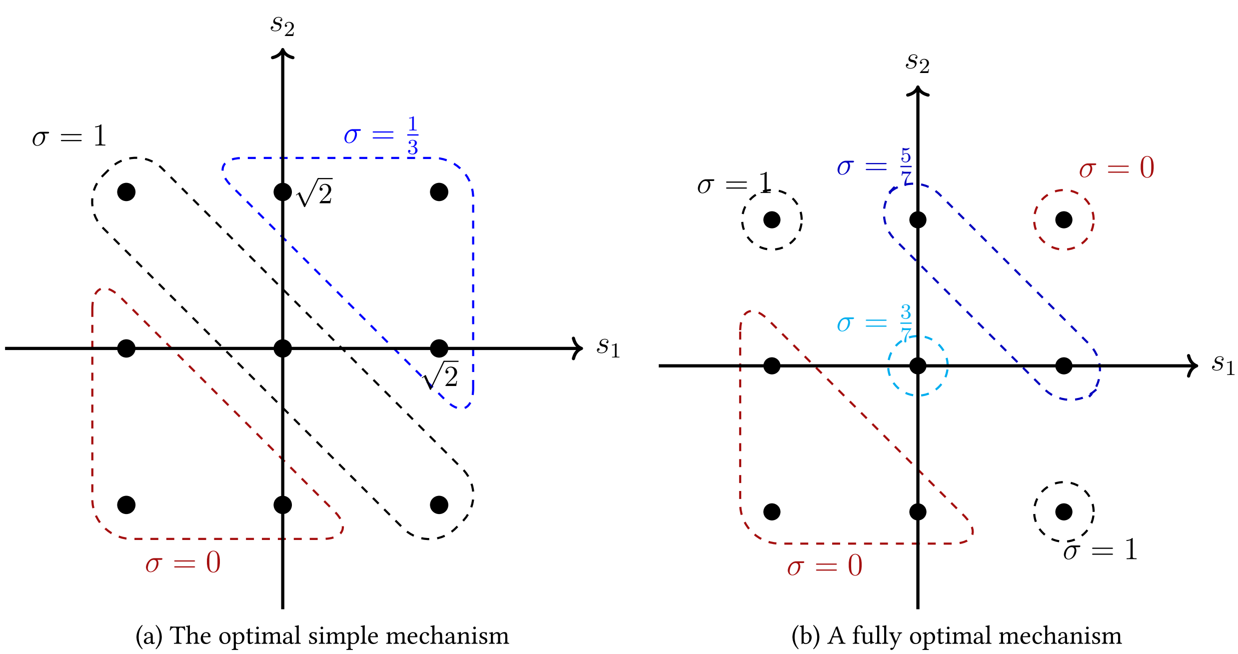

Figure 2: Illustration of mechanisms: Figure 2a shows the best incentive compatible simple mechanism for the receiver in this environment; Figure 2b presents a mechanism which is optimal in the class of all truthful direct mechanisms. No simple mechanism achieves the fully optimal payoff associated with the mechanism in Figure 2b.

This mechanism is optimal among all simple mechanisms and illustrates how informational division allows us to improve payoffs by reducing the set of deviations available to each sender.[8]

[7]Notice, this binding local incentive constraint is identical for both types l and m because $\mathbf{f}$ is uniform in this example. Moreover, the interim probability of acceptance is increasing in the sender's report $\tilde{s}_i$. Hence, by standard arguments these local incentive constraints imply global incentive compatibility.

[8]Notice from Figure 2a, when $N = 2$, each sender only chooses a row (or column) of acceptance probabilities

However, the restriction to simple mechanisms is costly. In particular, Figure 2b illustrates the optimal mechanism $\sigma^\star(\mathbf{s})$ which achieves a strictly greater payoff of $\frac{10}{63} > \frac{4}{27}$ for the receiver. Clearly, $\sigma^\star(\mathbf{s})$ is not simple: it provides a lower acceptance probability for the report (m, m) than for (l, h) and (h, l) despite all corresponding to $\omega = 0$. Relative to $\sigma^\dagger(\omega)$, reducing $\sigma(m, m)$ costs the receiver nothing—she is indifferent over a when $\omega = 0$. However, it serves as a punishment targeted at an m-report that strictly deters l from an upward lie. This allows us to increase the probability of $a = 1$ when $\omega = 1$ while still respecting l's incentives. Finally, while these changes do make an upward lie more appealing for the m-types, the mechanism recovers incentive compatibility with a punishment targeted at the top, $\tilde{\mathbf{s}} = (h, h)$, where the sender values $a = 1$ most.

This example makes clear why restricting to simple mechanisms can be costly: they cannot independently assign acceptance probabilities across reports. In particular, a simple mechanism could only reduce σ at (m, m) if it also made identical reductions at (l, h) and (h, l). While this would reduce the payoff from reporting m, it would for l and h too. In this example, such a reduction would actually encourage the l type to lie because it would reduce his truth-telling payoff by more than his payoff from masquerading as m. By contrast, $\sigma^\star(\mathbf{s})$ used the freedom to independently reduce the payoffs for higher reports (i.e., without also reducing the truth-telling payoffs for the lower types) to more effectively screen senders according to their type.

The optimality of complex mechanisms is not unique to this example. Indeed, in richer type spaces, effective screening of types can require mechanisms to be even more complex.[9] However, in the next section, we show that the importance of each sender's type becomes negligible as $N \to \infty$, which simplifies the characterization of incentive constraints. Then in Section 5, we use this result to show that in the large-N limit, the optimal mechanism is in fact simple. Moreover, this simple mechanism shares two important properties with $\sigma^\dagger(\omega)$ above: First, it uses surplus-burning punishments for high aggregate reports—where the sender wants $a = 1$ most—to buy much greater reductions in the probability of acceptance in the disagreement region. Second, an upward lie is pivotal only in two states: increasing the probability of accept in the lower state and decreasing it in the higher state.

4 Incentive Compatibility in Large Economies

The focus of this paper is on aggregating information when it is divided among a large population. To examine this, we study the solution to problem (P) as $N \to \infty$. To make this problem amenable

indexed by the other sender's type. By contrast, if a single sender could observe and report (s_1, s_2), he can select both the row and the column to secure acceptance, e.g., by reporting (m, m).

[9]See Online Appendix E for a numerical example.

to limiting arguments, we first reformulate it as a choice over mechanisms that map frequencies of signal realizations into the probability of recommending accept. This allows us to establish our main technical result (Theorem 1), which characterizes the limit set of incentive compatible mechanisms. Theorem 1 shows that the right way to extend incentives to the limit, is to treat each sender as if he is uninformed but can nevertheless have a small impact on the distribution of reports.

In section 5, we characterize the optimal mechanism at infinity subject to this characterization of incentive compatibility in the limit (Theorem 2). As this mechanism is the limit point to which all optimal mechanisms converge, it characterizes the asymptotic properties of optimal aggregation when information is divided across large finite populations.

4.1 Mechanisms in the Frequency Domain

Recall a direct mechanism is a mapping $\sigma : S^N \to [0,1]$. Hence, the domain of direct mechanisms changes in N and its dimensionality explodes as $N \to \infty$. To avoid these issues, we reframe the problem in terms of mechanisms that operate on the frequency domain—which is invariant in N. Formally, we define the normalized deviation of the sample frequency of signal t_k from its population frequency by

$$h_k(\mathbf{s}) = \sqrt{N}\left(\frac{|\{i|s_i = t_k\}|}{N} - f_k\right), \tag{3}$$

and the vector of these deviations by $\mathbf{h}^N(\mathbf{s}) = (h_1(\mathbf{s}), \cdots, h_K(\mathbf{s}))$. We refer to $\mathbf{h}^N(\mathbf{s})$ as *normalized empirical frequencies (NEF),* and to h_k as k-NEF. From here on, we consider mechanisms that are functions of the empirical frequencies $\mathbf{h}^N(\mathbf{s})$, i.e., $\sigma : \mathbb{R}^K_0 \to [0,1]$, where $\mathbb{R}^K_0$ is the subset of $\mathbb{R}^K$ whose elements sum to zero (because $\sum_{k=1}^K h_k^N(\mathbf{s}) = 0$). With a slight abuse of notation, we denote this function by $\sigma(\mathbf{h}^N)$ and suppress the dependence of $\mathbf{h}^N$ on $\mathbf{s}$ hereafter.[10]

It is without loss of generality to focus on mechanisms expressed on the $K-1$-dimensional frequency domain. To see this, consider that it is without loss to focus on symmetric recommendation mechanisms, for which $\sigma(\mathbf{s}) = \sigma(\pi(\mathbf{s}))$ for all $\pi \in \Pi$, where Π is the set of all permutations of $\{1, \cdots, N\}$. This is because, for any incentive compatible mechanism σ, the mechanism given by $\hat{\sigma}(\mathbf{s}) = \sum_{\pi\in\Pi}\sigma(\pi(\mathbf{s}))/N!$ is incentive compatible, symmetric, and delivers the same payoff to the receiver. Finally, notice that any symmetric mechanism is equivalent to one that depends on a simple count of each report, and the NEF maps one-to-one onto these counts.

[10]Notice, the object $\mathbf{h}^N(\mathbf{s})$ has a familiar interpretation. Each h_k is very similar to the classical test statistic (a "Z-test") for the population proportion of type k. Thus, one could also view the design of σ through the lens of hypothesis testing. Our approach is also closely related to the 'method of types' used in information theory; see, for instance, Cover and Thomas (2006). We use different language to avoid the obvious confusion.

As it is without loss to analyze problem (P) using mechanisms on the frequency domain, we use the fact that

$$\omega = \sum \frac{s_i}{\sqrt{N}} = \mathbf{h}^N \cdot \mathbf{t}.$$

to express incentive compatibility directly in terms of $\mathbf{h}^N$, as

$$\mathbb{E}\left[\sigma\left(\mathbf{h}^N\right)\left(\mathbf{h}^N \cdot \mathbf{t} + b\right) | s_i = t_k, \tilde{s}_i = t_k\right] \geq \mathbb{E}\left[\sigma\left(\tilde{\mathbf{h}}^N\right)\left(\mathbf{h}^N \cdot \mathbf{t} + b\right) | s_i = t_k, \tilde{s}_i = t_l\right], \quad \text{(F-IC)}$$

for all $t_k, t_l \in S$, where $\tilde{\mathbf{h}}^N$ is the reported NEF. Notice how the sender's expected payoffs depend on his signal and report. First, since his signal enters directly into the true NEF, it gives him private information about $\mathbf{h}^N$ and therefore the value of acceptance. Second, since his report enters directly into the reported NEF, $\tilde{\mathbf{h}}^N$, he can influence its distribution: if a k type lies and reports t_l, then he decreases $\tilde{h}_k^N$ and increases $\tilde{h}_l^N$. Of course, this is not profitable in an incentive compatible mechanism.

A benefit of reframing mechanisms in this way is that it ensures the domain of the mechanism is invariant in N. Moreover, the multidimensional central limit theorem implies the argument of the mechanism, the NEF, has appealing large sample properties:

Lemma 2. ***(Multidimensional Central Limit Theorem)*** *Let* $\mathbf{h}^N(\mathbf{s})$ *be the normalized empirical frequencies of* $\mathbf{s}$ *as defined in (3). Then, as* $N \to \infty$,

$$\mathbf{h}^N \xrightarrow{d} N(\mathbf{0}, \mathbf{\Sigma}), \tag{4}$$

with $\mathbf{\Sigma}_{kl} = f_k(\mathbf{1}(k = l) - f_l)$. *Furthermore,* $\omega \xrightarrow{d} N(0, \mathit{Var}(s_i))$.

Since the signals are independent, it is not surprising that a central limit theorem applies. However, given this independence, it may be surprising that there is (negative) correlation between the k-NEF and l-NEF ($\mathbf{\Sigma}_{kl} < 0$ for $k \neq l$). Yet, as in the discussion following condition (F-IC), more of one signal must result in less of another. Indeed, this correlation reflects both the sender's private information about $\mathbf{h}^N$, and the influence of his reports on $\tilde{\mathbf{h}}^N$. In the next section, we use Lemma 2 to analyze how incentive compatibility changes as the population becomes large and solve for the optimal mechanism in the limit.

Finally, in the N-sender problem, notice that a mechanism σ^N and the random variable $\mathbf{h}^N$ jointly induce a joint distribution over $(\mathbf{h}^N, a^N)$, where a^N is the recommended action and $\Pr\left[a^N = 1 \mid \mathbf{h}^N\right] = \sigma^N(\mathbf{h}^N)$; we refer to the joint distribution induced by the N-sender mechanism σ^N as its *outcome distribution.* Similarly, we define the outcome distribution of a mechanism σ in the large-N limit as the joint distribution over $(\mathbf{h}, a)$ induced by $\sigma(\mathbf{h}) = \Pr[a = 1 \mid \mathbf{h}]$ and

$\mathbf{h} \sim N(\mathbf{0}, \mathbf{\Sigma})$. For any incentive compatible direct mechanism, all payoff- and incentive-relevant information is captured by its outcome distribution; in particular, the outcome distribution determines all the expected values present in both the objective and the constraints of problem (P).

4.2 Incentive Compatibility in the Large

We are interested in the limiting properties of optimal mechanisms as N grows large. To characterize this limit, we study a design problem in which feasible mechanisms must satisfy a notion of *incentive compatibility in the large.*

Definition 1. A recommendation mechanism $\sigma : \mathbb{R}_0^K \to [0, 1]$ is *incentive compatible in the large (ICL)* if there exists a sequence of N-sender recommendation mechanisms $\{\sigma^N\}_{N\geq 1}$ satisfying the finite-N incentive compatibility condition (F-IC) whose induced outcome distributions converge

$$\left(\mathbf{h}^N, a^N\right) \xrightarrow{d} (\mathbf{h}, a) \, .$$

Thus, a mechanism σ is ICL if its outcome distributions can be approximated arbitrarily well by incentive compatible mechanisms σ^N in large, finite economies. Though the large-N limit in which $\mathbf{h} \sim N(\mathbf{0}, \mathbf{\Sigma})$ is a limiting construct (it does not correspond to any finite N-sender economy), understanding the set of ICL mechanisms is useful because it approximates what can be achieved when information is divided among sufficiently many senders. The following theorem provides a simple characterization of this set:

Theorem 1. *A recommendation mechanism $\sigma : \mathbb{R}_0^K \to [0, 1]$ is ICL if and only if it satisfies*

$$\mathbb{E}\left[\sigma\left(\mathbf{h}\right)\left(\mathbf{h}\cdot\mathbf{t}+b\right)h_k/f_k\right] = t_k\mathbb{E}[\sigma\left(\mathbf{h}\right)] \ \textit{for all } k, \tag{ENV}$$

$$\mathbb{E}\left[\sigma\left(\mathbf{h}\right)\left(\frac{h_k}{f_k}-\frac{h_{k-1}}{f_{k-1}}\right)\right] \geq 0 \ \textit{for all } k \geq 2, \tag{MON}$$

with $\mathbf{h} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}\right)$.

Theorem 1 is our main technical result and it will greatly simplify the analysis of optimal mechanisms in large economies. In essence, it proves that conditions (ENV) and (MON) are the limiting analogues of the envelope formulation of incentive compatibility and the interim monotonicity constraints common in the mechanism design literature. As such, they fully characterize the set of ICL mechanisms, i.e., those whose outcome distributions can be approximated by an incentive compatible mechanism for large enough N.

For some intuition, consider how a sender's incentives vary with N. Incentive compatibility requires that type k does not want to report t_l. Defining the truthful utility of type k as $U_k^N = \mathbb{E}\left[(\mathbf{h} \cdot \mathbf{t} + b)\,\sigma^N(\mathbf{h}) \mid s_i = \tilde{s}_i = t_k\right]$, this condition can be written as

$$U_k^N - U_l^N \geq \frac{t_k - t_l}{\sqrt{N}} \mathbb{E}\left[\sigma^N\left(\mathbf{h}^N\right) \mid \tilde{s}_i = t_l\right]. \tag{5}$$

Inequality (5) is the standard incentive constraint that arises in mechanism design problems where senders' types are independent. To fix ideas, let $t_l < t_k$. If type k reports t_l, he can induce the allocation offered to type l. However, since type k enjoys a higher private value for $a = 1$, inducing such an allocation would earn him an incremental information rent over type l's payoff, given by the right side of (5). In an incentive compatible mechanism, this payoff must be no more than type k's truthful utility U_k^N.

Since type l must not want to report t_k either, incentive compatibility implies:

$$\mathbb{E}\left[\sigma^N\left(\mathbf{h}^N\right) | \tilde{s}_i = t_k\right] \geq \sqrt{N} \frac{U_k^N - U_l^N}{t_k - t_l} \geq \mathbb{E}\left[\sigma^N\left(\mathbf{h}^N\right) \mid \tilde{s}_i = t_l\right]. \tag{N-ENV}$$

Notice that for adjacent types $k \geq 2$ and $l = k - 1$, this is the discrete version of the standard envelope condition: it bounds the *rate* at which the sender's interim utility increases in his type ($t_k/\sqrt{N}$). Moreover, (N-ENV) also implies that the usual monotonicity condition on the interim allocation, $\mathbb{E}\left[\sigma^N\left(\mathbf{h}^N\right) \mid \tilde{s}_i = t_k\right]$, holds. By the standard Spence-Mirrlees arguments, these envelope and monotonicity conditions are also sufficient for global incentive compatibility.[11]

To see how condition (N-ENV) behaves in the limit, we must establish how the sender's private information about $\mathbf{h}^N$ and the influence of his report on $\tilde{\mathbf{h}}^N$ depend on N. To do this, we update beliefs about $\mathbf{h}^N$ and $\tilde{\mathbf{h}}^N$ via Bayes' rule. For $\mathbf{h}^N$, this is[12]

$$\Pr\left(\mathbf{h}^N | s_i = t_l\right) = \Pr\left(\mathbf{h}^N\right) \frac{|\{i : s_i = t_l\}|}{f_l N} = \Pr\left(\mathbf{h}^N\right)\left(\frac{h_l^N}{\sqrt{N} f_l} + 1\right). \tag{6}$$

An identical expression holds for $\Pr\left(\tilde{\mathbf{h}}^N | \tilde{s}_i = t_l\right)$. Equation (6) is a reflection of the sender's private information, which diminishes in the number of senders N. Indeed, as $N \to \infty$ the sender's posterior collapses to the prior and he becomes essentially uninformed.

Using (6) and $U_k^N = \mathbb{E}\left[(\mathbf{h} \cdot \mathbf{t} + b)\,\sigma(\mathbf{h}) \mid s_i = \tilde{s}_i = t_k\right]$, the components of condition (N-ENV)

[11] Notice how this compares to mechanism design problems with transfers. There, every nondecreasing allocation would be incentive compatible, since transfers can always be constructed to appropriately bound the marginal utilities. In our setting, we do not have the additional tool of transfers to ensure that condition (N-ENV) is satisfied by every monotone mechanism; accordingly, the bounds (N-ENV) on marginal utilities impose additional restrictions on σ beyond those imposed by monotonicity.

[12] One can verify (6) after noting that (i) the population frequency of type t_l is f_l, and (ii) the realized sample frequency $\Pr\left(s_i = t_l \mid \mathbf{h}^N\right)$ is $|\{i : s_i = t_l\}|/N$. The prior $\Pr\left(\mathbf{h}^N(\tilde{\mathbf{s}})\right)$ is updated by their ratio.

can be expressed:

$$\mathbb{E}\left[\sigma^N(\mathbf{h}^N) \mid \tilde{s}_i = t_k\right] = \mathbb{E}\left[\sigma^N(\mathbf{h}^N)\left(1 + \frac{h_k^N}{f_k\sqrt{N}}\right)\right], \tag{7}$$

$$\sqrt{N}\frac{U_k^N - U_l^N}{t_k - t_l} = \frac{1}{t_k - t_l}\mathbb{E}\left[\left(\mathbf{h}^N \cdot \mathbf{t} + b\right)\sigma^N(\mathbf{h^N})\left(\frac{h_k^N}{f_k} - \frac{h_l^N}{f_l}\right)\right]. \tag{8}$$

We show in the proof that these expectations also converge appropriately, so that as $N \to \infty$ (N-ENV) becomes

$$\mathbb{E}\left[\left(\mathbf{h} \cdot \mathbf{t} + b\right)\sigma(\mathbf{h})\left(\frac{h_k}{f_k} - \frac{h_l}{f_l}\right)\right] = (t_k - t_l)\,\mathbb{E}\left[\sigma(\mathbf{h})\right]. \tag{9}$$

Intuitively, as N gets large and the sender's information shrinks, the impact of his report on the probability of acceptance becomes negligible (observe $1 + \frac{h_k^N}{f_k\sqrt{N}} \to 1$), so that (7) converges to $\mathbb{E}\left[\sigma(\mathbf{h})\right]$. Nonetheless, we see in (8) that the sender's utility does vary with his type $\frac{s_i}{\sqrt{N}}$ for all N: his report always has a *first-order* impact $\frac{h_k^N}{f_k} - \frac{h_l^N}{f_l}$ on the NEF, and so (8) becomes $\frac{1}{t_k - t_l}\mathbb{E}\left[\left(\mathbf{h} \cdot \mathbf{t} + b\right)\sigma(\mathbf{h})\left(\frac{h_k}{f_k} - \frac{h_l}{f_l}\right)\right]$ in the limit.[13] Hence, even as N becomes large, the sender's type continues to have a first-order impact on payoffs.

Condition (9) is equivalent to (ENV). Notice, as incentive compatibility holds for all l, (9) must also hold when taking expectations with respect to l (and fixing k). Moreover, since $\mathbb{E}[s_i] = 0$ and $\sum h_l = 0$, the terms in l drop out: equation (ENV) therefore expresses the constraint in terms of deviations of the sender's payoff from average. On the other hand, taking the difference of equation (ENV) for k and l yields (9).

Equation (MON) is motivated by similar (albeit simpler) reasoning. For finite N, it is easy to see from (7) that the monotonicity requirement is simply $\mathbb{E}\left[\sigma^N(\mathbf{h})\left(\frac{h_k^N}{f_k\sqrt{N}} - \frac{h_{k-1}^N}{f_{k-1}\sqrt{N}}\right)\right] \geq 0$. After multiplying by $\sqrt{N}$, we show the resulting expression converges to (MON) as $N \to \infty$. Hence, (ENV) and (MON) can be viewed as the limiting analogues of the envelope formulation of incentive compatibility and the interim monotonicity constraints.

The necessity of (ENV) and (MON) for a mechanism to be ICL follows almost immediately from the discussion above. Establishing sufficiency is more challenging. Though it is not difficult to find a sequence of finite-N mechanisms σ^N whose outcomes $(\mathbf{h}^N, a^N)$ converge in distribution to $(\mathbf{h}, a)$, we must also ensure it satisfies incentive compatibility condition (F-IC) for all N. The proof shows how to adjust any such sequence to ensure both that (F-IC) is respected and that the new sequence of induced outcomes still converges appropriately. The key observation is that, since (ENV) and (MON) are limiting analogues of (F-IC), the required adjustments to ensure

[13] One can see this by examining how the rate at which $\Pr\left(\mathbf{h}^N\left(\mathbf{s}\right)|s_i\right)$ changes in type $\frac{s_i}{\sqrt{N}}$, using (6).

incentive compatibility become negligible as $N \to \infty$, and hence so do their impact on outcomes. This ensures that the outcomes of the adjusted sequence of mechanisms also converge in distribution to $(\mathbf{h}, a)$.

Remark 2. Because types converge, $s_i/\sqrt{N} \to 0$, each sender's preference over allocations is independent of their type in the limit. Thus, one can view (ENV) as reflecting the incentive compatibility of an essentially uninformed sender who can nonetheless have a small, first-order impact on outcomes via his report: we describe a deviation by a sender whose misreport shifts the k-NEF h_k by an incremental amount dh_k as a '*nudge*' in 'direction' k. Indeed, (ENV) is equivalent to the first-order condition of an uninformed sender with the ability to nudge the distribution of the NEF $\mathbf{h}$ in 'direction' k; when this condition holds, the sender has no incentive to make such a nudge.[14] Under this interpretation, (MON) is a condition under which mechanisms obeying (ENV) are robust to the introduction of infinitesimal amounts of private information.

The convergence in types also makes clear why it was important for our analysis to frame incentive constraints in terms of rates of change. Had we expressed condition (N-ENV) in terms of differences $U_k^N - U_l^N$, the limit would yield $\lim U_k^N = \lim U_l^N$. But this trivially applies to *all* mechanisms at $N = \infty$ since the sender's types collapse to 0, and therefore $\lim U_k^N = \lim U_l^N$ would impose no discipline on incentives at all. For instance, this limit would permit the receiver-preferred allocation, yet obviously this mechanism is not incentive compatible for any finite N. By contrast, framing the constraint on the margins imposes meaningful restrictions. For instance, it is not hard to see that the receiver-preferred mechanism fails (ENV). While the first-best solution is therefore not feasible, Theorem 1 gives us the tools to solve the second-best problem (P) in the large.

5 Optimality in Large Economies

Theorem 1 implies that we can characterize the optimal mechanism in the limiting economy as $N \to \infty$ by solving the much simplified problem

$$\max_{\sigma:\mathbb{R}_0^K \to [0,1]} \mathbb{E}\left[\sigma(\mathbf{h})(\mathbf{h}\cdot\mathbf{t}+r)\right] \tag{P1}$$

subject to

$$\mathbb{E}\left[\sigma(\mathbf{h})\left((\mathbf{h}\cdot\mathbf{t}+b)\,h_k/f_k - t_k\right)\right] = 0, \mathbb{E}\left[\sigma(\mathbf{h})\frac{h_k}{f_k}\right] \geq \mathbb{E}\left[\sigma(\mathbf{h})\frac{h_{k-1}}{f_{k-1}}\right], \tag{10}$$

[14] We provide a formal version of this uninformed-sender interpretation of the problem (and its associated first-order conditions) in Online Appendix C.1.

where $\mathbf{h} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$.

The following theorem contains our main result on optimal mechanisms for large economies:

Theorem 2. *The optimum in (P1) is achieved by a recommendation mechanism that is a function of only the sample mean, $\omega = \mathbf{h} \cdot \mathbf{t}$, and satisfies*

$$\sigma^* (\mathbf{h}) = \begin{cases} 1 & \omega = \mathbf{h} \cdot \mathbf{t} \in [\underline{\omega}, \overline{\omega}] \,, \\ 0 & \omega \notin [\underline{\omega}, \overline{\omega}] \,. \end{cases}$$

Furthermore, when $r \in \left(\frac{b - \sqrt{b^2+4}}{2}, b \right)$, then the cutoffs $\underline{\omega}, \overline{\omega}$ satisfy $\underline{\omega} \in (-b, -r)$, $\overline{\omega} \in (-r, \infty)$. If $r < \frac{b-\sqrt{b^2+4}}{2}$, then $\underline{\omega} = \overline{\omega}$ and thus $\sigma^* (\mathbf{h}) = 0$ almost surely.

Moreover, σ^* is the unique recommendation mechanism (up to measure zero changes) that achieves the optimum in (P1).

Theorem 2 shows that a simple mechanism is optimal. It recommends accept when the aggregate reported state lies within a bounded interval. We call such mechanisms *interval mechanisms*. Figure 1 in the introduction depicts its key features: relative to the sender-preferred allocation $\sigma^S(\omega) = \mathbf{1}[\omega > -b]$, the mechanism (i) increases the threshold at which acceptance begins; and (ii) introduces a surplus-burning region of rejection beyond the second, higher threshold, $\overline{\omega}$. Despite all parties wanting accept when $\omega > \overline{\omega}$, this surplus burning at the top acts as a punishment for over-reporting that allows us to take the receiver's preferred action, reject, when $\omega \in [-b, \underline{\omega})$. Still, the receiver's payoff from σ^* is clearly bound away from the first-best payoff induced by σ^R. In addition, Theorem 2 establishes that when r is negative, the bias $b - r$ may be so large that the uninformative mechanism is optimal.

Next we provide intuition for the optimality of the interval mechanism in three parts: why it is incentive compatible; why it is optimal among simple mechanisms; and why it is without loss to restrict attention to simple mechanisms in the large. We conclude the section by showing that the essential uniqueness of the solution to problem (P1) implies that the optimal mechanisms for large finite populations converge to σ^* in terms of their outcome distributions.[15]

5.1 Incentive Compatibility and Pivotality

The structure of incentive compatibility for the interval mechanism σ^* mirrors the logic of the optimal simple mechanism in Section 3. Recall that in the limit, we can view a misreport by the essentially uninformed sender as a nudge of the reported k-NEF by an incremental amount dh_k in

[15] Notice, while convergence in outcome distributions does not require that the $\{\sigma^N\}$ converge pointwise to σ^*, it does imply that the conditional expectation of σ^N converges to that of σ on every hypercube in $\mathbb{R}_0^K$—including all arbitrarily small ones.

direction k (see Remark 2). However, in a simple mechanism, this nudge can only affect outcomes via its first-order impact on the reported state. That is, an upward lie nudges the reported state by some small increment $d\tilde{\omega} > 0$. Under σ^*, this nudge is pivotal in only two states: the lower and upper thresholds $\underline{\omega}$ and $\bar{\omega}$. At $\omega = \underline{\omega}$, such a lie benefits the sender because it nudges the reported state into the acceptance region. By contrast, at $\omega = \overline{\omega}$, the same lie induces the punishment of rejection. This reduces the incentive compatibility condition in (ENV) to a simple condition

$$\phi(\underline{\omega})(\underline{\omega} + b) = \phi(\bar{\omega})(\bar{\omega} + b), \tag{11}$$

where ϕ is just the standard normal density, and we have normalized $\text{var}(\omega) = 1$ for notational simplicity. As in the example, the sender values the action more at $\bar{\omega}$ than at $\underline{\omega}$ but the relative likelihood of $\bar{\omega}$ is lower, ensuring that when the sender is pivotal, the expected cost of the lie equals the benefit.

While this pivotality argument is intuitive, formally it follows from equation (ENV) applied to simple mechanisms:[16]

$$-\int_{-\infty}^{\infty} (\omega + b)\, \sigma\,(\omega)\, \phi'\,(\omega)\, d\omega - \mathbb{E}\left[\sigma\,(\omega)\right] = 0. \tag{ω-ENV}$$

In light of Remark (2), note that equation (ω-ENV) is the first-order condition ensuring that the uninformed sender does not wish to nudge the distribution of the reported state $\tilde{\omega}$ upward. The first term represents the sender's utility from a true upward shift in the distribution of ω, and the second term the marginal information rent.[17] Obviously, when the difference is zero, the sender has no incentive to misreport. We can derive the pivotal incentive compatibility condition (11) from (ω-ENV) using integration by parts and the fact that $\sigma(\omega)$ is constant on $[\underline{\omega}, \bar{\omega}]$.

Condition (11) also provides a simple intuition for why σ^* satisfies the monotonicity condition (MON): since the first-order effect of an increase in the sender's type is to shift the distribution of ω rightward by $d\omega$, $\phi(\underline{\omega}) \geq \phi(\bar{\omega})$ implies that the probability of acceptance weakly increases in type (to the first-order). When σ^* is informative, $\phi(\underline{\omega}) > \phi(\bar{\omega})$ and (MON) is in fact slack; otherwise, if σ^* is uninformative ($\underline{\omega} = \overline{\omega}$), then $\phi(\underline{\omega}) = \phi(\overline{\omega})$ and (MON) is binding.

[16]To see how (ω-ENV) follows from (ENV), note that $\mathbb{E}\left[(\omega + b)\sigma(\omega)\frac{h_k}{f_k}\right] = \mathbb{E}\left[(\omega + b)\sigma(\omega)\mathbb{E}\left[\frac{h_k}{f_k} \mid \omega\right]\right]$ by the law of iterated expectations. Moreover, by Lemma 2 h_k and ω are jointly Normal with $\text{cov}(h_k, \omega) = \sum t_l \text{cov}(h_k, h_l) = t_k f_k$. Hence, Normal updating gives $\mathbb{E}\left[\frac{h_k}{f_k} \mid \omega\right] = t_k \frac{\omega}{\text{var}(\omega)}$. Plugging in to (ENV), noting that the standard Normal density satisfies $\phi'(\frac{\omega}{\text{var}(\omega)}) = -\frac{\omega}{\text{var}(\omega)}\phi(\frac{\omega}{\text{var}(\omega)})$, and recalling the normalization $\text{var}(\omega) = 1$, establishes the claim.

[17]Formally, the first term is the total derivative of utility with respect to ω. The second is the partial derivative of utility with respect ω, i.e., holding the reported distribution constant. Their difference isolates the partial derivative of utility with respect to the reported state $\tilde{\omega}$.

5.2 The Value of Punishing at the Top

Unlike the example in Section 3, it is not immediately obvious that this interval mechanism is an improvement on the sender-preferred allocation σ^S. The total cost of the surplus-burning region could outweigh the benefit of reducing acceptance in the disagreement region. Indeed, there are many interval mechanisms whose thresholds satisfy incentive compatibility and yet reduce payoffs.[18] Moreover, it is even less obvious that interval mechanisms should be optimal among all simple mechanisms in the large. Fortunately, clear intuitions for both can be found in the simple economics of problem (P1).

To explain why the mechanism described in Theorem 2 improves on the sender-preferred mechanism σ^S, consider an incentive compatible perturbation from σ^S to a 'nearby' interval mechanism $\sigma' = \mathbf{1}[\underline{\omega}', \overline{\omega}']$ whose thresholds satisfy $\underline{\omega}' = -b + \varepsilon$ and $\overline{\omega}' < \infty$, where $\varepsilon > 0$ is chosen arbitrarily small and $\overline{\omega}'$ is correspondingly large. Consider first the effects of increasing the lower threshold from $-b$ to $\underline{\omega}'$. Doing so clearly benefits the receiver: she is spared a cost of approximately $b - r$ when $\omega \in [-b, \underline{\omega}']$, an event which occurs with probability close to $\phi(-b)\varepsilon$. However, this change in isolation also violates incentive compatibility, because a lie would now be pivotal only when the sender strictly prefers $a = 1$.

Indeed, increasing the lower threshold increases the left side of the incentive constraint (ω-ENV) by approximately

$$((\omega + b)\phi'(\omega) + \phi(\omega))\,\varepsilon, \tag{12}$$

evaluated at $\omega = \underline{\omega}'$. This expression reflects the effect of the change on the sender's payoff from an upward lie of magnitude $d\tilde{\omega} > 0$ relative to truth-telling. To see this, notice that it reduces the sender's payoffs from both truth-telling and lying. Under truth-telling, it reduces the sender's payoff by approximately $(\underline{\omega}' + b)\phi(\underline{\omega}')\varepsilon$. On the other hand, the change reduces the payoffs from an upwards lie by $(\underline{\omega}' - d\tilde{\omega} + b)\phi(\underline{\omega}' - d\tilde{\omega})\varepsilon$ because the true state at which $\tilde{\omega} = \underline{\omega}'$ is now $\underline{\omega}' - d\tilde{\omega}$. Notice, the reduction in σ affects the net value of an upward lie in two ways: it reduces the value of the state at which $a = 1$ is foregone by $d\tilde{\omega}$, and it shifts the likelihood of forgoing $a = 1$. By the product rule, the overall effect on incentives is (12). Of course, because $\underline{\omega}' \approx -b$, the effect of the change on the sender's incentives reduces to approximately $\phi(-b)\varepsilon > 0$, showing that the net value of a lie has increased. Thus, the increase in the lower threshold induces a benefit of approximately $b - r = \frac{(b-r)\phi(-b)\varepsilon}{\phi(-b)\varepsilon}$ per unit of distortion in the sender's incentives.

Of course, the upper threshold $\overline{\omega}'$—determined by condition (11)—is introduced as a deterrent to lying that restores the sender's incentives. The key observation is that the unit cost of restoring

[18] For instance, if $r = 0$, then we can maintain incentive compatibility while taking both thresholds to 0, and thus an uninformative payoff.

incentives in this way is negligible, so that overall the receiver benefits from the perturbation. Indeed, for a large threshold $\overline{\omega}'$, the unit cost is approximately[19]

$$\frac{-(\omega+r)\phi(\omega)}{(\omega+b)\phi'(\omega)+\phi(\omega)} = \frac{(\omega+r)}{(\omega+b)\omega-1}, \tag{13}$$

evaluated at $\omega = \overline{\omega}'$. The numerator represents the expected cost to the receiver; the denominator the slackening of the incentive constraint (ω-ENV); and the right hand side follows because ${}^{\phi'(\omega)}/_{\phi(\omega)} = -\omega$. It is easy to see that this unit cost converges to 0 as $\omega \to \infty$. Hence, this punishment restores incentives at almost no cost to the receiver. The cost of restoring incentives via surplus burning punishments at the top is small because the likelihood of punishment in the tail is very low under truth-telling; yet the relative deterrence effect is extremely large because the punishment is far more likely after a lie.[20]

The efficiency of punishing at the top is central to the optimality of interval mechanisms among all simple ones. This can be seen by examining the dual of problem (P1), where the simple envelope condition (ω-ENV) is used in place of (ENV). After inspection of (ω-ENV), it is easy to see that the derivative of the appropriate Lagrangian with respect to $\sigma(\omega)$ is:

$$\omega + r - \alpha\left(-(\omega+b)\frac{\phi'(\omega)}{\phi(\omega)} - 1\right) = \omega + r - \alpha\left((\omega+b)\,\omega - 1\right), \tag{14}$$

where $\alpha > 0$ is the associated multiplier for the constraint (ω-ENV). This first-order condition expresses the marginal value of *increasing* $\sigma(\omega)$ at state ω, net of any shadow cost for violating (ω-ENV). Moreover, it is easy to see that (14) is quadratic and hence positive on at most an intermediate interval of states. When ω is sufficiently small, increases in $\sigma(\omega)$ are both costly to the receiver and tighten the sender's incentive constraint. Hence, it is optimal to set $\sigma(\omega) = 0$ there. As ω increases, a tension arises between increasing the receiver's payoffs and providing the sender with incentives: per (14), this tension is resolved precisely by punishing the sender at at the top, where the unit cost (13) to the receiver is smallest.[21]

[19]This ratio arises from an exercise very similar to that described for calculating the unit benefit of raising the lower threshold, albeit using l'Hopital's rule to approximate the ratio of cost to incentives when $\overline{\omega}'$ is large.

[20]Formally, this vanishing unit cost relies on the log-concavity of the Normal distribution: the identity ${}^{\phi'(\omega)}/_{\phi(\omega)} = -\omega$ implies that the hazard-rate-like ratio in the denominator of (13) grows without bound, driving the cost to zero. In a moral hazard context, a similar intuition underlies Mirrlees's 'unpleasant theorem' (Mirrlees, [1975] 1999): punishment in the tails is effective because their relative likelihood (following a deviation) is largest. However, our problem involves no transfers, so punishments cannot all be loaded onto a single, vanishingly unlikely event. Instead, they must be applied to a range of events, leaving our solution bounded strictly away from the first best.

[21]Notice, that equation (13) implies that the argument for punishing at the top should extend to other preferences so long as the ratio of receiver-to-senders payoffs does not grow faster than $|\frac{\phi'(\omega)}{\phi(\omega)}|$.

5.3 Optimality of Simple Mechanisms in the Large

Given the example in Section 3, it is not at all obvious that it is without loss of optimality to focus on simple mechanisms. However, we will see that optimality of simple mechanisms arises as a consequence of the limiting incentive conditions expressed by (ENV) and (MON).

Consider replacing an ICL mechanism $\sigma(\mathbf{h})$ with its conditional mean $\bar{\sigma}(\omega) = \mathbb{E}[\sigma(\mathbf{h}) \mid \omega]$. Clearly, doing so collapses the mechanism to a simple one, and by the linearity of the receiver's expected utility in σ, leaves her payoff unchanged. Hence, the only issue is whether the change preserves incentives. For an optimal mechanism σ, this amounts to checking that (ω-ENV) holds for $\bar{\sigma}(\omega)$—because (MON) is slack whenever σ is informative, and the uninformative mechanism is already simple.[22] We argue that (ω-ENV) holds in two steps: first, (ENV) deters what we call *state-shifting nudges*—combined perturbations whose only aggregate effect is to shift the reported state $\tilde{\omega}$; and second, because the limit sender is uninformed, the effect of such a state-shifting nudge is identical whether it is applied through σ or through its state-conditional mean $\bar{\sigma}$.

Step 1: (ENV) Deters State-Shifting Nudges. Recall from Remark 2 that condition (ENV) is equivalent to saying that an uninformed sender could not profit from nudging the reported NEF $\tilde{\mathbf{h}}$ by an incremental amount dh_k, in any direction k. These first-order conditions also deter the sender from perturbing the empirical frequency distribution in *any* small way: that is, not only are nudges in each direction k deterred, but linear combinations of those nudges are also deterred. For instance, he must not wish to increase the frequency of two types at the expense of two others. In this way, incentive compatibility deters a large set of complex deviations by the sender. Importantly, this implies that (ENV) also deters simple state-shifting deviations, in which the sender's combined nudge only has the effect of shifting the (normalized) empirical mean of the distribution, $\tilde{\omega} = \mathbf{t} \cdot \tilde{\mathbf{h}}$.

We can construct state-shifting nudges via a linear combination of deviations. In particular, the sender can do so by nudging each h_k by the amount that would be expected following a $d\omega$ shift of the state, that is, by $d\tilde{h}_k = \mathbb{E}[h_k \mid \omega = d\omega]$. Importantly, this nudge does not change the conditional distribution of the reported NEF $\tilde{\mathbf{h}}$ given the reported state $\tilde{\omega}$, so its only effect is on $\tilde{\omega}$, which shifts by $d\tilde{\omega} = \sum t_k d\tilde{h}_k$.[23]

Step 2: Essentially Uninformed Sender. Since ICL mechanisms automatically deter these state-shifting nudges, $\bar{\sigma}$ must also deter them. From the perspective of an uninformed sender, a state-

[22] Notice, if $\bar{\sigma}$ satisfies (ω-ENV), it is feasible in a relaxed problem in which (MON) is dropped. But the solution to this relaxed problem is σ^*, which as we saw in Section 5.1, satisfies (MON). Hence, if $\bar{\sigma}$ satisfies (ω-ENV), then $\sigma^*(\omega)$ is a simple ICL mechanism with a weakly greater payoff than any ICL $\sigma(\mathbf{h})$.

[23] Formally, this is achieved by a nudge $d\mathbf{h} = \beta d\omega$ where direction β is a $K \times 1$ vector of regression coefficients $\beta_k = \frac{\partial \mathbb{E}[h_k|\omega]}{\partial \omega}$, so that $\beta_k d\omega$ represents the conditional expectation of h_k given an expected state of $d\omega$. Such a nudge implies the induced distribution of $\tilde{\mathbf{h}}$ conditional on state ω is equal to the distribution of the true NEF, $\mathbf{h}$, conditional on state $\omega + d\omega$. We provide the calculation in Online Appendix C.2.2.

shifting nudge implies that for each realization of the state, the probability of acceptance shifts by $\mathbb{E}\left[\sigma\left(\mathbf{h}\right) \mid \omega+d\omega\right]-\mathbb{E}\left[\sigma\left(\mathbf{h}\right) \mid \omega\right]=\bar{\sigma}(\omega+d\omega)-\bar{\sigma}(\omega)$. Since condition (ENV) implies this nudge is unprofitable, then it must also be unprofitable under mechanism $\bar{\sigma}$ to nudge the distribution of $\tilde{\omega}$ upwards by $d\omega$. Indeed, we show in Online Appendix C.2.1 that one can use this exercise to show that if σ satisfies (ENV), then $\bar{\sigma}\left(\omega\right)$ satisfies (ω-ENV).

Notice that the two-step argument above relied on the fact that the sender is essentially uninformed in the limit, i.e., $\mathbb{E}\left[\sigma\left(\mathbf{h}\right) \mid \omega, t_k\right]=E\left[\sigma\left(\mathbf{h}\right) \mid \omega\right]$; hence, from his perspective the state-shifting nudge described above had expected impact $\bar{\sigma}(\omega+d\omega)-\bar{\sigma}(\omega)$ on acceptance in state ω. By contrast, this equality does not hold for finite N as each sender has significant private information about $\mathbf{h}$ conditional on ω. For instance, in the simple example where $N=2$, a sender can infer from his own signal exactly what the other sender's signal must be for each realization of ω. As a result, incentive compatibility is violated by the simple mechanism $\bar{\sigma}(\omega)=\mathbb{E}\left[\sigma^{\star}\left(\mathbf{h}\right) \mid \omega\right]$ that collapses the optimal mechanism described in Figure 2b down to a simple one. Notice, this collapse only changes the probability of accept along the $\omega=0$ diagonal: it smoothes the probability of acceptance yielding a constant $\bar{\sigma}(0)={}^{17}/_{21}$. Now, $\bar{\sigma}(0)<\mathbb{E}\left[\sigma^{\star}\left(\mathbf{h}\right) \mid \omega=0, l\right]=\sigma^{\star}\big(l,h\big)$. Clearly, this reduces the value of an l-report. Similarly, it increases the value of an m-report as $\bar{\sigma}(0)>\sigma^{\star}\big(m,m\big)=\mathbb{E}\left[\sigma^{\star}\left(\mathbf{h}\right) \mid \omega=0, m\right]={}^{3}/_{7}$. Because the sender has material private information and influence when $N=2$, he knows that the effect of the smoothing is to reduce the chances of acceptance when $\tilde{\omega}=0$ and he reports l, and yet increase it when $\tilde{\omega}=0$ and he reports m. Since type l was indifferent between reporting l and m under $\sigma^{\star}$, he must strictly prefer to report m under $\bar{\sigma}$. Thus, the private information of the sender prevents us from collapsing $\sigma^{\star}$ down to a simple mechanism.

Remark 3. As it is without loss to reduce mechanisms down to simple ones in the limit, one might expect that any violation of incentive compatibility caused by such a reduction for finite N, should shrink as the population grows. Indeed, we show in Online Appendix C.2.3 that this is the case.

5.4 Convergence Properties of Optimal Finite Mechanisms

The uniqueness result in Theorem 2 implies that any sequence of optimal mechanisms, $\sigma^{N,*}$, converge in the appropriate sense. To see this, note that any sequence of optimal outcome distributions $(\mathbf{h}^N, a^{N,*})$ has a convergent subsequence. This follows from Prokhorov's theorem; see, for instance, Billingsley (2013). By the almost everywhere uniqueness of σ^*, every convergent subsequence converges in distribution to $(\mathbf{h}, a^*)$. But a sequence converges if and only if every subsequence has a further subsequence which converges to the same limit.[24] Thus:

[24]This argument is almost identical to the proof of Lemma 2.17 in Aliprantis and Border (2006).

Corollary 1. *Let* $(\mathbf{h}^N, a^{N,*})$ *be the outcome induced by optimal mechanism* $\sigma^{N,*}$ *for finite* N*, and* $(\mathbf{h}, a^*)$ *the outcome induced by* σ^**. Then,* $(\mathbf{h}^N, a^{N,*}) \xrightarrow{d} (\mathbf{h}, a^*)$.

This result implies that finding the optimal ICL mechanism was the correct exercise for thinking about large economies, because σ^* approximates the optimal mechanism for large finite populations.

6 Heterogeneous Bias

The main results of our paper are not tied to the common bias baseline. To illustrate this, as well as how our ICL approach can be applied to other settings, we consider a model in which senders' biases are heterogeneous. Formally, the senders observe independent signals and have payoffs given by

$$a \cdot (\omega + b_m)$$

where b_m is the bias of a sender of class m which is publicly observed by the mechanism and ω is the same as before. In this environment, an argument akin to that of Theorem 1 implies that the ICL holds if and only if, for all m:

$$\mathbb{E}\left[\sigma(\mathbf{h}_1, \cdots, \mathbf{h}_M)\left((\omega + b_m)\frac{h_{k,m}}{f_k} - t_k\right)\right] = 0, \tag{15}$$

$$\mathbb{E}\left[\sigma(\mathbf{h}_1, \cdots, \mathbf{h}_M)\frac{h_{k,m}}{f_k}\right] \text{is increasing in } k$$

where in the above $\mathbf{h}_m$ is the NEF vector among the senders of class m.

In this setting, a simple mechanism is one in which the mechanism σ depends on the aggregate report of each group ω_m with

$$\omega = \frac{\sum_{m=1}^{M} \sqrt{N_m}\omega_m}{\sqrt{N}}$$

where N_m is the number of senders of class m and N is the total number of senders. In the Online Appendix D, we show that an optimal mechanism is indeed simple in the sense mentioned here. Moreover, it can be represented by a hyperbola in the space of the average reports $(\omega_1, \cdots, \omega_M)$ inside which accept is recommended while reject is again recommended for extreme values outside of the hyperbola. The following example illustrates this when there are two bias classes.

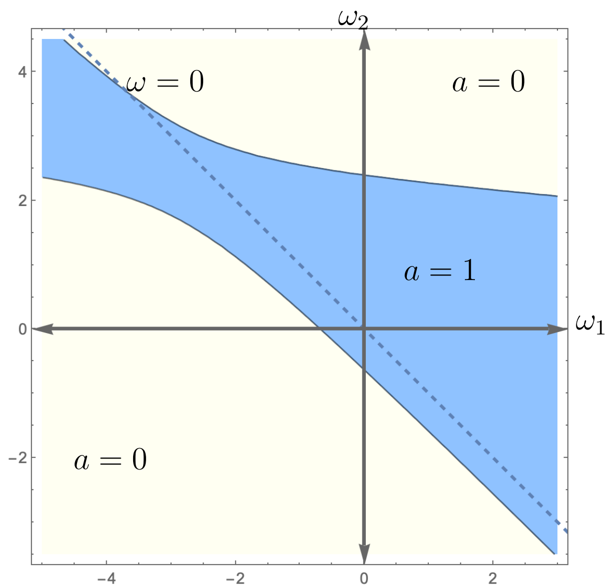

Figure 3: Regions of action with heterogeneous bias; in the white region, $a = 0$, while in the blue region, $a = 1$. The parameters are var$(s_i) = 1$, $b_1 = 0.1, b_2 = 0.3$, and $r = 0$.

Example. Suppose that there are two groups of senders with biases $b_1 = 0.1, b_2 = 0.3$ and that the two groups of senders are equal in size. In the large economy, the relevant statistics are the average reports within each class rather than the full report profile. Figure 3 depicts the optimal mechanism in the space of average signal of each group (ω_1, ω_2). The acceptance region in this mechanism sits between two parabolas. Similar to the optimal mechanism in Theorem 2, the senders are punished with rejection in the unlikely event that both group means are very high. However, there is a key difference from the common bias case: the mediator listens more to the less biased group— higher values of ω are more likely to be rewarded when ω_1 is high.

In the Online Appendix D, we provide the formal version of this model as well as one in which we allow for a more general class of preferences.

7 Related Literature

A large literature studies how information can be extracted by cross-checking when agents have correlated information. Crémer and McLean (1988) show that full surplus extraction can be achieved via transfers when each agent's information statistically identifies the others'. In multi-sender cheap talk, Krishna and Morgan (2001), Battaglini (2002), and Meyer et al. (2019) show how punishing disagreement can support full revelation when senders' information is almost identical.[25] Battaglini (2004) and Gerardi et al. (2009) extend this logic to imperfectly correlated

[25] A notable exception to the result of Battaglini (2002) is provided in Levy and Razin (2007). In their model, they show full revelation may be impeded in multi-dimensional talk when the states are correlated. In addition to the

information with large populations. The common force in these papers is that one sender's report can be compared with others'. That is impossible here, so we cannot implement the receiver-preferred allocation by cross-checking. Indeed, our optimal mechanisms may punish agreement rather than disagreement.

Similar to our paper, Wolinsky (2002) and Kattwinkel and Winter (2024) examine transfer-free mechanism design for a binary decision problem with many senders who have common bias. Like us, the senders in Wolinsky (2002) have *unconditionally* independent signals. However, the signals are binary and there is verifiable disclosure. He shows that there is an optimal non-monotone mechanism in his setting, but it consists of several acceptance intervals rather than one. Kattwinkel and Winter (2024) study a problem with binary *conditionally* independent signals and characterize the optimal mechanism for fixed N. Interestingly, despite allowing for correlated signals, they show that interval mechanisms are still optimal when the bias is large. While their setting features partially correlated binary signals, one can show that the reason for optimality in their setting is similar to that discussed in Section 5 and 5.2. By contrast, they show that the sender-preferred mechanism can be optimal when bias is small, whereas it is not optimal for any bias in the large-population limit that we examine.[26]

Beyond the differences in the optimal mechanisms, our paper has two further offsets. First, by allowing for a richer type-space we showed in Section 3 that it is typically with loss to restrict attention to simple mechanisms in finite populations. However, we showed that as the population grows large, it is without loss to use simple mechanisms that ignore higher dimensions of the NEF. In their papers, this distinction does not arise because the NEF is one-dimensional in the binary-signal case. Second, we examine informational division. This implies that the conflict of interest between the senders and receiver stays fixed as the population grows. By contrast, information is additive in their papers, so the probability of being in the disagreement region converges to zero and the sender-preferred allocation converges to the receiver's first best.

Non-monotonic decision rules can be optimal for other reasons. Chwe (2010) finds that non-monotonic mechanisms help when the informed agents have opposing preferences. In the context of protests, Gui and Ma (2025) show the government may not want to take an action when it is too popular because this can encourage revolt by revealing that there are large numbers of radicals in the population.

Our results can also be applied to voting (cf. Remark 1). In contrast to our results, several papers show how information aggregation by voters can be asymptotically efficient (Feddersen

multiple-sender case, and in contrast to Chakraborty and Harbaugh (2010), they show full revelation fails in a one sender version of their model. Carroll and Egorov (2019) show that full information extraction is possible if a single dimension of information can be verified.

[26] This is a feature of having an arbitrarily large population. As discussed in Section 2 and Online Appendix D.3, sender-preferred can be optimal when N is finite.

and Pesendorfer, 1997; McLennan, 1998; Gerardi et al., 2009; Barelli et al., 2022; Bobkova, 2024; Chen, 2026). The seminal paper on this, Feddersen and Pesendorfer (1997), examines a model of voters with signals about candidate competence (the common value state) and heterogeneous preferences over candidates. In equilibrium, voters split into swing voters who vote based on their information, and partisans who do not. As the population grows, the swing voters become increasingly homogenous and their proportion shrinks to zero. Even though the proportion shrinks, the absolute number of swing voters increases and their collective information perfectly identifies the state as $N \to \infty$. As a result, voting in their setting replicates the perfect-information outcome.

A key difference between our paper and the rest of the literature on information aggregation is that our limiting exercise captures informational division rather than addition. In Feddersen and Pesendorfer (1997) for instance, each voter adds to the aggregate information of the population. This means that a vanishing fraction of voters can still have almost perfect information collectively. By contrast, we keep aggregate information constant by reducing signal precision as the population grows. If information were similarly divided in their setting, the vanishing fraction of swing voters would be of almost no informational value.[27] We are not the first to identify barriers to efficient aggregation. Indeed, Feddersen and Pesendorfer (1997) show aggregate uncertainty about preferences is one such barrier.[28] Downs (1957) identifies costly information acquisition as another, and Martinelli (2006) studies its implications. Even when information is costless, Bobkova (2024) shows all threshold voting rules except majority voting fail when voters choose their information source.

Beyond the difference in results, we provide a new methodology for studying aggregation by examining the large-population limit directly. Much of the voting literature instead fixes a voting rule and studies its asymptotic properties as they *add* agents with *signals of fixed precision* (Feddersen and Pesendorfer, 1997; Martinelli, 2006; Gerardi et al., 2009; Bobkova, 2024). By contrast, our paper shows how one can use the infinite-population limit to approximate incentive compatibility for large populations via the notion of incentive compatibility in the large. This allows us to examine the full mechanism design problem and thus the optimal mechanism.

There is also a voting literature that examines the large-population limit in settings without private information (Persson et al., 2000; Persson and Tabellini, 2002). In these papers, people vote as if they are pivotal even though they have no chance of being so. In the absence of private information, this is innocuous, since it clearly captures incentive compatibility for large finite populations. However, when there is private information, voters condition their beliefs on the

[27] Of course, if one were to change the nature of the asymptotic exercise, then the actual equilibrium behavior would presumably be quite different.

[28] In light of this issue, Kawamura (2013) investigates the value of surveying strict subsets of the population.

fact that they are pivotal (Austen-Smith and Banks, 1996; Feddersen and Pesendorfer, 1996, 1998; Chwe, 2010; McMurray, 2013). This makes it less obvious how the large-population limit should serve as an approximation for large finite populations. We show how to adapt these pivotality-style arguments at the limit.[29]

Our work is also related to several other papers that examine the limiting properties of mechanisms as the population grows. Pesendorfer and Swinkels (1997, 2000); Atakan and Ekmekci (2014) study when common value auctions aggregate information efficiently. McLean and Postlewaite (2004) show mechanisms that use cross-checking with side-payments can attain first-best for any population size, but that the size of the payments goes to zero in the limit. Battaglini and Palfrey (2024) study collective action problems where agents hold private information about their cost of contribution and optimal mechanisms to mitigate free-riding. Apart from the obvious differences in subject matter, our paper provides methods to characterize the asymptotics by studying an unrestricted mechanism design problem at the limit.

Finally, there is a methodological connection to Frick et al. (2026b) and Frick et al. (2026a) which examine mechanisms as the quality of the designer's information becomes asymptotically perfect. In the limit, there is no private information and full surplus extraction is possible. However, rather than solving for the optimal mechanism, the authors show that certain fixed simple mechanisms converge quickly to the first-best. By contrast, we characterize the asymptotic properties of the optimal mechanism indirectly, by solving a non-trivial limit problem. Our approaches are complimentary: we identify the point to which mechanisms must converge if they are to do so optimally. Future work could combine their methods and ours to identify simple mechanisms with rapid convergence properties in environments where the limit problem is non-trivial.

8 Concluding Discussion

Our results show that a decision maker can benefit from information being divided across multiple senders even when verification by cross-checking is impossible. The optimal mechanism characterized in Theorem 2 achieves this gain by punishing consensus in the direction of the senders' bias.[30] However, despite the improvement, the optimal mechanism still falls short of the first best.

[29] Matějka and Tabellini (2021) is one of the few papers on voting where there the population is infinite and agents are privately informed. They avoid the problem by assuming that agents do not condition their beliefs on their own pivotality. While such an approach is not consistent with Bayesian information aggregation, this is not the focus of their paper.

[30] In a related applied work on using network structure to detect coordinated disinformation without direct content verification, Casillas et al. (2024) recommend that social media platforms such as X use a mechanism in which the platform gives warning when many suspicious accounts start the same information to detect and deter disinformation dissemination.

Beyond the impossibility of cross-checking, this is because our model features informational division. If information were additive instead, conflicts of interest would disappear in the limit and first best would be achievable even with uncorrelated signals.[31] This points to an open question: with correlated signals, can cross-checking recover first best when the relevant asymptotic exercise still captures division rather than addition? On one hand, keeping aggregate information constant would make each individual essentially uninformed in the limit, which is discouraging for cross-checking. On the other hand, one can imagine information structures in which conditional correlations remain strong—because groups of individuals observe essentially the same signals as each other—even as these signals become uninformative about the state. Thus, the potential for cross-checking in the limit may be sensitive to the particulars of the information environment.

This raises a broader question: how should the distribution of information in large populations be modeled more generally? So far, the literature has taken what we describe as an additive approach, which implies any fixed proportion of the population becomes perfectly informed in the limit. By contrast, informational division implies a fixed proportion's information is constant. A general framework that nests these cases may help us better understand information aggregation in large populations and open up new problems for investigation.

These considerations also suggest that new insights may be gained by reexamining the literature on information aggregation through the lens of informational division (or a general framework that nests it). The papers discussed above span many economic environments, including voting, common-value auctions, and collective action, and the notion of incentive compatibility in the large can be applied in them all. Hence, similar methods to those developed in Theorem 1 offer an approach to analyzing division in these settings as well. Likewise, these methods may also prove useful in online settings where fake reviews, deep fakes, and bots create new problems for information aggregation.

References

Aliprantis, C. D. and K. C. Border (2006): *Infinite dimensional analysis: A Hitchhiker's Guide*, Springer.

Ambrus, A., E. M. Azevedo, and Y. Kamada (2013): "Hierarchical Cheap Talk," *Theoretical Economics*, 8, 233–261.

[31] In our setting, informational addition would be a state that is just the sum of signals, $\omega = \sum_{i=1}^{N} s_i$. This would imply that the probability of the state being in the disagreement region would go to zero as $N \to \infty$.

Antic, N., A. Chakraborty, and R. Harbaugh (2025): "Subversive conversations," *Journal of Political Economy*, 133, 1621–1660.

Atakan, A. E. and M. Ekmekci (2014): "Auctions, actions, and the failure of information aggregation," *American Economic Review*, 104, 2014–2048.

Aumann, R. J. and S. Hart (2003): "Long cheap talk," *Econometrica*, 71, 1619–1660.

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

Ball, I. (2024): "Scoring Strategic Agents," *American Economic Journal: Microeconomics. Forthcoming.*

Barelli, P., S. Bhattacharya, and L. Siga (2022): "Full information equivalence in large elections," *Econometrica*, 90, 2161–2185.

Battaglini, M. (2002): "Multiple referrals and multidimensional cheap talk," *Econometrica*, 70, 1379–1401.

——— (2004): "Policy advice with imperfectly informed experts," *Advances in Theoretical Economics*, 4.

Battaglini, M. and T. R. Palfrey (2024): "Organizing for Collective Action: Olson Revisited," *Journal of Political Economy*, 132, 2881–2936.

Best, J. and D. Quigley (2024): "Persuasion for the long run," *Journal of Political Economy*, 132, 1740–1791.

Billingsley, P. (1995): *Probability and measure*, John Wiley & Sons, 3. ed ed.

——— (2013): *Convergence of probability measures*, John Wiley & Sons.

Bobkova, N. (2024): "The Optimality of Majority Rule: An Information-Choice Perspective," *Available at SSRN 5018717.*

Carroll, G. and G. Egorov (2019): "Strategic communication with minimal verification," *Econometrica*, 87, 1867–1892.

Casillas, A., M. Farboodi, L. Hashemi, M. Saeedi, and S. Wilson (2024): "(Dis)Information Wars," NBER Working Paper 32896, National Bureau of Economic Research.

Chakraborty, A. and R. Harbaugh (2010): "Persuasion by cheap talk," *American Economic Review*, 100, 2361–2382.

Chen, K. (2026): "Communication as Voting," *arXiv preprint arXiv:2505.14639*.

Chwe, M. S.-Y. (2010): "Anonymous procedures for Condorcet's model: Robustness, nonmonotonicity, and optimality," *Quarterly Journal of Political Science*, 5, 45–70.

Corrao, R. and Y. Dai (2023): "The Bounds of Mediated Communication," *arXiv preprint arXiv:2303.06244*.

Cover, T. M. and J. A. Thomas (2006): *Elements of information theory*, John Wiley & Sons, 2. ed ed.

Crawford, V. and J. Sobel (1982): "Strategic information transmission," *Econometrica*, 50, 1431–1451.

Crémer, J. and R. P. McLean (1988): "Full extraction of the surplus in Bayesian and dominant strategy auctions," *Econometrica*, 1247–1257.

Downs, A. (1957): "An economic theory of political action in a democracy," *Journal of political economy*, 65, 135–150.

Feddersen, T. and W. Pesendorfer (1997): "Voting behavior and information aggregation in elections with private information," *Econometrica: Journal of the Econometric Society*, 1029–1058.

——— (1998): "Convicting the innocent: The inferiority of unanimous jury verdicts under strategic voting," *American Political Science Review*, 92, 23–35.

Feddersen, T. J. and W. Pesendorfer (1996): "The swing voter's curse," *The American economic review*, 408–424.

Frick, M., R. Iijima, and Y. Ishii (2026a): "Monitoring with rich data," *arXiv preprint arXiv:2312.16789*.

——— (2026b): "Multidimensional Screening with Precise Seller Information," *Econometrica*, 94, 35–70.

Gerardi, D., R. McLean, and A. Postlewaite (2009): "Aggregation of expert opinions," *Games and Economic Behavior*, 65, 339–371.

Golosov, M., V. Skreta, A. Tsyvinski, and A. Wilson (2014): "Dynamic Strategic Information Transmission," *Journal of Economic Theory*, 151, 304–341.

Goltsman, M., J. Hörner, G. Pavlov, and F. Squintani (2009): "Mediation, arbitration and negotiation," *Journal of Economic Theory*, 144, 1397–1420.

Gui, Z. and Z. Ma (2025): "Slippery Protests and Information Aggregation," *Available at SSRN 5315735*.

He, K., F. Sandomirskiy, and O. Tamuz (2026): "Private private information," *Journal of Political Economy*, 134, 000–000.

Kattwinkel, D. and A. Winter (2024): "Optimal Decision Mechanisms for Committees: Acquitting the Guilty," *arXiv preprint arXiv:2407.07293*.

Kawamura, K. (2013): "Eliciting information from a large population," *Journal of Public Economics*, 103, 44–54.

Krishna, V. and J. Morgan (2001): "A model of expertise," *The Quarterly Journal of Economics*, 116, 747–775.

——— (2004): "The art of conversation: eliciting information from experts through multi-stage communication," *Journal of Economic Theory*, 117, 147–179.

Levy, G. and R. Razin (2007): "On the Limits of Communication in Multidimensional Cheap Talk: a Comment," *Econometrica*, 75, 885–893.

Luenberger, D. G. (1997): *Optimization by vector space methods*, John Wiley & Sons.

Martinelli, C. (2006): "Would rational voters acquire costly information?" *Journal of economic theory*, 129, 225–251.

Matějka, F. and G. Tabellini (2021): "Electoral competition with rationally inattentive voters," *Journal of the European Economic Association*, 19, 1899–1935.

McLean, R. and A. Postlewaite (2004): "Informational size and efficient auctions," *The Review of Economic Studies*, 71, 809–827.

McLennan, A. (1998): "Consequences of the Condorcet jury theorem for beneficial information aggregation by rational agents," *American Political Science Review*, 92, 413–418.

McMurray, J. C. (2013): "Aggregating information by voting: The wisdom of the experts versus the wisdom of the masses," *Review of Economic Studies*, 80, 277–312.

Meyer, M., I. Moreno de Barreda, and J. Nafziger (2019): "Robustness of full revelation in multisender cheap talk," *Theoretical Economics*, 14, 1203–1235.

Migrow, D. (2021): "Designing communication hierarchies," *Journal of Economic Theory*, 198, 105349.

Mirrlees, J. A. ([1975] 1999): "The Theory of Moral Hazard and Unobservable Behaviour: Part I," *The Review of Economic Studies*, 66, 3–21, paper completed 1975, published 1999.

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

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

Persson, T., G. Roland, and G. Tabellini (2000): "Comparative politics and public finance," *Journal of political Economy*, 108, 1121–1161.

Persson, T. and G. Tabellini (2002): *Political economics: explaining economic policy*, MIT press.

Pesendorfer, W. and J. M. Swinkels (1997): "The loser's curse and information aggregation in common value auctions," *Econometrica: Journal of the Econometric Society*, 1247–1281.

——— (2000): "Efficiency and information aggregation in auctions," *American Economic Review*, 90, 499–525.

Salamanca, A. (2021): "The value of mediated communication," *Journal of Economic Theory*, 192, 105191.

Strack, P. and K. H. Yang (2024): "Privacy-Preserving Signals," *Econometrica*, 92, 1907–1938.

Whitmeyer, M. (2024): "Bayesian Elicitation," *Arizona State University Working Paper*.

Wolinsky, A. (2002): "Eliciting information from multiple experts," *Games and Economic Behavior*, 41, 141–160.

A Proofs

A.1 Proof of Theorem 1

Proof. Let $a^N \in \{0,1\}$ denote the recommended action induced by finite-N mechanism σ^N. That is, a^N is a random variable with conditional distribution

$$\Pr\left(a^N = 1 \mid \mathbf{h}^N\right) = \sigma^N\left(\mathbf{h}^N\right).$$

For any integrable function g, the Law of Iterated Expectations implies that

$$\mathbb{E}^N\left[\sigma^N\left(\mathbf{h}^N\right)g\left(\mathbf{h}^N\right)\right] = \mathbb{E}^N\left[a^N g\left(\mathbf{h}^N\right)\right]. \tag{16}$$

In what follows, this observation will be useful for two reasons. First, when g is a continuous function of $\mathbf{h}^N$, the function $a^N g(\mathbf{h}^N)$ is also continuous in $\left(\mathbf{h}^N, a^N\right) \in \mathbb{R}_0^K \times \mathbb{R}$. Second, since $\{0,1\}$ is metrized by the discrete metric, the space $\mathbb{R}_0^K \times \{0,1\}$ is metrizable. As will become clear, these facts will aid convergence arguments.

"Only If" Direction. Assume that σ is ICL in the sense of Definition 1. Then, there exists a sequence of incentive compatible mechanisms $\{\sigma^N\}_{N\geq 1}$ satisfying (F-IC) such that

$$\left(\mathbf{h}^N, a^N\right) \xrightarrow{d} (\mathbf{h}, a),$$

where $\mathbf{h} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ (recall Lemma 2) and $\Pr(a = 1 \mid \mathbf{h}) = \sigma(\mathbf{h})$. Similar to a^N above, a is the random variable induced by the limit mechanism σ and the random variable $\mathbf{h}$.

For any $k > l$, incentive compatibility requires that σ^N satisfy the bounds (N-ENV). Using (8), (7) and (16) to substitute terms, (N-ENV) can be equivalently expressed

$$\begin{aligned}\mathbb{E}^N\left[a^N\left(1+\frac{h_k^N}{f_k\sqrt{N}}\right)\right] &\geq \frac{1}{t_k - t_l}\mathbb{E}^N\left[a^N\left(\mathbf{h}^N\cdot\mathbf{t}+b\right)\left(\frac{h_k^N}{f_k}-\frac{h_l^N}{f_l}\right)\right] \\ &\geq \mathbb{E}^N\left[a^N\left(1+\frac{h_l^N}{f_l\sqrt{N}}\right)\right].\end{aligned} \tag{17}$$

We now show that (ENV) is satisfied by passing (17) to the limit. Each of the three integrands in (17) is continuous in the variables $(a^N, \mathbf{h}^N)$. Though these integrands are clearly not bounded (so we cannot appeal to Portmanteau's theorem directly), we show that they are *uniformly integrable*, and as a result the desired convergence follows from Theorem 25.12 in Billingsley (1995) and the continuous mapping theorem. Recall that a sequence of random variables Y_N are said to be uniformly integrable if they satisfy

$$\lim_{M\to\infty}\sup_N \mathbb{E}^N\left[|Y_N|\,\mathbf{1}\left[|Y_N| > M\right]\right] = 0$$

We will do this only for $Y_N = a^N\left(\mathbf{h}^N\cdot\mathbf{t}+b\right)\frac{h_k^N}{f_k}$. The argument for the other terms in (17) are identical. Note that

$$|Y_N| \leq \frac{1}{f_k}\left|\mathbf{h}^N\cdot\mathbf{t}+b\right|\left\|\mathbf{h}^N\right\| \leq \frac{1}{f_k}\left(\left\|\mathbf{h}^N\right\|\|\mathbf{t}\|+b\right)\left\|\mathbf{h}^N\right\| \leq C(1+\left\|\mathbf{h}^N\right\|^2)$$

where in the above we have used the Cauchy–Schwartz inequality, the fact that $\left\|\mathbf{h}^N\right\| \leq 1 +$

$\left\|\mathbf{h}^N\right\|^2$ and $C = (\|t\| + b) / \min f_k$.

The above inequality implies that $|Y_N| > M \Rightarrow \sqrt{\frac{M-C}{C}} \leq \left\|\mathbf{h}^N\right\|$. We can apply Hoeffding's inequality to $\mathbf{h}^N$ and thus we have

$$\Pr\left(\|\mathbf{h}^N\| \geq z\right) \leq \Pr\left(\max_k \left|h_k^N\right| \geq \frac{z}{\sqrt{K}}\right) \leq \sum_{k=1}^{K} \Pr\left(\left|h_k^N\right| \geq \frac{z}{\sqrt{K}}\right) \leq 2Ke^{-\frac{2z^2}{K}}.$$

Hence, we can write

$$\mathbb{E}^N\left[|Y_N|\,\mathbf{1}\left[|Y_N| > M\right]\right] \leq \mathbb{E}^N\left[C(1 + \left\|\mathbf{h}^N\right\|^2)\mathbf{1}\left[\left\|\mathbf{h}^N\right\| > \sqrt{\frac{M-C}{C}}\right]\right]$$

if μ^N is the probability distribution of $\mathbf{h}^N$, we can use the earlier version of Hoeffding's inequality and write the above as

$$\begin{aligned}
\mathbb{E}^N\left[|Y_N|\,\mathbf{1}\left[|Y_N| > M\right]\right] &\leq \int_{\|\mathbf{h}^N\| \geq \sqrt{\frac{M-C}{C}}} C\left(1 + \|\mathbf{h}^N\|^2\right) d\mu^N \\
&= \mu^N\left(\left\{\mathbf{h} \mid \|\mathbf{h}\| \geq \sqrt{\frac{M-C}{C}}\right\}\right) M + \int_{M/C}^{\infty} \mu^N\left(\left\{\mathbf{h} \mid \|\mathbf{h}\|^2 + 1 \geq z\right\}\right) dz \\
&\leq 2KMe^{-2\frac{M-C}{C}} + 2K\int_{M/C}^{\infty} e^{-2(z-1)}dz = K\left(2M+1\right)e^{-2\frac{M-C}{C}}. \qquad (18)
\end{aligned}$$

The RHS of (18) converges to 0 as $M \to \infty$. This concludes the proof of asymptotic uniform integrability. Therefore, we may repeatedly apply Theorem 25.12 of Billingsley (1995) to the terms in (17) to recover (ENV) as we pass to the limit. Hence, the argument in the text is valid and (ENV) is established. The monotonicity constraint is established by considering the outermost inequalities in (17) and multiplying by $\sqrt{N}$. The same convergence argument as above establishes the claim.

"If" Direction. Assume that σ satisfies (ENV) and (MON). We construct a sequence of finite-N incentive compatible mechanisms $\{\sigma^N\}$ whose induced joint laws converge to the joint law generated by σ.

Step 1: Constructing a sequence of mechanisms that approximate σ**.** In this step, we find a sequence of mechanisms $\tilde{\sigma}^N$ whose induced joint laws over $(\mathbf{h}^N, a^N)$ converge in distribution to $(\mathbf{h}, a)$; in subsequent steps we make small adjustments along this sequence to ensure incentive compatibility is satisfied while preserving the appropriate convergence in distribution to $(\mathbf{h}, a)$.

To aid this task, it is convenient to construct a probability space on which the $\mathbf{h}^N$ satisfy a strong convergence property. Since $\mathbf{h}^N \xrightarrow{d} \mathbf{h}$, we may apply the Skorokhod representation theorem (Billingsley (1995), Theorem 25.6) to establish the existence of a collection $\left\{\left(\hat{\mathbf{h}}^N\right)_{N=1}^{\infty}, \hat{\mathbf{h}}\right\}$

of random vectors on a common probability space such that $\hat{\mathbf{h}}^N \overset{d}{=} \mathbf{h}^N$ for each N, $\hat{\mathbf{h}} \overset{d}{=} \mathbf{h}$, and the sequence $\left\{\hat{\mathbf{h}}^N\right\}$ converges to $\mathbf{h}$ almost surely. For notational simplicity, we relabel $(\hat{\mathbf{h}}^N, \hat{\mathbf{h}})$ as $(\mathbf{h}^N, \mathbf{h})$ below.

On this space, we may also define appropriate action recommendations as follows. Let $y \sim \mathrm{Unif}[0,1]$ be independent of $\mathbf{h}$ and define

$$a := \mathbf{1}\left\{y \leq \sigma(\mathbf{h})\right\}.$$

For each N, let $\tilde{\sigma}^N : \mathbb{R}_0^K \to [0,1]$ be a (measurable) version of the conditional probability

$$\tilde{\sigma}^N(\mathbf{h}) := \Pr\big(a = 1 \mid \mathbf{h}^N = \mathbf{h}\big),$$

that is, the conditional probability that $a = 1$ given observation of the random variable $\mathbf{h}^N$. Let $y^N \sim \mathrm{Unif}[0,1]$ be a sequence of independent uniform random variables and set[32]

$$\tilde{a}^N := \mathbf{1}\left\{y^N \leq \tilde{\sigma}^N\big(\mathbf{h}^N\big)\right\}.$$

By construction, the pair $(\mathbf{h}^N, \tilde{a}^N)$ has the same distribution as $(\mathbf{h}^N, a)$, and since $\mathbf{h}^N \underset{a.s.}{\longrightarrow} \mathbf{h}$, we have $(\mathbf{h}^N, a) \xrightarrow{d} (\mathbf{h}, a)$,[33] and therefore

$$\big(\mathbf{h}^N, \tilde{a}^N\big) \xrightarrow{d} (\mathbf{h}, a). \tag{19}$$

It remains to show that $\tilde{\sigma}^N$ can be adjusted so that *(i)* the resulting mechanisms satisfy the finite-N incentive constraints, and *(ii)* the resulting sequence of distributions over $(\mathbf{h}^N, a^N)$ also converge to the distribution over $(\mathbf{h}, a)$.

Step 2: A convenient reframing of incentive compatibility. For each $k \geq 2$, define

$$\begin{aligned} w_k(\mathbf{h}) &:= g_1^k(\mathbf{h}) - (t_k - t_{k-1}), \\ w_{k,N}(\mathbf{h}) &:= g_1^k(\mathbf{h}) - (t_k - t_{k-1})\, g_2^k(\mathbf{h}, N), \end{aligned}$$

where $g_1^k(\mathbf{h}^N) = \big(\mathbf{h}^N \cdot \mathbf{t} + b\big)\left(\frac{h_k^N}{f_k} - \frac{h_{k-1}^N}{f_{k-1}}\right)$ and $g_2^k(\mathbf{h}^N) = 1 + \frac{h_{k-1}^N}{f_{k-1}\sqrt{N}}$. For adjacent types k and $k-1$, the lower bound in constraint (N-ENV) can be written (see (17)) in terms of $w_{k,N}$ as

$$\mathbb{E}^N\big[\sigma^N(\mathbf{h}^N) w_{k,N}(\mathbf{h}^N)\big] \geq 0, \tag{20}$$

[32] y and the $\{y^N\}_{N=1}^{\infty}$ are independent of each other and all other variables, including $\mathbf{h}$ and the $\left\{\mathbf{h}^M\right\}_{M=1}^{\infty}$.

[33] To see this, observe that the almost sure convergence of $\mathbf{h}^N$ to $\mathbf{h}$ implies that for each open hypercube $H \subset \mathbb{R}_0^K$ and $a \in \{0,1\}$ (and hence for each open subset of $\mathbb{R}_0^K \times \{0,1\}$), we have $\Pr(a, \mathbf{h}^N \in H) \to \Pr(a, \mathbf{h} \in H)$.

for all $k = 2, \dots, K$. If σ^N satisfies (20) with equality, then the upper bound in (N-ENV) is satisfied if and only if

$$\mathbb{E}^N\left[\sigma^N(\mathbf{h}^N)(h_k^N/f_k - h_{k-1}^N/f_{k-1})\right] \geq 0, \tag{21}$$

for all $k = 2, \dots, K$.[34] In the next step, we work with this characterization of incentive compatibility to identify our sequence of incentive compatible mechanisms σ^N whose outcome distributions converge appropriately.

There is no reason that the mechanisms $\tilde{\sigma}^N$ identified in step 1 should satisfy (20) (or indeed, (21)). We measure the extent to which (20) differs from 0 for $\tilde{\sigma}^N$, by the difference

$$\delta_{k,N} := \mathbb{E}^N\left[\tilde{\sigma}^N\left(\mathbf{h}^N\right) w_{k,N}\left(\mathbf{h}^N\right)\right] = \mathbb{E}^N\left[\tilde{a}^N w_{k,N}\left(\mathbf{h}^N\right)\right], \qquad k = 2, \dots, K,$$

Since the outcome distributions of $\tilde{\sigma}^N$ satisfy (19), identical arguments to those developed in the 'only if' part apply so that the expectations of the $w_{k,N}$ converge, and hence

$$\delta_{k,N} \to \mathbb{E}[a\, w_k(\mathbf{h})] = \mathbb{E}[\sigma(\mathbf{h})\, w_k(\mathbf{h})] = 0,$$

where the last equality is simply a rewriting of ((ENV)).

Step 3: Correcting the $\tilde{\sigma}^N$ to satisfy incentive compatibility. We now adjust the $\tilde{\sigma}^N$ to develop a new sequence σ^N which satisfies (20) with equality and (21) for all sufficiently large N. This will establish existence of the desired limiting sequence. To focus on the main substantive arguments, we first show how to construct the σ^N under two assumptions, whose purpose is to avoid boundary issues: (i) there exists a $\varepsilon > 0$ such that σ and the $\tilde{\sigma}^N$ satisfy $\sigma(\mathbf{h}), \tilde{\sigma}^N(\mathbf{h}) \in [\varepsilon, 1-\varepsilon]$, for all $\mathbf{h}$ and N, and (ii) σ satisfies the monotonicity constraints (MON) with strict inequality. After proving the result for this case (see step 4 below), we describe how to extend this argument to apply to any σ which satisfies (ENV) and (MON) in step 5.

We identify values of $\mathbf{h}$ around which we may adjust $\tilde{\sigma}^N$ to restore (20) with equality, for $k = 2, \dots, K$, as follows. Define the map $W : \mathbb{R}_0^K \to \mathbb{R}^{K-1}$ by $W(\mathbf{h}) = (w_2(\mathbf{h}), \dots, w_K(\mathbf{h}))$. A direct calculation shows that for any $\mathbf{v} \in \mathbb{R}_0^K$,

$$Dw_k(0) \cdot \mathbf{v} = b\left(\frac{v_k}{f_k} - \frac{v_{k-1}}{f_{k-1}}\right), \qquad k = 2, \dots, K.$$

The matrix $DW(0)$ is invertible. To see this, note that if $DW(0)\mathbf{v} = 0$, then $v_k/f_k = c$ for all k, and so $\sum_k v_k = c\sum_k f_k = c = 0$ (recall $\sum_k v_k = 0$ for $\mathbf{v} \in \mathbb{R}_0^K$), which means $\mathbf{v} = 0$. Since $\dim(\mathbb{R}_0^K) = K-1$, $DW(0)$ has full rank (i.e., is invertible). Thus, by the inverse function theorem the image of W contains an open set in $\mathbb{R}^{K-1}$. Thus we may choose points $\bar{\mathbf{h}}^2, \dots, \bar{\mathbf{h}}^K \in \mathbb{R}_0^K$

[34] Note that (21) is just constraint (MON).

whose corresponding vectors $W(\bar{\mathbf{h}}^j)$, $j = 2, \ldots, K$, are linearly independent. These $\bar{\mathbf{h}}^2, \ldots, \bar{\mathbf{h}}^K$ are the locations around which our perturbations to the $\tilde{\sigma}^N$ will be made.

We adjust $\tilde{\sigma}^N$ as follows. Letting μ denote the measure on $\mathbb{R}_0^K$ corresponding to the distribution $\mathbf{h} \sim N(0, \Sigma)$, choose disjoint open balls B_j around the $\bar{\mathbf{h}}^j$ with $\mu(B_j) > 0$ and pick bounded, continuous functions $\psi_j : \mathbb{R}_0^K \to [0, 1]$ with $\mathrm{supp}(\psi_j) \subset B_j$ and $\psi_j(\bar{\mathbf{h}}^j) = 1$.[35] Since the supports are disjoint, we have $\sum_{j=2}^K \psi_j(\mathbf{h}) \leq 1$ for all $\mathbf{h}$. By appropriately scaling these functions, we will be able to make the desired modifications to the $\tilde{\sigma}^N$.

To help identify the appropriate scalings, define the matrix $M \in \mathbb{R}^{(K-1)\times(K-1)}$ whose $(k-1, j-1)^{th}$ entry is

$$M_{k-1,j-1} := \mathbb{E}[w_k(\mathbf{h})\, \psi_j(\mathbf{h})]\,, \qquad k, j = 2, \ldots, K.$$

Since w_k is continuous and the B_j can be taken arbitrarily small, M can be made arbitrarily close to the matrix with columns $W(\bar{\mathbf{h}}^j)\mathbb{E}[\psi_j(\mathbf{h})]$ and therefore made invertible.[36]

For each N, define the analogous matrix

$$(M_N)_{k-1,j-1} := \mathbb{E}^N\left[w_{k,N}\left(\mathbf{h}^N\right)\psi_j\left(\mathbf{h}^N\right)\right], \qquad k, j = 2, \ldots, K.$$

Because ψ_j is bounded and continuous, the same convergence argument used in the “only if” direction applies here and so $M_N \to M$ pointwise. Hence M_N is also invertible for all sufficiently large N.

Choose the scaling vector $\alpha^N := -M_N^{-1}\delta_N$, where δ_N is the $(K-1) \times 1$ vector whose entries $\delta_{k-1,N}$, $k = 2, \ldots, K$, are defined in step 2. Since $\delta_N \to 0$ and $M_N \to M$—where M is invertible, we have $\alpha^N \to 0$. We are now ready to define the corrected mechanism

$$\sigma^N(\mathbf{h}) := \tilde{\sigma}^N(\mathbf{h}) + \sum_{j=2}^K \alpha_j^N \psi_j(\mathbf{h})\,. \tag{22}$$

Because the supports of ψ_j are disjoint, we have $|\sum_j \alpha_j^N \psi_j| \leq \max_j |\alpha_j^N|$. Since, by hypothesis, $\tilde{\sigma}^N \in [\varepsilon, 1-\varepsilon]$ for all N, $\alpha^N \to 0$ implies that there exists an N' such that for all $N \geq N'$, $\sigma^N(\mathbf{h}) \in [0, 1]$. Moreover, by construction,

$$\mathbb{E}^N\left[\sigma^N\left(\mathbf{h}^N\right) w_{k,N}\left(\mathbf{h}^N\right)\right] = \delta_{k,N} + M_{(k-1,\cdot);N}\alpha^N = 0, \qquad k = 2, \ldots, K.$$

where $M_{(k-1,\cdot);N}$ is the $k-1^{th}$ row of M_N. Thus, the sequence $\{\sigma^N\}$ satisfies (20) with equality

[35] That is, $\psi_j(\mathbf{h}) = 0$ for $\mathbf{h} \notin B_j$.

[36] Recall that the $W(\bar{\mathbf{h}}^j)$ are linearly independent. As is well known, scaling vectors by nonzero constants (in this case the $\mathbb{E}[\psi_j(\mathbf{h})]$) preserves their linear independence.

for all $N \geq N'$.

We still need to show that the sequence $\{\sigma^N\}$ satisfies (21) for N sufficiently large. By hypothesis, the limits

$$m_k := \mathbb{E}\left[\sigma(\mathbf{h})\left(\frac{h_k}{f_k} - \frac{h_{k-1}}{f_{k-1}}\right)\right], \qquad k = 2, \ldots, K,$$

are strictly positive. Applying the same convergence arguments used in the "only if" direction, we have

$$\mathbb{E}^N\left[\tilde{\sigma}^N(\mathbf{h}^N)\left(\frac{h_k^N}{f_k} - \frac{h_{k-1}^N}{f_{k-1}}\right)\right] = \mathbb{E}^N\left[\tilde{a}^N\left(\frac{h_k^N}{f_k} - \frac{h_{k-1}^N}{f_{k-1}}\right)\right] \to m_k.$$

Since $\alpha^N \to 0$ and the ψ_j are continuous and bounded, the same conclusion extends to σ^N and thus, there exists an N'' such that σ^N satisfies (21) with strict inequality for all $N \geq N''$. Taking $N \geq N^\star := \max\{N', N''\}$, the subsequence $\{\sigma^N\}_{N \geq N^\star}$ satisfies both (20) (with equality) and (21) for all $N \geq N^\star$. By the usual Spence-Mirrlees arguments, this means that, for $N \geq N^\star$, σ^N satisfies the full set (N-ENV) (equivalently, (17)) of incentive constraints.

Step 4: Convergence of the induced joint distributions. It remains to show that the joint distributions $(\mathbf{h}^N, a^N)$ induced by the sequence $\{\sigma^N\}_{N \geq N^\star}$ converge to that of $(\mathbf{h}, a)$.[37] To this end, first define the recommendation a^N induced by σ^N as

$$a^N := \mathbf{1}\left\{y^N \leq \sigma^N(\mathbf{h}^N)\right\},$$

where the y^N are those defined in part 1. Then $\Pr(a^N \neq \tilde{a}^N \mid \mathbf{h}^N) = |\sigma^N(\mathbf{h}^N) - \tilde{\sigma}^N(\mathbf{h}^N)|$. Hence for any bounded continuous function $g : \mathbb{R}_0^K \times \{0, 1\} \to \mathbb{R}$ where $\sup|g(\mathbf{h}, a)| := G < \infty$,

$$\left|\mathbb{E}^N\left[g\left(\mathbf{h}^N, a^N\right)\right] - \mathbb{E}^N\left[g\left(\mathbf{h}^N, \tilde{a}^N\right)\right]\right| \leq 2G\,\mathbb{E}^N\left[\left|\sigma^N\left(\mathbf{h}^N\right) - \tilde{\sigma}^N\left(\mathbf{h}^N\right)\right|\right].$$

The right-hand side converges to zero because $|\sigma^N - \tilde{\sigma}^N| \leq \max_j |\alpha_j^N|$ and $\alpha^N \to 0$ (see equation (22)). Combining this with (19) yields

$$\left(\mathbf{h}^N, a^N\right) \xrightarrow{d} (\mathbf{h}, a).$$

which is exactly ICL. This proves the result for all σ which satisfy *(i)* and *(ii)* described in step 3.

Step 5: extension to any σ satisfying (ENV)-(MON). Finally, we explain how to adapt the arguments from steps 3 and 4 for all σ which satisfy (ENV) -(MON). Take any such σ, and construct

[37]One can extend the sequence to all N per Definition (1), by simply choosing $\sigma^N = \sigma^S$ for all $N < N^\star$.

the nearby mechanism

$$\sigma'_{\varepsilon,\varphi}(\mathbf{h}) = (1 - 2\varepsilon - \varphi)\,\sigma(\mathbf{h}) + \varepsilon + \varphi\sigma^S(\mathbf{h}),$$

where the scalars ε and φ satisfy $\varepsilon, \varphi > 0$ and $2\varepsilon + \phi < 1$. $\sigma'_{\varepsilon,\varphi}$ is a mixture over three mechanisms: σ, the sender-preferred mechanism σ^S, and $\underline{\sigma} = \frac{1}{2}$, which randomizes between $a \in \{0, 1\}$ independently of $\mathbf{h}$. The mechanisms σ^S and $\underline{\sigma}$ are trivially incentive compatible for all N, and so unsurprisingly $\sigma'_{\varepsilon,\varphi}$ satisfies (ENV) and (MON). Moreover, σ^S satisfies (MON) with strict inequality, and therefore so too does $\sigma'_{\varepsilon,\varphi}$. Let $(\mathbf{h}, a')$ be the joint distribution induced by $\sigma'_{\varepsilon,\varphi}$ in the obvious way.

Similarly, replace the approximating sequence $\tilde{\sigma}^N$ defined in step 1 with

$$\tilde{\sigma}'^{,N}_{\varepsilon,\varphi} = (1 - 2\varepsilon - \varphi)\,\tilde{\sigma}^N(\mathbf{h}) + \varepsilon + \varphi\sigma^S(\mathbf{h}).$$

Notice that the $\tilde{\sigma}'^{,N}_{\varepsilon,\varphi}$ satisfy $\tilde{\sigma}'^{,N}_{\varepsilon,\varphi}(\mathbf{h}) \in [\varepsilon, 1-\varepsilon]$ for all $\mathbf{h} \in \mathbb{R}^K_0$. Using essentially the same arguments in step 1, it is easy to see that the $(\mathbf{h}^N, \tilde{a}'^{,N})$ induced by $\tilde{\sigma}'^{,N}_{\varepsilon,\varphi}$ converge in distribution to the $(\mathbf{h}, a'_{\varepsilon,\varphi})$ induced by σ'. Moreover, defining the natural analogue of the differences δ_N from step 2 by

$$\delta'_{k,N} := \mathbb{E}^N\left[\tilde{\sigma}'^{,N}_{\varepsilon,\varphi}\left(\mathbf{h}^N\right) w_{k,N}\left(\mathbf{h}^N\right)\right] = \mathbb{E}^N\left[\tilde{a}'^{,N}_{\varepsilon,\varphi} w_{k,N}\left(\mathbf{h}^N\right)\right], \qquad k = 2, \ldots, K,$$

it is easily verified that we again have $\delta'_{k,N} \to \mathbb{E}[\sigma'(\mathbf{h})\, w_k(\mathbf{h})] = 0$.[38] By step 3, this implies that, for each ε, φ, there is a sequence $\alpha^N_{\varepsilon,\varphi} \to 0$ which makes (20) hold with equality for $\sigma'^{,N}_{\varepsilon,\varphi}$ (defined in the obvious way, per (22)). Since $\sigma'_{\varepsilon,\varphi}$ satisfies *(i)* and *(ii)* from step 3, the $\sigma'^{,N}_{\varepsilon,\varphi}$ furthermore satisfy $\sigma'^{,N}_{\varepsilon,\varphi}(\mathbf{h}) \in [0, 1]$, for all $\mathbf{h}$, and (21), for all N greater than a threshold $N^\star_{\varepsilon,\varphi}$.

Finally, we identify a sequence σ^N that approximates σ via a diagonal argument. Specifically, set sequences $\varepsilon_m = \varphi_m = \frac{1}{2^{m+1}}$, $m = 1, 2, \ldots$, and for each m choose a $N^m \geq N^\star_{\varepsilon_m,\varphi_m}$ such that $\alpha^N_m = \alpha^N_{\varepsilon_m,\varphi_m}$ further satisfies $\max_k \left|(\alpha^N_m)_k\right| < \frac{1}{2^{m+1}}$ for all $N \geq N^m$, where $(\alpha^N_m)_k$ is the k^{th} entry of α^N_m. Furthermore, for each $N \geq N^1$, let $\underline{m}(N) = \max\{N^m : N^m \leq N\}$, write $\sigma'^{,N}_{\varepsilon_m,\varphi_m} = \sigma'^{,N}_m$, and define the sequence

$$\sigma^N = \begin{cases} \sigma'^{,N}_{\underline{m}(N)}, & \text{for } N \geq N^1 \\ \sigma^S, & \text{otherwise.} \end{cases}$$

For $N < N^1$, σ^N is trivially incentive compatible. Thereafter, by construction, the sequence $\sigma^N = \sigma'^{,N}_{\underline{m}(N)} = \tilde{\sigma}'^{,N}_{\underline{m}(N)} + \sum_{j=2}^K (\alpha^N_{\underline{m}(N)})_j \psi_j$, satisfies (20) (with equality), (21), and $\sigma^N \in [0, 1]$ for

[38] Note, we suppress the dependence of $\delta'_{k,N}$ on ε, φ for notational ease.

each N. Moreover, since $\varepsilon_{\underline{m}(N)}, \varphi_{\underline{m}(N)}, \alpha^{N}_{\underline{m}(N)} \to 0$, essentially the same convergence argument in step 4 applies, so that the $(\mathbf{h}^N, a^N)$ induced by the σ^N convergence in distribution to $(\mathbf{h}, a)$. This concludes the proof. □

A.2 Proof of Theorem 2

Proof. Note that any $\sigma(\mathbf{h})$ that is ICL satisfies

$$\mathbb{E}\left[\sigma(\mathbf{h})\left((\mathbf{h}\cdot\mathbf{t}+b)\sum_k h_k t_k - \sum_k f_k t_k^2\right)\right] = 0 \Rightarrow \mathbb{E}\left[\sigma(\mathbf{h})\left((\omega(\mathbf{h})+b)\,\omega(\mathbf{h}) - \text{Var}(s)\right)\right] = 0. \tag{23}$$

Now, consider the relaxed optimization where the ICL requirement, (ENV), is replaced with the above. Optimality implies that any optimal mechanism σ^* should satisfy

$$\sigma^*(\mathbf{h}) = 1 \Leftrightarrow \omega(\mathbf{h}) + r - \alpha\left((\omega(\mathbf{h})+b)\,\omega(\mathbf{h}) - \text{Var}(s)\right) \geq 0$$

for some Lagrange multiplier associated with the relaxed constraint (23). This condition implies that $\sigma^*(\mathbf{h})$ depends only on $\omega(\mathbf{h})$.

Note that if $\alpha > 0$, the above condition defines two cutoffs for $\underline{\omega} < \overline{\omega}$ that satisfy the following quadratic equation:

$$\omega + r + \alpha\text{Var}(s) - \alpha(\omega+b)\,\omega = 0. \tag{24}$$

For the above to have two roots, we need to have $(1-\alpha b)^2 + 4\alpha(r+\alpha\text{Var}(s)) \geq 0$. As we will show in the Online Appendix, Section D.4.1, when $b > r > \frac{b-\sqrt{b^2+4}}{2}$, for any $\alpha > 0$, this is the case. Moreover, it has to be that $-b < \underline{\omega} < 0 < -\underline{\omega} < \overline{\omega}$. Finally, there is a unique α such that the relaxed version of ICL, (23), holds.

This implies that the optimal mechanism in the relaxed problem must satisfy $\sigma^*(\mathbf{h}) = 1$ when $\omega = \mathbf{h}\cdot\mathbf{t} \in [\underline{\omega}, \overline{\omega}]$, and $\sigma^*(\mathbf{h}) = 0$, otherwise. Note that when $\sigma(\mathbf{h})$ is only a function of the sample mean, basic properties of the normal distribution implies that $\mathbb{E}[h_k|\omega] = \frac{f_k t_k}{\text{Var}(s)}\omega$. We can use this to show that

$$\begin{aligned}
\mathbb{E}\left[\sigma^*(\mathbf{h})\left((\omega(\mathbf{h})+b)\frac{h_k}{f_k} - t_k\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[\sigma^*(\mathbf{h})\left((\omega(\mathbf{h})+b)\frac{h_k}{f_k} - t_k\right)|\omega(\mathbf{h}) = \omega\right]\right] \\
&= \mathbb{E}\left[\sigma^*(\mathbf{h})\left((\omega(\mathbf{h})+b)\,\mathbb{E}\left[\frac{h_k}{f_k}|\omega(\mathbf{h}) = \omega\right] - t_k\right)\right] \\
&= \mathbb{E}\left[\sigma^*(\mathbf{h})\left((\omega(\mathbf{h})+b)\frac{t_k\omega(\mathbf{h})}{\text{Var}(s)} - t_k\right)\right] \\
&= \frac{t_k}{\text{Var}(s)}\mathbb{E}\left[\sigma^*(\mathbf{h})\left((\omega(\mathbf{h})+b)\,\omega(\mathbf{h}) - \text{Var}(s)\right)\right] = 0,
\end{aligned}$$

where we have used the law of iterated expectations and the fact that $\sigma^*(\mathbf{h})$ depends only on $\omega(\mathbf{h})$. Similarly, we can write

$$\mathbb{E}\left[\sigma^*(\mathbf{h})\frac{h_k}{f_k}\right] = \mathbb{E}\left[\sigma^*(\mathbf{h})\,\mathbb{E}\left[\frac{h_k}{f_k}|\omega\right]\right] = \frac{t_k}{\mathrm{Var}(s_i)}\mathbb{E}\left[\sigma^*(\mathbf{h})\,\omega\right].$$

Since $\overline{\omega} + \underline{\omega} > 0$, $\mathbb{E}\sigma^*(\mathbf{h})\,\omega(\mathbf{h}) > 0$ and thus the above expression is increasing in k. Therefore, when $r > \frac{b-\sqrt{b^2+4}}{2}$, the solution of the relaxed problem is also the optimal mechanism.

In the Online Appendix, Section D.4.1, we show that $\alpha < 0$ is not a possibility. Moreover, whenever $r \leq \frac{b-\sqrt{b^2+4}}{2}$, we must have $\underline{\omega} = \overline{\omega}$, and thus the solution of the relaxed problem as well as the optimal mechanism is $\sigma^* \equiv 0$. □

B Three-type Example

This subsection verifies: (i) the simple mechanism $\sigma^\dagger$ in (2) is optimal within the class of truthful obedient simple mechanisms; and (ii) the mechanism $\sigma^\star$ in (2) solves P. As a preliminary, we establish some useful conditions for reference. For $s_i \in \{l, m, h\}$, the interim truth-telling utility of sender s_i is

$$U(s_i) := \mathbb{E}\big[\sigma(\mathbf{s})\big(\omega(\mathbf{s}) + b\big) \mid s_i\big] = \frac{1}{3}\sum_{s_{-i}\in\{l,m,h\}} \sigma(s_i, s_{-i})\Big(\frac{s_i + s_{-i}}{\sqrt{2}} + 3\Big).$$

The probability of acceptance given report $\tilde{s}$ is:

$$\mathbb{E}[\sigma(\mathbf{s}) \mid \tilde{s}] = \frac{1}{3}\sum_{s_{-i}\in\{l,m,h\}} \sigma(\tilde{s}, s_{-i}).$$

Finally, the incentive compatibility condition (IC) in this example is equivalent to

$$U(s_i) - U(\tilde{s}) \;\geq\; \frac{s_i - \tilde{s}}{\sqrt{2}}\,\mathbb{E}[\sigma(\mathbf{s}) \mid \tilde{s}] \qquad \forall\, s_i, \hat{s} \in \{l, m, h\}. \tag{25}$$

To see how we get this formulation of incentive compatibility, notice that if a sender with signal s_i reports $\tilde{s}$, then $\omega(s_i, s_{-i}) = \omega(\tilde{s}, s_{-i}) + \frac{s_i - \tilde{s}}{\sqrt{2}}$ for every s_{-i}, so his interim payoff from the lie $\tilde{s}$ is just $U(\tilde{s}) + \frac{s_i-\tilde{s}}{\sqrt{2}}\mathbb{E}[\sigma(\mathbf{s}) \mid \tilde{s}]$.

B.1 The Optimal Simple Mechanism $\sigma^\dagger(\omega)$

To show $\sigma^\dagger(\omega)$ is optimal, we first consider a relaxed problem in which we only have to satisfy the two incentive constraints for the adjacent upward deviations l to m and m to h. We then verify that incentive compatibility holds globally for the solution to the relaxed problem. Applying (25), and rearranging, these incentive constraints are:

$$\sigma(-2) + \sigma(-1) + \sigma(0) \geq 3\sigma(1), \tag{IC-$l \to m$}$$

$$2\sigma(-1) + \sigma(0) + \sigma(1) \geq 4\sigma(2). \tag{IC-$m \to h$}$$

The receiver's expected payoff is

$$\mathbb{E}[\omega\, \sigma(\omega)] \quad = \quad \frac{1}{9}\Big(-2\sigma(-2) - 2\sigma(-1) + 2\sigma(1) + 2\sigma(2)\Big). \tag{26}$$

As $\omega = 0$ does not enter the objective, we may set $\sigma(0) = 1$ without loss (it only relaxes (IC-$l \to m$)–(IC-$m \to h$)). With $\sigma(0) = 1$, constraints (IC-$l \to m$)–(IC-$m \to h$) imply

$$\sigma(1) \leq \frac{1 + \sigma(-2) + \sigma(-1)}{3}, \qquad \sigma(2) \leq \frac{1 + \sigma(1) + 2\sigma(-1)}{4}.$$

Since the objective is increasing in $\sigma(1)$ and $\sigma(2)$, these constraints bind at an optimum. Substituting them into the objective yields (up to the constant factor $1/9$)

$$-2\sigma(-2) - 2\sigma(-1) + 2\sigma(1) + 2\sigma(2) = \frac{4}{3} - \frac{7}{6}\sigma(-2) - \frac{1}{6}\sigma(-1),$$

which is strictly decreasing in $\sigma(-2)$ and $\sigma(-1)$. Therefore $\sigma(-2) = \sigma(-1) = 0$ at an optimum for this relaxed problem, and then $\sigma(1) = \sigma(2) = 1/3$. This is exactly σ in (2).

Finally, this mechanism is incentive compatible for all deviations. First, $\mathbb{E}[\omega\, \sigma(\mathbf{s})]$ is increasing in reports and incentive compatibility binds for upward adjacent lies, so incentive compatibility is slack for downward adjacent lies. Second, a single-crossing property implies local incentive compatibility is sufficient for global.

B.2 The Mechanism $\sigma^\star$ is Optimal

Recall, it is without loss to restrict attention to symmetric mechanisms with $\sigma(s_1, s_2) = \sigma(s_2, s_1)$. Under symmetry, it is therefore without loss to use the simple notation $\sigma(\omega)$ for all $\omega(\mathbf{s}) \neq 0$. Let $\sigma(-2) := \sigma(l, l)$, $\sigma(-1) := \sigma(l, m) = \sigma(m, l)$, $\sigma(1) := \sigma(m, h) = \sigma(h, m)$, and $\sigma(2) := \sigma(h, h)$. However, where $\omega(\mathbf{s}) = 0$ we separate out the profiles into $\sigma(l, h) = \sigma(h, l)$ and $\sigma(m, m)$, which need not coincide. Notice, that this implies the receiver's expected payoff is still given by equation (26). The only entries that matter directly for the receiver payoffs are $\sigma(-2), \sigma(-1), \sigma(1), \sigma(2)$; the $\omega = 0$ entries $\sigma(l, h)$ and $\sigma(m, m)$ matter only through the incentive constraints.

We first establish an upper bound via the relaxed problem where we only satisfy feasibility and the two adjacent upward deviations l to m and m to h; then, we show the upper-bound is tight. Applying (25), and rearranging, these incentive constraints are:

$$6\sigma(m,m) \leq 3\sigma(-2) + 3\sigma(-1) + 9\sigma(l,h) - 9\sigma(1), \tag{IC$^\star l \to m$}$$

$$6\sigma(m,m) \geq 4\sigma(l,h) + 8\sigma(2) - 4\sigma(-1) - 2\sigma(1). \tag{IC$^\star m \to h$}$$

Combining these inequalities yields

$$7\sigma(1) + 8\sigma(2) \leq 3\sigma(-2) + 7\sigma(-1) + 5\sigma(l,h). \tag{27}$$

Using (27),

$$\begin{aligned}
\sigma(1) + \sigma(2) - \sigma(-1) - \sigma(-2) &= \frac{1}{7}\Big(7\sigma(1) + 7\sigma(2) - 7\sigma(-1) - 7\sigma(-2)\Big) \\
&= \frac{1}{7}\Big((7\sigma(1) + 8\sigma(2)) - \sigma(2) - 7\sigma(-1) - 7\sigma(-2)\Big) \\
&\leq \frac{1}{7}\Big(3\sigma(-2) + 7\sigma(-1) + 5\sigma(l,h) - \sigma(2) - 7\sigma(-1) - 7\sigma(-2)\Big) \\
&= \frac{1}{7}\Big(5\sigma(l,h) - 4\sigma(-2) - \sigma(2)\Big) \\
&\leq \frac{5}{7},
\end{aligned}$$

where the last inequality uses feasibility: $0 \leq \sigma(l,h) \leq 1$ and $\sigma(-2) \geq 0$, $\sigma(2) \geq 0$. In combination with equation (26), this implies $\mathbb{E}[\omega\,\sigma(\mathbf{s})] \leq \frac{2}{9} \cdot \frac{5}{7} = \frac{10}{63}$.

To see this upper bound is tight observe: 1) the payoff of $\sigma^\star$ is $\mathbb{E}[\omega\,\sigma^\star(\mathbf{s})] = \frac{2}{9}\frac{5}{7} = \frac{10}{63}$; 2) incentive compatibility constraints for upward adjacent lies are tight; 3) as $\mathbb{E}[\sigma^\star(\tilde{\mathbf{s}}) \mid \tilde{s_i}]$ is weakly increasing in reports, incentive compatibility must be slack for downward adjacent lies; and 4) a single-crossing property implies local incentive compatibility is sufficient for global incentive compatibility.

Online Appendix

C Supporting Materials

C.1 Remark 2

To see the alternative interpretation (ENV) of in Remark (2), note by Lemma (2) that such an uninformed sender believes $\mathbf{h}$ follows distribution $N(\mathbf{0}, \Sigma)$. Suppose this sender can shift the mean of $\mathbf{h}$ by an amount $\tilde{\beta}$, so that the reported NEF $\tilde{\mathbf{h}}$ will follow distribution $N(\tilde{\beta}, \Sigma)$. For this nudged distribution to correspond to a NEF, its entries must sum to 0. Thus, we consider nudges which satisfy $\tilde{\beta}_l = -\sum_{j \neq l} \tilde{\beta}_j$.

We show that (ENV) corresponds to the claimed set of first order conditions. To do so, note that if this uninformed sender chooses nudge $\tilde{\beta}$, his expected utility in a mechanism σ is $U\left(\tilde{\beta}\right) = \mathbb{E}\left[(\boldsymbol{h} \cdot \boldsymbol{t} + b)\sigma(\mathbf{h} + \tilde{\beta})\right]$. Using the change of variables $h = \tilde{\boldsymbol{h}} - \tilde{\beta}$ and plugging in the constraints $h_l = -\sum_{j \neq l} h_j$ and $\tilde{\beta}_l = -\sum_{j \neq l} \tilde{\beta}_j$, this can be re-expressed

$$U\left(\tilde{\beta}_{-l}\right) = \mathbb{E}\left[(\tilde{\boldsymbol{h}}_{-l} \cdot \hat{\boldsymbol{t}} + b)\sigma(\tilde{\mathbf{h}}_{-l})\right] - \left(\tilde{\beta}_{-l} \cdot \hat{\mathbf{t}}\right) \mathbb{E}\left[\sigma\left(\tilde{\boldsymbol{h}}_{-l}\right)\right], \tag{28}$$

where $\mathbf{h}_{-l}$, $\tilde{\boldsymbol{h}}_{-l}$ and $\tilde{\beta}_{-l}$ are $(K-1)\times 1$ sub-vectors of $\mathbf{h}$, $\tilde{\mathbf{h}}$ and $\tilde{\beta}$ whose l^{th} row has been removed, and $\hat{\mathbf{t}} = t_{-l} - t_l \mathbf{1}_{-l}$, where t_{-l} and $\mathbf{1}_{-l}$ are the corresponding $(K-1) \times 1$ sub-vectors of $\mathbf{t}$ and $\mathbf{1}$. $\mathbf{h}_{-l}$ is distributed according to $\mathcal{N}(\mathbf{0}, \hat{\Sigma})$, where $\hat{\Sigma}$ is the $(K-1) \times (K-1)$ matrix formed by deletion of the l^{th} row and column from Σ. Moreover, covariance matrix $\hat{\Sigma}$ is invertible, with entry $\alpha_{kj} = \frac{1}{f_k}\mathbf{1}(k=j) + \frac{1}{f_l}$ in the k^{th} row, and j^{th} column of $\hat{\Sigma}$.[39] We are now ready to differentiate (28) with respect to $\tilde{\beta}_k$. First consider the term

$$\mathbb{E}\left[(\tilde{\boldsymbol{h}}_{-l} \cdot \hat{\boldsymbol{t}} + b)\sigma(\tilde{\mathbf{h}}_{-l})\right] = C \int_{\mathbb{R}^K} (\tilde{\boldsymbol{h}}_{-l} \cdot \hat{\boldsymbol{t}} + b)\sigma(\tilde{\boldsymbol{h}}_{-l}) e^{-\frac{1}{2}\left(\tilde{\boldsymbol{h}}_{-l} - \tilde{\beta}_{-l}\right)^\top \hat{\Sigma}^{-1}\left(\tilde{\boldsymbol{h}}_{-l} - \tilde{\beta}_{-l}\right)} d\tilde{\boldsymbol{h}}_{-l},$$

where $C = \frac{1}{(2\pi)^{\frac{K-1}{2}} \det(\hat{\Sigma})^{\frac{1}{2}}}$ is a coefficient of the joint Normal distribution. Differentiation with respect to $\tilde{\beta}_k$, evaluated at $\tilde{\beta}_{-l} = \mathbf{0}$, yields

$$\begin{aligned} & C \int_{\mathbb{R}^K} (\tilde{\boldsymbol{h}}_{-l} \cdot \hat{\boldsymbol{t}} + b)\sigma(\tilde{\boldsymbol{h}}) e^{-\frac{1}{2}\tilde{\boldsymbol{h}}_{-l}\hat{\Sigma}^{-1}\tilde{\boldsymbol{h}}_{-l}} \times \boldsymbol{e}_k^\top \hat{\Sigma}^{-1} \tilde{\boldsymbol{h}}_{-l} d\tilde{\boldsymbol{h}}_{-l} \\ = & \mathbb{E}\left[(\boldsymbol{h}_{-K} \cdot \hat{\boldsymbol{t}} + b)\sigma(\boldsymbol{h})\left(\frac{h_k}{f_k} - \frac{h_l}{f_l}\right)\right] \end{aligned}$$

where $\mathbf{e}_k$ is a basis vector for the row in which $\tilde{\beta}_k$ enters $\tilde{\beta}$. The first line follows after differenti-

[39] Indeed, direct calculation easily verifies that $\hat{\Sigma} \cdot \hat{\Sigma}^{-1} = \hat{\Sigma}^{-1}\hat{\Sigma} = I$.

ation of the quadratic form in $\tilde{\beta}_k$, and the second from $\boldsymbol{e}_k^\top \hat{\Sigma}^{-1} \tilde{\boldsymbol{h}}_{-l} = \frac{h_k}{f_k} - \frac{h_l}{f_l}$.[40] As for the second term in (28), $-\left(\tilde{\beta}_{-l} \cdot \hat{\boldsymbol{t}}\right) \mathbb{E}\left[\sigma(\tilde{\boldsymbol{h}})\right]$, a trivial application of the product rule for differentiation shows that this differentiates to

$$-(t_k - t_l)\mathbb{E}[\sigma(\boldsymbol{h})].$$

Thus, the first order condition at $d\mathbf{h}_{-l} = 0$ is

$$\mathbb{E}\left[(\boldsymbol{h}_{-l} \cdot \hat{\boldsymbol{t}} + b)\sigma(\boldsymbol{h})\left(\frac{h_k}{f_k} - \frac{h_l}{f_l}\right)\right] - (t_k - t_l)\mathbb{E}[\sigma(\boldsymbol{h})].$$

To get (28), integrate out t_l as described in section (4).

C.2 Simple Mechanisms in the Large Economy Limit

C.2.1 If $\sigma(\mathrm{h})$ satisfies ENV, then $\bar{\sigma}(\omega)$ satisfies ω-ENV

Consider a marginal shift in the reported frequency distribution so that $\tilde{\boldsymbol{h}} = \boldsymbol{h} + \boldsymbol{\beta} d\omega$, where $\boldsymbol{\beta}$ is a vector of regression coefficients $\beta_k = \frac{\partial \mathbb{E}[h_k|\omega]}{\partial \omega} = \frac{\mathrm{cov}(h_k,\omega)}{\mathrm{var}(\omega)} = \frac{t_k f_k}{var(\omega)}$. After some algebra, we can directly evaluate the following β_k-weighted sum as:

$$\begin{aligned}
\sum_{k=1}^{K} \beta_k \mathbb{E}\left[(\omega + b)\,\sigma\left(\mathbf{h}\right)\frac{h_k}{f_k}\right] &= \mathbb{E}\left[(\omega + b)\,\sigma\left(\mathbf{h}\right)\sum_{k=1}^{K}\frac{t_k f_k}{\mathrm{var}\left(\omega\right)}\frac{h_k}{f_k}\right] \\
&= \frac{1}{\mathrm{var}\left(\omega\right)}\mathbb{E}\left[(\omega + b)\,\omega\sigma\left(\mathbf{h}\right)\right] \\
&= -\int (\omega + b)\;\mathbb{E}\left[\sigma\left(\mathbf{h}\right) \mid \omega\right]\phi'\left(\frac{\omega}{\mathrm{var}\left(\omega\right)}\right) d\omega.
\end{aligned}$$

where the final line follows after noting that the Normal density ϕ satisfies $\phi'\left(\frac{\omega}{\mathrm{var}(\omega)}\right) = -\frac{\omega}{\mathrm{var}(\omega)}\phi\left(\frac{\omega}{\mathrm{var}(\omega)}\right)$ and uses the Law of Iterated Expectations. On the other hand, evaluating the same sum using (ENV) yields

$$\begin{aligned}
\sum_{k=1}^{K} \beta_k \mathbb{E}\left[(\omega + b)\,\sigma\left(\mathbf{h}\right)\frac{h_k}{f_k}\right] &= \sum_{k=1}^{K} \beta_k t_k \mathbb{E}\left[\sigma\left(\mathbf{h}\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\sigma\left(\mathbf{h}\right) \mid \omega\right]\right],
\end{aligned}$$

where the second line uses $\sum_{k=1}^{K} t_k \beta_k = 1$ and the Law of Iterated Expectations. Equating the two derived expressions, recalling that $\mathbb{E}\left[\sigma(\mathbf{h}) \mid \omega\right]$ and comparing to the state-dependent version

[40] Observe that $\boldsymbol{e}_k^\top \hat{\Sigma}^{-1} \tilde{\boldsymbol{h}}_{-l} = \sum_{j \neq l} \alpha_{kj} \tilde{h}_j = \frac{h_k}{f_k} + \frac{1}{f_l}\sum_{j \neq l} h_j = \frac{h_k}{f_k} - \frac{h_l}{f_l}$.

(ω-ENV) of the envelope condition, establishes the claim.

C.2.2 Verification that nudge in direction β shifts $\tilde{\omega}$ but preserves conditional distributions $\tilde{\mathrm{h}} \mid \tilde{\omega}$

Though the derivation (C.2.1) already proves that $\bar{\sigma}$ satisfies (ω-ENV), described an underlying intuition based on the statistical properties of the considered nudge. Here, we briefly verify the described effects of a nudge in the direction $\tilde{\beta} = \beta dw$.

$\mathbf{h}$ and ω are jointly normal (Lemma (2)), and the conditional distribution of $\mathbf{h} \mid \omega$ is well-known to be $\mathcal{N}(\boldsymbol{\beta}\omega, \Sigma - \boldsymbol{\beta}\boldsymbol{\beta}^\top \mathrm{Var}(\omega))$. Consider the random variables $\tilde{\boldsymbol{h}} = \boldsymbol{h} + \boldsymbol{\beta} d\omega$, which distorts the reported frequencies in the direction of vector $\tilde{\beta} = \boldsymbol{\beta} d\omega$, and $\tilde{\omega} = \tilde{\boldsymbol{h}} \cdot \boldsymbol{t}$. Observe that $\sum t_k \beta_k = 1$, and therefore $\tilde{\omega} = \omega + d\omega$.

Now consider the conditional distribution of $\tilde{\boldsymbol{h}}$ given $\tilde{\omega}$. In the move from $\boldsymbol{h}$ to $\tilde{\boldsymbol{h}}$, and ω to $\tilde{\omega}$, the means of these random variables have changed in the obvious fashion, while the covariances are unaffected. The variables are still jointly Normal, and the conditional distribution is

$$\tilde{\boldsymbol{h}} \mid \tilde{\omega} \sim \mathcal{N}\left(\boldsymbol{\beta} d\omega + \boldsymbol{\beta}\left(\tilde{\omega} - d\omega\right), \Sigma - \boldsymbol{\beta}\boldsymbol{\beta}^\top \mathrm{Var}(\omega)\right) = \mathcal{N}\left(\boldsymbol{\beta}\tilde{\omega}, \Sigma - \boldsymbol{\beta}\boldsymbol{\beta}^\top \mathrm{Var}(\omega)\right).$$

Hence, given a reported state $\tilde{\omega}$, the joint distribution over $\tilde{h}$ is exactly the same as it would have been if the agent did not distort the mechanism. Of course, since the conditional distributions are unchanged, this implies that the expectations of any function of $\boldsymbol{h}$ (such as σ) are also unchanged.

C.2.3 Remark 3

Consider a mechanism σ^N that is incentive compatible. Then, consider an alternative mechanism $\bar{\sigma}^N(\hat{\omega}) = \mathbb{E}^N\left[\sigma^N(\mathbf{h}) \mid \omega = \omega(\mathbf{h})\right]$ that replaces the probability of accept for each realization $\mathbf{h}$ with the probability of accept conditional on its associated state value $\omega(\mathbf{h})$. Since the receiver only cares about ω, this mechanism provides her with the same payoffs. However, this modification of the mechanism may no longer satisfy sender incentive compatibility. Nonetheless, it has a notable property, specifically, that the effect of an individual sender's information, s_i, on the mechanism or the payoffs is an order of magnitude lower than it would be if we were simply controlling for ω.

To see this, consider the payoff of sender i, whose signal realization s_i takes a value $t_k \in S$ when he reports this signal truthfully. This is given by

$$U_k = \mathbb{E}^N\left[\sigma^N(\mathbf{h})(\omega + b) \mid s_i = t_k\right].$$

Using the law of iterated expectations, we can write the above as

$$U_k = \mathbb{E}^N \left[\mathbb{E}^N \left[\sigma^N (\mathbf{h}) | \omega, s_i = t_k \right] (\omega + b) | s_i = t_k \right].$$

We can also write

$$\begin{aligned} \mathbb{E}^N \left[\sigma^N (\mathbf{h}) | \omega, s_i = t_k \right] &= \frac{\sum_{\mathbf{h}:\mathbf{h}\cdot\mathbf{t}=\omega} \Pr^N (\mathbf{h} | s_i = t_k) \sigma^N (\mathbf{h})}{\sum_{\mathbf{h}:\mathbf{h}\cdot\mathbf{t}=\omega} \Pr^N (\mathbf{h} | s_i = t_k)} \\ &= \frac{\sum_{\mathbf{h}:\mathbf{h}\cdot\mathbf{t}=\omega} \left(\frac{h_k}{f_k \sqrt{N}} + 1 \right) \Pr^N (\mathbf{h}) \sigma^N (\mathbf{h})}{\sum_{\mathbf{h}:\mathbf{h}\cdot\mathbf{t}=\omega} \left(\frac{h_k}{f_k \sqrt{N}} + 1 \right) \Pr^N (\mathbf{h})}. \end{aligned}$$

We can view the above as a function of $\frac{1}{\sqrt{N}}$ (without considering the effect of changes of N on the probabilities, $\Pr^N$), and a Taylor approximation around $1/\sqrt{N} = 0$ implies that

$$\mathbb{E}^N \left[\sigma^N (\mathbf{h}) | \omega, s_i = t_k \right] = \bar{\sigma}^N (\omega) + \frac{1}{\sqrt{N}} \mathrm{cov}^N \left(\frac{h_k}{f_k}, \sigma^N (\mathbf{h}) | \omega \right) + O \left(\frac{1}{N} \right).$$

The above states that controlling for ω, the additional information contained in $s_i = t_k$ affects the expected probability of recommending $a = 1$ only by an order of $1/\sqrt{N}$, and thus this impact shrinks at that rate as N converges to infinity. A similar property holds for the payoff of the senders of a given type under truth-telling:

$$\begin{aligned} U_k &= \mathbb{E}^N \left[\mathbb{E}^N \left[\sigma^N (\mathbf{h}) | \omega, s_i = t_k \right] (\omega + b) | s_i = t_k \right] \\ &= \mathbb{E}^N \left[\left(\bar{\sigma}^N (\omega) + \frac{\mathrm{cov}^N \left(\frac{h_k}{f_k}, \sigma^N (\mathbf{h}) | \omega \right) + O \left(1/\sqrt{N} \right)}{\sqrt{N}} \right) (\omega + b) | s_i = t_k \right]. \\ &= \mathbb{E}^N \left[\left(\frac{h_k}{f_k \sqrt{N}} + 1 \right) \bar{\sigma}^N (\omega) (\omega + b) \right] + O (1/N) \end{aligned}$$

In arriving at the last expression, we have used the typical property of the multinomial distribution in (6). Recall that we can cast the incentive compatibility using its envelope form and write

$$U_k - U_l \geq \frac{t_k - t_l}{\sqrt{N}} \mathbb{E}^N \left[\sigma^N (\mathbf{h}) | s_i = t_l \right].$$

The right-hand side represents the extra surplus that a sender of type k is able to guarantee by pretending to be of type l. Therefore, it determines the marginal information rent captured by that sender. Applying the aforementioned logic to this incentive compatibility has a few implications as N becomes large. First, the right-hand side of the above can be replaced by $\frac{t_k - t_l}{\sqrt{N}} \mathbb{E}^N \left[\bar{\sigma}^N (\omega) \right] + O (1/N)$. Second, the left-hand side can be replaced by $\mathbb{E}^N \left[\left(\frac{h_k}{f_k \sqrt{N}} - \frac{h_l}{f_l \sqrt{N}} \right) \bar{\sigma}^N (\omega) (\omega + b) \right] +$

$O(1/N)$. In other words, the conditional mean of σ^N is the main determinant of incentive provision up to a first order, and all of its higher moments—which fully determine the distribution of σ^N—affect the incentives with a lower order of magnitude. Since as we take limit, only the first-order effects become relevant, we can simply replace σ^N with $\bar{\sigma}^N$, and the resulting limit (as $N \to \infty$) will satisfy the conditions in Theorem 1.

D Extensions

In this section, we extend our analysis in two directions. In Section D.1, we study optimal mechanisms in the large when bias is heterogeneous across senders. In Section D.2, we study the problem under more general preferences.

D.1 Heterogeneous Bias

Here, we consider an extension where the senders' bias is heterogeneous but observable. We show that the general structure of the optimal mechanisms in the large does not change.

Formally, there are M bias classes, where a sender in class $m \in \{1, \cdots, M\}$ has bias b_m. There are N_m senders in bias class m, and $N = \sum_m N_m$. The relative size of class m is $\nu_m = N_m/N$, with $\nu_1 + \cdots + \nu_M = 1$. We assume that the senders' signals are independent and distributed according to a discrete distribution. That is, for a sender i of bias class m, $s_{i,m} \in S_m = \{t_{1,m} < \cdots < t_{K,m}\}$, with $f_{k,m} = \Pr(s_{i,m} = t_{k,m})$ such that

$$\sum_{k=1}^{K} f_{k,m} t_{k,m} = 0, \sum_{k=1}^{K} f_{k,m} t_{k,m}^2 = \text{var}(s_{i,m}) = \eta.$$

Moreover, the payoff of the sender is given by

$$\left(\frac{\sum_{m=1}^{M} \sum_{i=1}^{N_m} s_{i,m}}{\sqrt{N}} + b_m \right) a = (\omega + b_m) a,$$

and the payoff of the receiver is $(\omega + r) a$. As in the main model, we can define $\mathbf{h}_m \in \mathbb{R}_0^K$ as the deviations of the sample distribution of signals among the senders of type m from the true distribution $\mathbf{f}_m$ multiplied by $\sqrt{N_m}$. By the central limit theorem, $\mathbf{h}_m \to_d \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_m)$, where the k, k' element of Σ_m is $f_{k,m}(\mathbf{1}[k = k'] - f_{k',m})$. We will refer to $\omega_m = \mathbf{h}_m \cdot \mathbf{t}_m$ as the mean of group m. Again, we focus on symmetric mechanisms, which are of the form $\sigma(\mathbf{h}_1, \cdots, \mathbf{h}_M) \in [0, 1]$. We can write

$$\omega = \frac{\sum_{m=1}^{M} \mathbf{h}_m \cdot \mathbf{t}_m \sqrt{N_m}}{\sqrt{N}} = \sum_{m=1}^{M} \omega_m \sqrt{\nu_m}.$$

The central limit theorem implies that $\omega_m \to \mathcal{N}(0,\eta)$, and by assumption, ω_m's are independent across groups. In this environment, an argument akin to that of Theorem 1 implies that the ICL holds if and only if, for all m:

$$\mathbb{E}\left[\sigma\left(\mathbf{h}_1,\cdots,\mathbf{h}_M\right)\left(\left(\omega+b_m\right)\frac{h_{k,m}}{f_{k,m}}-t_{k,m}\right)\right]=0, \tag{29}$$
$$\mathbb{E}\left[\sigma\left(\mathbf{h}_1,\cdots,\mathbf{h}_M\right)\frac{h_{k,m}}{f_{k,m}}\right]\text{is increasing in } k.$$

With this result in hand, the rest of the characterization follows that of Theorem 2. The following proposition states the main result for this extension:

Proposition 1. *The receiver-optimal ICL mechanism is a function of the sample mean for each class $\omega_m = \mathbf{h}_m \cdot \mathbf{b}_m$. Moreover, two vectors, $\lambda = (\lambda_1,\cdots,\lambda_M) \in \mathbb{R}^M$ and $\zeta = (\zeta_1,\cdots,\zeta_M) \in \mathbb{R}^M$, exist such that the optimal mechanism σ^* satisfies*

$$\sigma^*\left(\omega_1,\cdots,\omega_M\right)=1 \Leftrightarrow \omega+r+\sum_m \lambda_m\left[\omega\omega_m+b_m\omega_m-\eta\right]+\sum_m \zeta_m\omega_m \geq 0, \tag{30}$$
$$\mathbb{E}\left[\sigma^*\cdot\left(\left(\omega+b_m\right)\omega_m-\eta\right)\right]=0, \tag{31}$$
$$\zeta_m\mathbb{E}\left[\sigma^*\cdot\omega_m\right]=0, \zeta_m\geq 0, \mathbb{E}\left[\sigma^*\cdot\omega_m\right]\geq 0. \tag{32}$$

In Proposition 2, the λ_m's are the Lagrange multipliers associated with the aggregated version of the incentive compatibility, (29), in each bias class. Additionally, the ζ_m's are the Lagrange multipliers on the monotonicity constraints, which become $\mathbb{E}\left[\sigma^*\cdot\omega_m\right] \geq 0$. Thus, (32) is the complementary slackness associated with this constraint. The proof of Proposition 1 closely follows that of Theorems 1 and 2, and is relegated to the Appendix.

The variables λ,ζ that determine the region for which $\sigma = 1$ can be found by solving the system of equations defined by the incentive constraints, (31) and (32). Their existence is guaranteed by the existence of the solution of the mechanism design problem, as we show in the proof of Proposition 1.

D.2 General Preferences

In this section, we show how our main results on optimal aggregation in the large extend to more general preferences.

As before, there are N senders, and each has type s_i independently drawn from $\{t_1,\cdots,t_K\}$ with probability f_k. Again, let h_k be the k-NEF. Focusing on symmetric mechanisms, we can assume that a mechanism is a function of NEF. Hence, we let the payoffs of the senders and of the receiver satisfy

$$u_R(a,\mathbf{h}) = \begin{cases} u_r(\mathbf{h}) & a=1 \\ 0 & a=0 \end{cases}, u_S(a,\mathbf{h}) = \begin{cases} u_s(\mathbf{h}) & a=1 \\ 0 & a=0 \end{cases}. \tag{33}$$

We make the following assumption on u_s, u_r:

Assumption 1. *The payoff functions $u_r, u_s : \mathbb{R}_0^K \to \mathbb{R}$ satisfy the following properties:*

1. *They are continuous and differentiable.*
2. *The marginal value of h_k for the sender, $\frac{\partial u_s(\mathbf{h})}{\partial h_k}$, is higher for higher values of k.*
3. *There exists $p > 1$ such that $u_s(\mathbf{h}) = O\left(\|\mathbf{h}\|_K^p\right)$ and $\|\nabla u_s(\mathbf{h})\|_K = O\left(\|\mathbf{h}\|_K^p\right)$.*
4. *The functions $\left\{u_s(\mathbf{h})\frac{h_k}{f_k} - \frac{\partial u_s(\mathbf{h})}{\partial h_k}\right\}_{k=1}^K$ are linearly independent.*

In our model so far, the payoff functions are linear and thus satisfy the aforementioned assumption. The monotonicity assumption on the senders' marginal value of an additional type k sender is akin to the standard single crossing assumption used in mechanism design. The last parts of the assumption are technical ones that allow us to show the equivalent of Theorem 1 in this general setting.

The following proposition is the equivalent of Theorem 1 for arbitrary payoffs:

Proposition 2. *Suppose payoffs are given by (33) and satisfy Assumption 1. A recommendation mechanism $\sigma(\mathbf{h})$ is incentive compatible in the large if and only if there exists U such that it satisfies*

$$\mathbb{E}\left[\sigma(\mathbf{h})\left(u_s(\mathbf{h})\frac{h_k}{f_k} - \frac{\partial u_s(\mathbf{h})}{\partial h_k}\right)\right] = U, \forall k \tag{34}$$

$$\mathbb{E}\left[\sigma(\mathbf{h})\left(\frac{h_k}{f_k} - \frac{h_l}{f_l}\right)\left(\frac{\partial u_s(\mathbf{h})}{\partial h_k} - \frac{\partial u_s(\mathbf{h})}{\partial h_l}\right)\right] \geq \mathbb{E}\left[\sigma(\mathbf{h})\left(\frac{\partial^2 u_s(\mathbf{h})}{\partial h_k^2} + \frac{\partial^2 u_s(\mathbf{h})}{\partial h_l^2} - 2\frac{\partial^2 u_s(\mathbf{h})}{\partial h_l \partial h_k}\right)\right] \tag{35}$$

for all $k > l$.

Proposition 2 describes how incentive compatibility in the large is affected by having arbitrary payoffs. The first equality is the equivalent of the standard envelope condition as in (ENV). The second inequality is the standard monotonicity condition. Notably, with general payoffs, the second derivative or curvature of the payoff function is also relevant to the standard monotonicity condition.

The complications arising from arbitrary payoffs imply that a general characterization of optimal mechanisms is not so straightforward. However, once the assumption of monotonicity is not binding, the characterization of the resulting optimal mechanism is simple:

Proposition 3. *Let $\sigma^* : \mathbb{R}_0^K \to \{0,1\}$ be a mechanism that is ICL, i.e., satisfies the conditions in Proposition 2. If $\{\lambda_k\}_{k=1}^K$ exist such that*

$$\sigma^*(\mathbf{h}) = 1 \Leftrightarrow u_r(\mathbf{h}) + \sum_{k=1}^{K} \lambda_k \left(\frac{h_k}{f_k} u_s(\mathbf{h}) - \frac{\partial u_s}{\partial h_k}(\mathbf{h}) \right) \geq 0; \tag{36}$$

then σ^* is an optimal mechanism.

This condition is equivalent to the interval condition in the case of our initial payoffs.

A particularly tractable specification is when u_r and u_s are linear and satisfy the specification in Proposition 4. Notably, this model is different from our baseline example, because the senders and the receiver evaluate different signal realizations differentially. For example, in the context of voting, the receiver's preferences can represent how an uninformed voter cares about different issues, while the senders' preferences represent the interests of the differentially informed political elite in terms of the same issues. In this case, optimal aggregation can be described by the following Proposition:

Proposition 4. *Suppose that $u_r(\mathbf{h}) = \mathbf{t}_r \cdot \mathbf{h}$ and that $u_s(\mathbf{h}) = \mathbf{t}_s \cdot \mathbf{h}$ with $\mathbf{t}_s, \mathbf{t}_r \in \mathbb{R}^K$ such that*

$$\sum_k f_k t_{s,k} = \sum_k f_k t_{r,k} = 0,$$
$$\sum_k f_k t_{s,k}^2 = \sum_k f_k t_{r,k}^2 = 1, \sum_k f_k t_{r,k} t_{s,k} = \gamma$$

and $t_{s,k}, t_{r,k}$ are increasing in k. Then, the optimal mechanism σ^* is a function of $\omega_s = \mathbf{t}_s \cdot \mathbf{h}, \omega_r = \mathbf{t}_r \cdot \mathbf{h}$ and satisfies

$$\begin{aligned}\sigma^*(\omega_r, \omega_s) = 1 \Leftrightarrow \omega_r + (\lambda_r \omega_r + \lambda_s \omega_s)\, \omega_s - \gamma \lambda_r - \lambda_s \\ + \omega_r \sum_k (t_{r,k} - \gamma t_{s,k})(\zeta_k - \zeta_{k+1}) \\ + \omega_s \sum_k (t_{s,k} - \gamma t_{r,k})(\zeta_k - \zeta_{k+1}) \geq 0\end{aligned}$$

for some α_r, α_s and $\zeta_k \geq 0$ with $\zeta_1 = \zeta_{K+1} = 0$ such that

$$\mathbb{E}\left[\sigma^* \cdot \left(\omega_s^2 - 1\right)\right] = 0 \tag{37}$$
$$\mathbb{E}\left[\sigma^* \cdot (\omega_s \omega_r - \gamma)\right] = 0 \tag{38}$$

and

$$\zeta_k \left[\left(t_{s,k} - t_{s,k-1}\right) \mathbb{E}\sigma^* \cdot \left(\omega_s - \gamma\omega_r\right) + \left(t_{r,k} - t_{r,k-1}\right) \mathbb{E}\sigma^* \cdot \left(\omega_r - \gamma\omega_s\right) \right] = 0. \tag{39}$$

Proposition 4 illustrates that the optimal mechanism takes a tractable form and is only a function of the ex post payoffs of the senders and of the receiver. Moreover, (39) is the usual complementary slackness associated with the monotonicity constraint (35). Hence, much in line with standard models of mechanism design, Proposition 4 illustrates a procedure for finding the optimal mechanism. That is, one can conjecture that the monotonicity constraints are slack, i.e., $\zeta_k = 0$, and find λ_r, λ_s so that ICL constraints (37) and (38) are satisfied. If the resulting mechanism σ^* satisfies the monotonicity constraint, it is the optimum. Otherwise, a procedure akin to ironing is required to find which monotonicity constraints are binding.

The next example illustrates this result.

Example 1. Consider an example in which $K = 3$ and $u_r(\mathbf{h}) = \rho\left(-2h_1 - h_2 + 3h_3\right), u_s(\mathbf{h}) = \rho\left(-\sqrt{7}h_1 + \sqrt{7}h_3\right)$, where $\rho = \sqrt{3/14}$ and $f_1 = f_2 = f_3 = 1/3$. In this case, $\gamma = \frac{5}{2\sqrt{7}}$. Using Proposition 4, we can numerically find the values of λ_r, λ_s that satisfy the conditions in Proposition 4. Figure 4 depicts the set of ω_s, ω_r's for which $a = 1$. In this example, the agreement regions are the positive and negative quadrants, while the disagreement regions are the remaining regions. The optimal mechanism creates value for the receiver by recommending $a = 1$ when $\omega_r > 0$ and $\omega_s < 0$—the disagreement quadrant preferred by the receiver. Surplus burning is occurring in the positive and negative quadrants. In the positive quadrant, high reports of ω_s and low reports of ω_r are punished by recommending $a = 0$. In the negative quadrant, surplus burning is happening by recommending $a = 1$ when both the senders and the receiver prefer $a = 0$.

D.3 Optimality of Sender Preferred for N>1

Let each sender's signal be distributed according to a common cumulative distribution function $F(s)$, with density $f(s)$ on an interval $S = [\underline{s}, \overline{s}] \subset \mathbb{R}$, that satisfies $\int_S sf(s)\,ds = 0$. We use $\mathbf{s} = (s_1, s_2, ..., s_N)$ to denote the senders' type profile. Additionally, with a slight abuse of notation, we let $F(\mathbf{s}) = \prod_j F(s_j)$ and $F_{-i}(\mathbf{s}_{-i}) = \prod_{j\neq i} F(s_j)$ be the distribution functions of $\mathbf{s}$ and $\mathbf{s}_{-i}$, respectively, with the obvious corresponding density functions.

Proposition 5. *Let $f(\cdot)$ be a C^1 function with a finite derivative and full support, and define $\ell = \inf_{s_i \in S} f'(s_i)(\overline{s} - s_i)/f(s_i)$. Then, for any sender and receiver payoff parameters, (b, r), there exists a positive finite value $\underline{N}(b, r, \ell, \overline{s})$ such that the sender-preferred mechanism $\sigma^S(\mathbf{s}) = \mathbf{1}[\omega(\mathbf{s}) + b \geq 0]$ is a solution to (P)whenever*

$$N \leq \underline{N}(b, r, \ell, \overline{s}).$$

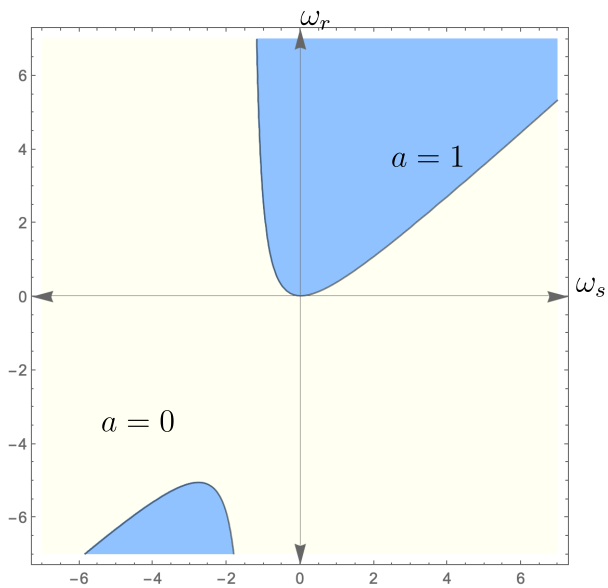

Figure 4: Optimal mechanism in Example 1. The sender's payoff ω_s is on the x-axis, and the receiver's payoff ω_r is on the y-axis.

Proposition 5 states that if the number of senders is low enough, then the best incentive compatible mechanism from the receiver's perspective is the same as that of the senders. It can be shown that the finite bound on the number of senders for the optimality of the sender-preferred mechanism, $\underline{N}(b, r, \ell, \overline{s})$, is decreasing in the bias of the sender, $b - r > 0$. Of course, for large levels of bias, even the uninformative allocation may be preferable to σ^S. Perhaps also unsurprising, when σ^S tends to the receiver first best, i.e., $b - r \to 0$, then this bound tends to infinity.[41] As we discuss in more detail later, one way to interpret $N \leq \underline{N}(b, r, \ell, \overline{s})$ is as a condition under which unrestricted private communication among the senders is indeed the best outcome from the receiver's perspective. Conversely, it identifies a necessary condition for the commitment of the mediator to be able to create value for the receiver by exploiting the informational division between the senders.

Proof. We prove the claim in two steps: First, we show that one can use standard Lagrangian techniques to characterize the solution of problem (P). Second, we construct the Lagrange multipliers – members of an appropriate dual space to be precisely defined below – so that the sender-preferred allocation, $\sigma_S^*(\mathbf{s})$, maximizes the appropriate Lagrangian.

[41]The bound is decreasing in $0 \geq \ell > -\infty$. The value of ℓ is the upper bound on the negative curvature of F. Thus, for highly negative values of ℓ, there may be a very large probability mass in the disagreement region relative to the probability of being in the agreement region. By considering the $N = 1$ case, one can see how this could easily imply that the cost of disagreement is just too high relative to the value of accepting when all parties agree.

We begin by rewriting the problem (P) as follows:

$$\max_{\sigma:S^N\to[0,1],\underline{v}} \int_{S^N} (\omega(\mathbf{s})+r)\,\sigma(\mathbf{s})\,f(\mathbf{s})\,d\mathbf{s}$$

subject to

$$\mathbb{E}\left[(\omega(\mathbf{s})+b)\,\sigma(\mathbf{s})\,|s_i\right] = \underline{v} + \frac{1}{\sqrt{N}}\int_{-1}^{s_i} \mathbb{E}\left[\sigma(\mathbf{s})\,|\tilde{s}_i\right] d\tilde{s}_i, \forall s_i \in S, i \in \{1,\cdots,N\}, \tag{40}$$

$$\mathbb{E}\left[\sigma(\mathbf{s})\,|s_i\right] \leq \mathbb{E}\left[\sigma(\mathbf{s})\,|s_i'\right], \forall s_i \leq s_i' \tag{41}$$

$$0 \leq \sigma(\mathbf{s}) \leq 1, \mathbf{s} \in S^N. \tag{42}$$

The first constraint is the integral form of the envelope version of incentive compatibility (IC), while the second constraint is the standard monotonicity of the interim allocations $\mathbb{E}\left[\sigma(\mathbf{s})\,|s_i\right]$. Finally, the last condition is the upper and lower bounds on $\sigma(\mathbf{s})$.[42] We can thus frame the problem in a convenient way for establishing weak duality or standard sufficiency conditions for optimality—see Theorem 1 in Section 8.4 in Luenberger (1997).

More specifically, let $x = (\sigma, \underline{v}) \in X = L_2\left(S^N\right) \times \mathbb{R}$ and $Z = L_2(S)^{3N} \times L_2\left(S^N\right)^2$. The above program is a linear programming problem whose constraint set is of the form $G(x) \leq 0$ for some $G : X \to Z$, where $\leq$ is defined below. The function $G(x)$ is given by the quintuple $(G_1(x), \cdots, G_5(x))$. In this quintuple, $G_1, G_2 : X \to L_2(S)^N$ and are given by

$$G_{1,i}(x)(s_i) = \mathbb{E}\left[(\omega(\mathbf{s})+b)\,\sigma(\mathbf{s})\,|s_i\right] - \underline{v} - \frac{1}{\sqrt{N}}\int_{-1}^{s_i} \mathbb{E}\left[\sigma(\mathbf{s})\,|\tilde{s}_i\right] d\tilde{s}_i \leq 0$$

$$G_{2,i}(x)(s_i) = -G_{1,i}(x)(s_i) \leq 0.$$

That is, instead of using an equality constraint in (40), we use two inequality constraints that have the same implication; $G_{3,i} : X \to L_2(S)$ is given by $G_{3,i}(x)(s_i) = -\mathbb{E}\left[\sigma(\mathbf{s})\,|s_i\right]$ (associated with (42)), and $G_4, G_5 : X \to L_2\left(S^N\right)$ are simply $G_4(x)(\mathbf{s}) = -\sigma(\mathbf{s})$ and $G_5(x)(\mathbf{s}) = 1 - \sigma(\mathbf{s})$ (associated with (41)). The relation $\leq$ is associated with the cone of members of Z whose first, second, fourth, and fifth elements are nonnegative while its third element is a non-decreasing function.

Given this formulation, we can use Theorem 1 in Section 8.4 in Luenberger (1997) (switched to maximization), which states that if there exist $x_0 \in X$ and $z_0^* \in Z^*$ such that the Lagrangian

[42] That incentive compatibility is equivalent to the envelope condition (40), and the monotonicity of $\mathbb{E}\left[\sigma(\mathbf{s})\,|s_i\right]$ is standard. See, for example, Myerson (1981).

$L(x, z^*) = w(x) + \langle z^*, G(x) \rangle$ has a saddle point at x_0, z_0^*,

$$L(x, z_0^*) \leq L(x_0, z_0^*) \leq L(x_0, z^*), \forall x \in X, z^* \geq 0.$$

Then $x_0 \in \arg\max_{x \in X, G(x) \leq 0} w(x)$. In our setting, G is as defined before, while $w(x) = \mathbb{E}[\sigma(\mathbf{s}) \omega(\mathbf{s})]$. If x_0 satisfies $\langle z_0^*, G(x_0) \rangle = 0$, then the second inequality is satisfied by the definition of the dual cone in Z^*. Thus, we have to construct the Lagrange multiplier z_0^* such that $\langle z_0^*, G(x_0) \rangle = 0$, and the first of the aforementioned inequalities holds.

Using the Riesz representation theorem—see Theorem 14.12 in Aliprantis and Border (2006)—we can write the Lagrangian as

$$\begin{aligned} L(x, z^*) = &\int (\omega(\mathbf{s}) + r) \sigma(\mathbf{s}) f(\mathbf{s}) d\mathbf{s} \\ &+ \sum_i \int_S G_{1,i}(x)(s_i) d\Lambda_i^1(s_i) + \sum_i \int_S G_{2,i}(x)(s_i) d\Lambda_i^2(s_i) \\ &+ \sum_i \int_S G_{3,i}(s_i) d\Omega_i(s_i) + \int \sigma(\mathbf{s}) d\underline{\zeta}(\mathbf{s}) + \int (1 - \sigma(\mathbf{s})) d\overline{\zeta}. \end{aligned}$$

Here, Λ_i^1, Λ_i^2's are positive Borel measures over S, and $\underline{\zeta}, \overline{\zeta}$ are positive Borel measures over S^N, since z_0^* should be a member of the dual cone of $\geq$ in Z^*. Additionally, Ω_i is a signed measure that satisfies $\Omega_i([s, \overline{s}]) \geq 0, \forall s \in S$.[43] Let $\Delta\Lambda_i = \Lambda_i^1 - \Lambda_i^2$. We can use integration by parts to write the Lagrangian as

$$\begin{aligned} L(x, z^*) = &\int (\omega(\mathbf{s}) + r) \sigma(\mathbf{s}) f(\mathbf{s}) d\mathbf{s} \\ &+ \sum_i \int_S \int_{S^{N-1}} [(\omega(s_i; s_{-i}) + b) \sigma(s_i; s_{-i}) f_{-i}(s_{-i}) ds_{-i} - \underline{v}] d\Delta\Lambda_i \\ &- \frac{1}{\sqrt{N}} \sum_i \int_S \int_{S^{N-1}} \sigma(s_i; s_{-i}) f_{-i}(s_{-i}) ds_{-i} \Delta\Lambda_i([s_i, \overline{s}]) ds_i \\ &+ \int \sigma(\mathbf{s}) d\underline{\zeta}(\mathbf{s}) + \int (1 - \sigma(\mathbf{s})) d\overline{\zeta}(\mathbf{s}) + \sum_i \int_S \int_{S^{N-1}} \sigma(s_i; s_{-i}) f_{-i}(s_{-i}) ds_{-i} d\Omega_i(s_i). \end{aligned}$$

We are now ready to identify the relevant Lagrange multipliers and verify that—under the upper bound on N—they support the sender-preferred allocation $x_0 = \left(\sigma^S(\mathbf{s}), v^S(\underline{s})\right)$ as a saddle point

[43] Recall that the dual cone of increasing functions is the set of signed Borel measures Ω so that $\int f(s_i) d\Omega \geq 0$ for all increasing f. This coincides with the set of measures that satisfy $\Omega([s, \overline{s}]) \geq 0$.

of L. To that end, set

$$\Delta\Lambda_i\left([\underline{s}, s_i]\right) = \begin{cases} 0 & s_i = \underline{s} \\ f\left(s_i\right) \frac{b-r}{b+\sqrt{N}\overline{s}}\left(\overline{s} - s_i\right) & s_i > \underline{s} \end{cases}$$
$$\Omega_i\left([\underline{s}, s_i]\right) = 0.$$

By the Jordan decomposition theorem, any signed measure can be written as the difference of two positive measures. Therefore, consider two positive measures Λ_i^1, Λ_i^2 such that $\Delta\Lambda_i = \Lambda_i^1 - \Lambda_i^2$. Given the aforementioned Lagrangian, let us define the "first order condition" as follows:[44]

$$\begin{aligned}\forall i, s_i > \underline{s} : \lambda\left(\mathbf{s}\right) = &\left(\omega\left(\mathbf{s}\right) + r\right) f\left(\mathbf{s}\right) \\ &+ \sum_{i=1}^{N}\left(\omega\left(\mathbf{s}\right) + b\right) \frac{f\left(\mathbf{s}\right)}{f\left(s_i\right)} \frac{b-r}{b+\sqrt{N}\overline{s}} \frac{d}{ds_i}\left(f\left(s_i\right)\left(\overline{s} - s_i\right)\right) \\ &+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} f\left(\mathbf{s}\right) \frac{b-r}{b+\sqrt{N}\overline{s}}\left(\overline{s} - s_i\right).\end{aligned}$$

Moreover, we can write

$$\mathbf{s}, \exists i, s_i = \underline{s} : \frac{\lambda\left(\mathbf{s}\right)}{\left(\omega\left(\mathbf{s}\right) + b\right) f\left(\mathbf{s}\right)} = \frac{b-r}{b+\sqrt{N}\overline{s}}\left(\overline{s} - \underline{s}\right) \sum_{j=1}^{N} \mathbf{1}\left[s_j = \underline{s}\right].$$

This holds since Λ_i has a mass point at $\underline{s}$, and when calculating the Fréchet–Gateaux derivative of the Lagrangian, the effect of changes in the direction of the Dirac's delta at s, only mass points are relevant.

In order to show optimality of sender-preferred allocation, it is sufficient to show that

$$\lambda\left(\mathbf{s}\right) \geq 0 \iff \omega\left(\mathbf{s}\right) + b \geq 0.$$

[44] Formally, $\lambda\left(\mathbf{s}\right)$ is the Fréchet derivative of $L\left(x, z\right)$ evaluated at x^* and in the direction of the Dirac delta function at s.

When $\forall i, s_i > \underline{s}$, we have

$$\begin{aligned}
s_i > \underline{s} : \lambda(\mathbf{s}) &= (\omega(\mathbf{s}) + r) f(\mathbf{s}) \\
&\quad + \frac{b-r}{b+\sqrt{N}\overline{s}} \sum_{i=1}^{N} (\omega(\mathbf{s}) + b) f(\mathbf{s}) \left(-1 + \frac{f'(s_i)(\overline{s} - s_i)}{f(s_i)}\right) \\
&\quad + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} f(\mathbf{s}) \frac{b-r}{b+\sqrt{N}\overline{s}} (\overline{s} - s_i) \\
&= (\omega(\mathbf{s}) + b) f(\mathbf{s}) + (r-b) f(\mathbf{s}) \\
&\quad + (\omega(\mathbf{s}) + b) f(\mathbf{s}) \frac{b-r}{b+\sqrt{N}\overline{s}} \sum_{i=1}^{N} \left(-1 + \frac{f'(s_i)(\overline{s} - s_i)}{f(s_i)}\right) \\
&\quad + f(\mathbf{s}) \frac{b-r}{b+\sqrt{N}\overline{s}} \sum_{i=1}^{N} \frac{(\overline{s} - s_i)}{\sqrt{N}} \\
&= (\omega(\mathbf{s}) + b) f(\mathbf{s}) \left[1 + \frac{b-r}{b+\sqrt{N}\overline{s}} \sum_{i=1}^{N} \left(-1 + \frac{f'(s_i)(\overline{s} - s_i)}{f(s_i)}\right)\right] \\
&\quad + (r-b) f(\mathbf{s}) + f(\mathbf{s}) \frac{b-r}{b+\sqrt{N}\overline{s}} \left(\sqrt{N}\overline{s} - \omega(\mathbf{s})\right) \\
&= \left[1 - \frac{b-r}{b+\sqrt{N}\overline{s}} + \frac{b-r}{b+\sqrt{N}\overline{s}} \times \sum_{i=1}^{N} \left(-1 + \frac{f'(s_i)(\overline{s} - s_i)}{f(s_i)}\right)\right] \times \\
&\quad (\omega(\mathbf{s}) + b) f(\mathbf{s}) + (r-b) f(\mathbf{s}) + f(\mathbf{s}) \frac{b-r}{b+\sqrt{N}\overline{s}} \left(\sqrt{N}\overline{s} + b\right) \\
&= \frac{b-r}{b+\sqrt{N}\overline{s}} \sum_{i=1}^{N} \left(-1 + \frac{f'(s_i)(\overline{s} - s_i)}{f(s_i)} + \frac{1}{N} \frac{r+\sqrt{N}\overline{s}}{b-r}\right) \times . \\
&\quad (\omega(\mathbf{s}) + b) f(\mathbf{s})
\end{aligned}$$

This implies that if the term in the last brackets is always positive, $\lambda(\mathbf{s}) \geq 0$ if and only $\omega(\mathbf{s}) + b \geq 0$. Note that the term in the brackets is always positive if and only if

$$\frac{1}{N} \frac{r+\sqrt{N}\overline{s}}{b-r} \geq 1 - \inf_{s_i \in S} \frac{f'(s_i)(\overline{s} - s_i)}{f(s_i)} = 1 - \ell \rightarrow r + \sqrt{N}\overline{s} \geq N(b-r)(1-\ell)$$

The above is a quadratic equation in $\sqrt{N}$ and since $1 - \ell > 0$, it should hold when $\sqrt{N}$ is below its higher root. The squared value of this higher root is given by $\underline{N}(b, r, \ell, \overline{s}) = \left(\frac{\sqrt{\overline{s}^2 + 4r(b-r)(1-\ell)} + \overline{s}}{2(b-r)(1-\ell)}\right)^2$ which gives us the (sufficient) bound on N.

When for some i, $s_i = \underline{s}$, then from the calculation of L, it is clear that $\lambda(\mathbf{s}) \geq 0$ if and only if $\omega(\mathbf{s}) + b \geq 0$. This concludes the proof. □

D.4 Other Proofs

D.4.1 Existence of the Multiplier in Theorem 2

Proof. Here, we show that when $r > \frac{b-\sqrt{b^2+4\text{Var}(s)}}{2}$ there is a unique α for which

$$\int_{D_\alpha}\left(1-\frac{\omega\left(\omega+b\right)}{\text{Var}\left(s\right)}\right)\phi\left(\frac{\omega}{\sqrt{\text{Var}\left(s\right)}}\right)d\omega=0$$
$$\left\{\omega|\psi\left(\omega,\alpha\right)=\omega+r+\alpha\text{Var}\left(s\right)-\alpha\left(\omega+b\right)\omega\geq 0\right\}=D_\alpha$$

Moreover $\alpha\in(0,1/b)$ which implies that $D_\alpha=[\underline{\omega},\overline{\omega}]$ where

$$-b<\underline{\omega}<-r<-\underline{\omega}<\overline{\omega}.$$

Additionally, we show that if $r\leq\frac{b-\sqrt{b^2+4}}{2}$, then whenever the above holds, $\underline{\omega}=\overline{\omega}$. We do so via a series of claims.

Claim 1. *For any value of* α, $\int_{D_\alpha}\left(1-\frac{\omega(\omega+b)}{Var(s)}\right)\phi\left(\frac{\omega}{\sqrt{Var(s)}}\right)d\omega$ *is increasing in* α.

To prove this claim, note that if $\alpha<0$, then either $\psi\left(\omega,\alpha\right)$ is always positive or at most two values of $\underline{\omega},\overline{\omega}$ exist such that $\psi\left(\underline{\omega},\alpha\right)=\psi\left(\overline{\omega},\alpha\right)=0$ in which case $D_\alpha=\mathbb{R}\backslash\left(\underline{\omega},\overline{\omega}\right)$. On the other hand, if $\alpha>0$, either $D_\alpha=[\underline{\omega},\overline{\omega}]$ with $\psi\left(\underline{\omega},\alpha\right)=\psi\left(\overline{\omega},\alpha\right)=0$ or $D_\alpha=\emptyset$. Evidently, if $\psi\left(\omega,\alpha\right)$ does not have a zero, then the integral is 0.

Note that when the zeros exist, they are given by

$$\underline{\omega}=\frac{1-b\alpha}{2\alpha}-\frac{1}{2\alpha}\sqrt{\left(1-b\alpha\right)^2+4r\alpha+4\alpha^2\text{Var}\left(s\right)}=\frac{1-\alpha b-\sqrt{\Delta}}{2\alpha}\tag{43}$$
$$\overline{\omega}=\frac{1-\alpha b}{2\alpha}+\frac{1}{2\alpha}\sqrt{\left(1-b\alpha\right)^2+4r\alpha+4\alpha^2\text{Var}\left(s\right)}=\frac{1-\alpha b+\sqrt{\Delta}}{2\alpha}\tag{44}$$

We thus have that

$$\frac{d\overline{\omega}}{d\alpha}=-\frac{\psi_\alpha\left(\overline{\omega},\alpha\right)}{\psi_\omega\left(\overline{\omega},\alpha\right)}=-\frac{\text{Var}\left(s\right)-\overline{\omega}\left(\overline{\omega}+b\right)}{1-\alpha\left(2\overline{\omega}+b\right)}=-\frac{-\frac{\overline{\omega}+r}{\alpha}}{-\sqrt{\Delta}}=-\frac{\overline{\omega}+r}{\alpha\sqrt{\Delta}}$$
$$\frac{d\underline{\omega}}{d\alpha}=-\frac{\psi_\alpha\left(\underline{\omega},\alpha\right)}{\psi_\omega\left(\underline{\omega},\alpha\right)}=-\frac{\text{Var}\left(s\right)-\underline{\omega}\left(\underline{\omega}+b\right)}{1-\alpha\left(2\underline{\omega}+b\right)}=-\frac{-\frac{\underline{\omega}+r}{\alpha}}{\sqrt{\Delta}}=\frac{\underline{\omega}+r}{\alpha\sqrt{\Delta}}$$

When $\alpha > 0, \overline{\omega} \geq \underline{\omega}$ and we have

$$\frac{d}{d\alpha}\int_{D_\alpha}\left[1-\frac{\omega\left(\omega+b\right)}{\text{Var}\left(s\right)}\right]\phi\left(\frac{\omega}{\sqrt{\text{Var}\left(s\right)}}\right)d\omega =$$

$$\frac{d}{d\alpha}\int_{\underline{\omega}}^{\overline{\omega}}\left[1-\frac{\omega\left(\omega+b\right)}{\text{Var}\left(s\right)}\right]\phi\left(\frac{\omega}{\sqrt{\text{Var}\left(s\right)}}\right)d\omega =$$

$$\phi\left(\frac{\overline{\omega}}{\sqrt{\text{Var}\left(s\right)}}\right)\frac{d\overline{\omega}}{d\alpha}\left[1-\frac{\overline{\omega}\left(\overline{\omega}+b\right)}{\text{Var}\left(s\right)}\right]-\phi\left(\frac{\underline{\omega}}{\sqrt{\text{Var}\left(s\right)}}\right)\frac{d\underline{\omega}}{d\alpha}\left[1-\frac{\underline{\omega}\left(\underline{\omega}+b\right)}{\text{Var}\left(s\right)}\right] =$$

$$\phi\left(\frac{\overline{\omega}}{\sqrt{\text{Var}\left(s\right)}}\right)\left(-\frac{\overline{\omega}+r}{\alpha\sqrt{\Delta}}\right)\left(-\frac{\overline{\omega}+r}{\alpha\text{Var}\left(s\right)}\right)-\phi\left(\frac{\underline{\omega}}{\sqrt{\text{Var}\left(s\right)}}\right)\left(\frac{\underline{\omega}+r}{\alpha\sqrt{\Delta}}\right)\left(-\frac{\underline{\omega}+r}{\alpha\text{Var}\left(s\right)}\right) =$$

$$\frac{1}{\sqrt{\Delta}\text{Var}\left(s\right)}\left[\phi\left(\frac{\overline{\omega}}{\sqrt{\text{Var}\left(s\right)}}\right)\left(\frac{\overline{\omega}+r}{\alpha}\right)^2+\phi\left(\frac{\underline{\omega}}{\sqrt{\text{Var}\left(s\right)}}\right)\left(\frac{\underline{\omega}+r}{\alpha}\right)^2\right] > 0$$

where the above holds since both terms are squares multiplied by a density. Similarly when $\alpha < 0, \overline{\omega} < \underline{\omega}$ and thus

$$\int_{D_\alpha}\left[1-\frac{\omega\left(\omega+b\right)}{\text{Var}\left(s\right)}\right]\phi\left(\frac{\omega}{\sqrt{\text{Var}\left(s\right)}}\right)d\omega =$$

$$\int_{\mathbb{R}\setminus(\overline{\omega},\underline{\omega})}\left[1-\frac{\omega\left(\omega+b\right)}{\text{Var}\left(s\right)}\right]\phi\left(\frac{\omega}{\sqrt{\text{Var}\left(s\right)}}\right)d\omega =$$

$$\int_{\underline{\omega}}^{\overline{\omega}}\left[1-\frac{\omega\left(\omega+b\right)}{\text{Var}\left(s\right)}\right]\phi\left(\frac{\omega}{\sqrt{\text{Var}\left(s\right)}}\right)d\omega$$

where in the above we have used the fact that the integral over the entire real line is 0. Then a logic similar to the case of $\alpha > 0$ implies that the above is increasing.

Claim 2. *For all values* $\alpha < 0$, $\int_{D_\alpha}\left[1-\frac{\omega(\omega+b)}{Var(s)}\right]\phi\left(\frac{\omega}{\sqrt{Var(s)}}\right)d\omega < 0.$

To show this, it is sufficient to show that it holds as $\alpha \to 0$ from below. Note that as $\alpha \nearrow 0$, $\underline{\omega} \to -r$ and $\overline{\omega} \to -\infty$. Hence,

$$\int_{D_\alpha}\left[1-\frac{\omega\left(\omega+b\right)}{\text{Var}\left(s\right)}\right]\phi\left(\frac{\omega}{\sqrt{\text{Var}\left(s\right)}}\right)d\omega =$$

$$\int_{-r}^{-\infty}\left[1-\frac{\omega\left(\omega+b\right)}{\text{Var}\left(s\right)}\right]\phi\left(\frac{\omega}{\sqrt{\text{Var}\left(s\right)}}\right)d\omega =$$

$$-\int_{-\infty}^{-r}\left[1-\frac{\omega\left(\omega+b\right)}{\text{Var}\left(s\right)}\right]\frac{e^{-\frac{\omega^2}{2\text{Var}(s)}}}{\sqrt{2\pi}}d\omega = -\left(\omega+b\right)\frac{e^{-\frac{\omega^2}{2\text{Var}(s)}}}{\sqrt{2\pi}}\Bigg|_{-\infty}^{-r} = -\left(b-r\right)\frac{e^{-\frac{r^2}{2\text{Var}(s)}}}{\sqrt{2\pi}} < 0$$

which proves the claim.

Claim 3. *If* $r > \frac{b-\sqrt{b^2+4Var(s)}}{2}$, *then* $\exists!\alpha \in (0,1/b)$ *such that* $\int_{D_\alpha}\left[1-\frac{\omega(\omega+b)}{Var(s)}\right]\phi\left(\omega/\sqrt{Var(s)}\right)d\omega = 0$. *At this value of* α, $-b < \underline{\omega} < -r < \overline{\omega}$ *and* $0 < \underline{\omega}+\overline{\omega}$.

Note that as $\alpha \searrow 0, \underline{\omega} \to -r$ and $\overline{\omega} \to \infty$. Hence and as above, $\int_{D_\alpha}\left[1-\frac{\omega(\omega+b)}{\mathrm{Var}(s)}\right]\phi\left(\omega/\sqrt{\mathrm{Var}(s)}\right)d\omega < 0$ holds for $\alpha = 0$. In contrast, at $\alpha = 1/b$,

$$\underline{\omega} = -\frac{b}{2}\sqrt{4r/b + 4\mathrm{Var}(s)/b^2} = -\sqrt{rb+\mathrm{Var}(s)}$$
$$\overline{\omega} = \sqrt{rb+\mathrm{Var}(s)}$$

At these values

$$\begin{aligned}\int_{D_\alpha}\left[1-\frac{\omega(\omega+b)}{\mathrm{Var}(s)}\right]\phi\left(\omega/\sqrt{\mathrm{Var}(s)}\right)d\omega &= \int_{-\sqrt{rb+\mathrm{Var}(s)}}^{\sqrt{rb+\mathrm{Var}(s)}}\left[1-\frac{\omega(\omega+b)}{\mathrm{Var}(s)}\right]\phi\left(\omega/\sqrt{\mathrm{Var}(s)}\right)d\omega\\ &= \left(\sqrt{rb+\mathrm{Var}(s)}+b\right)\frac{e^{-\frac{rb+\mathrm{Var}(s)}{2\mathrm{Var}(s)}}}{\sqrt{2\pi}}\\ &\quad -\left(b-\sqrt{rb+\mathrm{Var}(s)}\right)\frac{e^{-\frac{rb+\mathrm{Var}(s)}{2\mathrm{Var}(s)}}}{\sqrt{2\pi}}\\ &=2\sqrt{rb+\mathrm{Var}(s)}\frac{e^{-\frac{rb+\mathrm{Var}(s)}{2\mathrm{Var}(s)}}}{\sqrt{2\pi}} > 0\end{aligned}$$

Hence, there exists a unique $\alpha \in (0,1/b)$ for which $\int_{D_\alpha}\left[1-\frac{\omega(\omega+b)}{\mathrm{Var}(s)}\right]\phi\left(\omega/\sqrt{\mathrm{Var}(s)}\right)d\omega = 0$. Since $\underline{\omega},\overline{\omega}$ satisfy (43) and (44), their sum is $\frac{1-\alpha b}{\alpha} > 0$. Moreover, since $\Delta = (1-b\alpha)^2 + 4r\alpha + 4\alpha^2\mathrm{Var}(s)$, we can sign $\underline{\omega},\overline{\omega}$ by signing Δ. The expression for Δ is a quadratic function of α whose discriminant is $(4r-2b)^2 - 4\left(b^2+4\mathrm{Var}(s)\right) = 16\left(r^2-rb-\mathrm{Var}(s)\right)$. When $b > r > \frac{b-\sqrt{b^2+4\mathrm{Var}(s)}}{2}$, this expression is always negative which means that for all values of α, $\Delta > 0$. Since $\alpha < 1/b$, we must have that $\overline{\omega} > 0 > \underline{\omega}$. Finally, $\psi(-r,\alpha) = \alpha\left(\mathrm{Var}(s)+rb-r^2\right)$. When $b > r > \frac{b-\sqrt{b^2+4\mathrm{Var}(s)}}{2}$, this expression is positive which means that $\underline{\omega} < -r < \overline{\omega}$.

Claim 4. *If* $r \leq \frac{b-\sqrt{b^2+4Var(s)}}{2}$, *then* $\exists\alpha > 0$ *such that* $\int_{D_\alpha}\left[1-\frac{\omega(\omega+b)}{Var(s)}\right]\phi\left(\omega/\sqrt{Var(s)}\right)d\omega = 0$. *For all such values of* α, $\underline{\omega} = \overline{\omega}$.

The proof of this claim follows from the same argument as before together with the fact that when $r < \frac{b-\sqrt{b^2+4\mathrm{Var}(s)}}{2}$, there is an $1/b > \alpha > 0$ for which $\Delta = (1-b\alpha)^2+4r\alpha+4\alpha^2\mathrm{Var}(s) = 0$. To see this, note that the values of α for which this holds are given by

$$\alpha = \frac{b-2r\pm2\sqrt{r^2-rb-\mathrm{Var}(s)}}{b^2+4\mathrm{Var}(s)}$$

Since $r \leq \frac{b-\sqrt{b^2+4\text{Var}(s)}}{2}$, we must have that $2r \leq b-\sqrt{b^2+4\text{Var}\,(s)}$ which implies that the lower root above is positive. Since as we have argued before, $\int_{D_\alpha}\left[1-\frac{\omega(\omega+b)}{\text{Var}(s)}\right]\phi\left(\omega/\sqrt{\text{Var}\,(s)}\right)d\omega$ is strictly increasing in α whenever $\alpha > 0$ and $\underline{\omega}, \overline{\omega}$ are real and at the above values of α, $\underline{\omega} = \overline{\omega}$ and $\int_{D_\alpha}\left[1-\frac{\omega(\omega+b)}{\text{Var}(s)}\right]\phi\left(\omega/\sqrt{\text{Var}\,(s)}\right)d\omega = 0$, this establishes the claim. □

D.5 Extensions: Proofs

Heterogeneous Biases: Proof of Proposition 1

Proof. The fact that ICL is equivalent to 1 follows a similar existence proof as of that of Theorem 1. Such an extension is possible because the number of bias types and signal realizations is finite. Hence, we can apply the same technique – consider a linear operator for all signal realization in each bias group and show the existence of a mechanism for any finite number of senders.

Now, consider the problem of choosing $\sigma(\mathbf{h}_1,\cdots,\mathbf{h}_M)$ to maximize the receiver's payoff $\mathbb{E}\left[\sigma(\mathbf{h}_1,\cdots,\mathbf{h}_M)(\omega+r)\right]$ subject to the ICL requirements:

$$\mathbb{E}\left[\sigma(\mathbf{h}_1,\cdots,\mathbf{h}_M)\left(\frac{h_{k,m}}{f_{k,m}}(\omega+b_m)-t_{k,m}\right)\right]=0$$
$$\mathbb{E}\left[\sigma(\mathbf{h}_1,\cdots,\mathbf{h}_M)\frac{h_{k,m}}{f_{k,m}}\right]\geq\mathbb{E}\left[\sigma(\mathbf{h}_1,\cdots,\mathbf{h}_M)\frac{h_{k-1,m}}{f_{k-1,m}}\right], k>1$$

By multiplying the top condition by $f_{k,m}t_{k,m}$ and summing over k for a fixed m and using the fact that $\sum_k f_{k,m}t_{k,m}^2=\eta$, we arrive at

$$\mathbb{E}\left[\sigma(\mathbf{h}_1,\cdots,\mathbf{h}_M)(\omega_m(\omega+b_m)-\eta)\right]=0 \tag{45}$$

Moreover, since $t_{k,m}$ and $\mathbb{E}\left[\sigma\frac{h_{k,m}}{f_{k,m}}\right]$ are both increasing in k, their covariance has to be positive. In other words,

$$\sum_k f_{k,m}t_{k,m}\mathbb{E}\left[\sigma\frac{h_{k,m}}{f_{k,m}}\right]\geq\sum_k f_{k,m}t_{k,m}\sum_k f_{k,m}\mathbb{E}\left[\sigma\frac{h_{k,m}}{f_{k,m}}\right]$$

The left hand side of the above is $\mathbb{E}[\sigma\omega_m]$ while the right hand side is 0. Hence, we must have that for all m, $\mathbb{E}[\sigma\omega_m]\geq 0$.

In other words, if a mechanism is ICL, then it must satisfy (45) and $\mathbb{E}[\sigma\omega_m]\geq 0$. One can thus focus on the relaxed problem of maximizing $\mathbb{E}[\sigma\cdot(\omega+r)]$ subject to (45) and $\mathbb{E}[\sigma\cdot\omega_m]\geq 0$. In this relaxed problem, all constraint only depend on $(\mathbf{h}_1,\cdots,\mathbf{h}_M)$ via $\omega_1,\cdots,\omega_M$ which implies that the solution should only depend on $\omega_1,\cdots,\omega_M$. Additionally, σ^* is a solution to this relaxed problem if and only if multipliers $\lambda_1,\cdots,\lambda_M$ associated with (45) and $\zeta_m\geq 0$ associated with

$\mathbb{E}[\sigma \cdot \omega_m] \geq 0$ exists that satisfy the conditions provided in statement of the proposition.

It therefore remains to be shown that if σ^* solves the relaxed problem, then it is indeed ICL. To see this, Since $\mathbf{h}_m \sim \mathcal{N}(0, \Sigma_m)$ and $h_{k,m}$'s are independent across different bias groups, we must have that

$$\mathbb{E}[h_{k,m}|\omega_1, \cdots, \omega_M] = \mathbb{E}[h_{k,m}|\omega_m] = \frac{\mathbb{E}[h_{k,m}\omega_m]}{\mathbb{E}[\omega_m^2]}\omega_m = \frac{f_{k,m}t_{k,m}}{\eta}\omega_m$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left[\sigma^* \cdot \left(\frac{h_{k,m}}{f_{k,m}}(\omega + b_m) - t_{k,m}\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[\sigma^* \cdot \frac{h_{k,m}}{f_{k,m}}(\omega + b_m)|\omega_1, \cdots, \omega_M\right] - \sigma^* \cdot t_{k,m}\right] \\
&= \mathbb{E}\left[\sigma^* \cdot (\omega + b_m)\,\mathbb{E}\left[\frac{h_{k,m}}{f_{k,m}}|\omega_1, \cdots, \omega_M\right] - \sigma^* \cdot t_{k,m}\right] \\
&= \mathbb{E}\left[\sigma^* \cdot (\omega + b_m)\frac{t_{k,m}}{\eta}\omega_m - \sigma^* \cdot t_{k,m}\right] \\
&= \frac{t_{k,m}}{\eta}\mathbb{E}[\sigma^* \cdot (\omega + b_m)\,\omega_m - \eta\sigma^*] = 0
\end{aligned}$$

where in the above we have used the law of iterated expectations and that σ^* is the solution to the relaxed problem. Similarly,

$$\begin{aligned}
\mathbb{E}\left[\sigma^* \cdot \frac{h_{k,m}}{f_{k,m}}\right] &= \mathbb{E}\left[\sigma^* \cdot \mathbb{E}\left[\frac{h_{k,m}}{f_{k,m}}|\omega_1, \cdots, \omega_M\right]\right] \\
&= \frac{t_{k,m}}{\eta}\mathbb{E}[\sigma^* \cdot \omega_m]
\end{aligned}$$

Since σ^* is the solution to the relaxed problem, it must satisfy $\mathbb{E}[\sigma^* \cdot \omega_m]$ which implies that the above is increasing in k. Therefore, σ^* is ICL. □

D.5.1 General Preferences: Proof of Proposition 2

Proof. Similar to the first part of Theorem 1, we can write

$$U_{l,k}^N = \sum_{\mathbf{s}_- \in S^{N-1}} \Pr_{N-1}(\mathbf{s}_-)\,\sigma^N\left(\mathbf{h}^N(\mathbf{s}_- + \mathbf{e}_l)\right) u_s\left(\mathbf{h}^N(\mathbf{s}_- + \mathbf{e}_k)\right) \tag{46}$$

where in the above $\mathbf{e}_k$ is a K−dimensional vector whose elements are 0 except for its k-th element which is 1

Consider a vector of realizations $\mathbf{s}_- \in S^{N-1}$ for which $\mathbf{h}^N(\mathbf{s}_- + \mathbf{e}_k) = \mathbf{h}$. In this case, the count of s_i's which are of type m is given by $n_m = \sqrt{N}h_m + Nf_m$ with $\sqrt{N}h_k + Nf_k \geq 1$. Then probability of this occurring – using the multi–nomial distribution – is given by

$$\begin{aligned}
\Pr_{N-1}(\mathbf{s}_-) &= \binom{N-1}{n_1,\cdots,n_k-1,\cdots,n_K} f_1^{n_1}\cdots f_k^{n_k-1}\cdots f_K^{n_K} \\
&= \binom{N}{n_1,\cdots,n_K} f_1^{n_1}\cdots f_K^{n_K}\frac{n_k}{f_k N} \\
&= \Pr_N(\mathbf{s}_- + \mathbf{e}_k)\frac{h_k\sqrt{N}+Nf_k}{f_k N}
\end{aligned}$$

Using this adjustment of the probabilities, we can write

$$\begin{aligned}
U_{l,k}^N &= \sum_{\mathbf{s}_-\in S^{N-1}} \Pr_{N-1}(\mathbf{s}_-)\,\sigma^N\left(\mathbf{h}^N(\mathbf{s}_-+\mathbf{e}_l)\right)u_s\left(\mathbf{h}^N(\mathbf{s}_-+\mathbf{e}_k)\right) \\
&= \sum_{\mathbf{s}_-\in S^{N-1}} \Pr_{N-1}(\mathbf{s}_-)\,\sigma^N\left(\mathbf{h}^N(\mathbf{s}_-+\mathbf{e}_l)\right)u_s\left(\mathbf{h}^N(\mathbf{s}_-+\mathbf{e}_l)\right) + \\
&\quad \sum_{\mathbf{s}_-\in S^{N-1}} \Pr_{N-1}(\mathbf{s}_-)\,\sigma^N\left(\mathbf{h}^N(\mathbf{s}_-+\mathbf{e}_l)\right)\left[u_s\left(\mathbf{h}^N(\mathbf{s}_-+\mathbf{e}_k)\right)-u_s\left(\mathbf{h}^N(\mathbf{s}_-+\mathbf{e}_l)\right)\right] \\
&= U_{l,l}^N + \sum_{\mathbf{s}_-\in S^{N-1}} \Pr_{N-1}(\mathbf{s}_-)\,\sigma^N\left(\mathbf{h}^N(\mathbf{s}_-+\mathbf{e}_l)\right)\left[u_s\left(\mathbf{h}^N(\mathbf{s}_-+\mathbf{e}_k)\right)-u_s\left(\mathbf{h}^N(\mathbf{s}_-+\mathbf{e}_l)\right)\right]
\end{aligned}$$

Let $\mu^N(\mathbf{h})$ be the probability of the adjusted frequencies being equal to $\mathbf{h}$. If $\mathbf{h} = \mathbf{h}^N(\mathbf{s})$, then it must be that $\mu^N(\mathbf{h}) = \Pr_N(\mathbf{s})$. Let us also define $H^N \subset \mathbb{R}_0^K$ to be the support of μ^N. We can thus write the above as

$$\begin{aligned}
U_{l,k}^N =& U_{l,l}^N + \sum_{\mathbf{s}_-\in S^{N-1}} \Pr_{N-1}(\mathbf{s}_-)\,\sigma^N\left(\mathbf{h}^N(\mathbf{s}_-+\mathbf{e}_l)\right)\left[u_s\left(\mathbf{h}^N(\mathbf{s}_-+\mathbf{e}_k)\right)-u_s\left(\mathbf{h}^N(\mathbf{s}_-+\mathbf{e}_l)\right)\right] \\
=& U_{l,l}^N + \sum_{\mathbf{s}_-\in S^{N-1}} \Pr_N(\mathbf{s}_-+\mathbf{e}_l)\frac{h_l^N(\mathbf{s}_-+\mathbf{e}_l)\sqrt{N}+Nf_l}{f_l N}\sigma^N\left(\mathbf{h}^N(\mathbf{s}_-+\mathbf{e}_l)\right)\times \\
&\qquad\qquad \left[u_s\left(\mathbf{h}^N(\mathbf{s}_-+\mathbf{e}_k)\right)-u_s\left(\mathbf{h}^N(\mathbf{s}_-+\mathbf{e}_l)\right)\right] \\
=& U_{l,l}^N + \sum_{\mathbf{h}\in H^N, h_l\geq\frac{1-f_l\sqrt{N}}{N}} \mu^N(\mathbf{h})\,\sigma^N(\mathbf{h})\frac{h_l\sqrt{N}+Nf_l}{f_l N}\left[u_s\left(\mathbf{h}+\frac{\mathbf{e}_k-\mathbf{e}_l}{\sqrt{N}}\right)-u_s(\mathbf{h})\right] \\
=& U_{l,l}^N + \sum_{\mathbf{h}\in H^N, h_l\geq\frac{1-f_l\sqrt{N}}{N}} \mu^N(\mathbf{h})\,\sigma^N(\mathbf{h})\frac{h_l\sqrt{N}+Nf_l}{f_l N}\left[u_s\left(\mathbf{h}+\frac{\mathbf{e}_k-\mathbf{e}_l}{\sqrt{N}}\right)-u_s(\mathbf{h})\right] + \\
& \sum_{\mathbf{h}\in H^N, h_l=-f_l\sqrt{N}} \mu^N(\mathbf{h})\,\sigma^N(\mathbf{h})\frac{h_l\sqrt{N}+Nf_l}{f_l N}\left[u_s\left(\mathbf{h}+\frac{\mathbf{e}_k-\mathbf{e}_l}{\sqrt{N}}\right)-u_s(\mathbf{h})\right]
\end{aligned}$$

If we use $\mathbb{E}^N$ to refer to the expectation with respect to μ^N, we can write the incentive constraints as

$$U_{k,k}^N - U_{l,l}^N \geq \mathbb{E}^N \left[\sigma^N(\mathbf{h}) \frac{h_l\sqrt{N} + f_l N}{f_l N} \left(u_s\left(\mathbf{h} + \frac{\mathbf{e}_k - \mathbf{e}_l}{\sqrt{N}}\right) - u_s(\mathbf{h}) \right) \right] \tag{47}$$

As N tends to infinity, $\sqrt{N}\left(u_s\left(\mathbf{h} + \frac{\mathbf{e}_k - \mathbf{e}_l}{\sqrt{N}}\right) - u_s(\mathbf{h})\right)$ converges to $\frac{\partial u_s}{\partial h_k} - \frac{\partial u_s}{\partial h_l}$. Similar to the specific case of section 4, we have

$$\begin{aligned}
\sqrt{N}\left(U_{k,k}^N - U_{l,l}^N\right) &= \sqrt{N}\mathbb{E}^N \left[\sigma(\mathbf{h}) \left(\frac{h_k\sqrt{N} + f_k N}{f_k N} - \frac{h_l\sqrt{N} + f_l N}{f_l N} \right) u_s(\mathbf{h}) \right] \\
&= \mathbb{E}^N \left[\sigma(\mathbf{h}) \left(\frac{h_k}{f_k} - \frac{h_l}{f_l} \right) u_s(\mathbf{h}) \right]
\end{aligned}$$

Now taking a limit in (47) gives us

$$\mathbb{E}\left[\sigma(\mathbf{h}) u_s(\mathbf{h}) \left(\frac{h_k}{f_k} - \frac{h_l}{f_l} \right)\right] \geq \mathbb{E}\left[\sigma(\mathbf{h}) \left(\frac{\partial u_s(\mathbf{h})}{\partial h_k} - \frac{\partial u_s(\mathbf{h})}{\partial h_l} \right)\right]$$

Since this has to hold for all k, l, it should hold with equality and thus

$$\mathbb{E}\left[\sigma(\mathbf{h}) \left(u_s(\mathbf{h}) \frac{h_k}{f_k} - \frac{\partial u_s(\mathbf{h})}{\partial h_k} \right)\right] = \overline{U}$$

Similar to before, we also need that when $k > l$,

$$\begin{aligned}
&\mathbb{E}^N \left[\sigma^N(\mathbf{h}) \left(\frac{h_k}{f_k\sqrt{N}} + 1 \right) \left(u_s(\mathbf{h}) - u_s\left(\mathbf{h} + \frac{\mathbf{e}_l - \mathbf{e}_k}{\sqrt{N}}\right) \right) \right] \geq \\
&\qquad \mathbb{E}^N \left[\sigma^N(\mathbf{h}) \left(\frac{h_l}{f_l\sqrt{N}} + 1 \right) \left(u_s\left(\mathbf{h} + \frac{\mathbf{e}_k - \mathbf{e}_l}{\sqrt{N}}\right) - u_s(\mathbf{h}) \right) \right]
\end{aligned}$$

Using Taylor's formula, we can write

$$\begin{aligned}
u_s\left(\mathbf{h} + \frac{\mathbf{e}_l - \mathbf{e}_k}{\sqrt{N}}\right) - u_s(\mathbf{h}) &= \frac{\mathbf{e}_l - \mathbf{e}_k}{\sqrt{N}} \nabla u_s(\mathbf{h})^{\text{tr}} + \frac{1}{2} \frac{\mathbf{e}_l - \mathbf{e}_k}{\sqrt{N}} \nabla^2 u_s(\mathbf{h}) \frac{\mathbf{e}_l^{\text{tr}} - \mathbf{e}_k^{\text{tr}}}{\sqrt{N}} + O\left(\frac{1}{N\sqrt{N}}\right) \\
u_s\left(\mathbf{h} + \frac{\mathbf{e}_k - \mathbf{e}_l}{\sqrt{N}}\right) - u_s(\mathbf{h}) &= \frac{\mathbf{e}_k - \mathbf{e}_l}{\sqrt{N}} \nabla u_s(\mathbf{h})^{\text{tr}} + \frac{1}{2} \frac{\mathbf{e}_k - \mathbf{e}_l}{\sqrt{N}} \nabla^2 u_s(\mathbf{h}) \frac{\mathbf{e}_k^{\text{tr}} - \mathbf{e}_l^{\text{tr}}}{\sqrt{N}} + O\left(\frac{1}{N\sqrt{N}}\right)
\end{aligned}$$

Thus the above inequality becomes

$$\mathbb{E}^N\left[\sigma^N(\mathbf{h})\left(\frac{h_k}{f_k\sqrt{N}}+1\right)\left(\frac{\mathbf{e}_k-\mathbf{e}_l}{\sqrt{N}}\nabla u_s(\mathbf{h})^{\mathrm{tr}}-\frac{1}{2}\frac{\mathbf{e}_k-\mathbf{e}_l}{\sqrt{N}}\nabla^2 u_s(\mathbf{h})\frac{\mathbf{e}_k^{\mathrm{tr}}-\mathbf{e}_l^{\mathrm{tr}}}{\sqrt{N}}+O\left(\frac{1}{N\sqrt{N}}\right)\right)\right]\geq$$
$$\mathbb{E}^N\left[\sigma^N(\mathbf{h})\left(\frac{h_l}{f_l\sqrt{N}}+1\right)\left(\frac{\mathbf{e}_k-\mathbf{e}_l}{\sqrt{N}}\nabla u_s(\mathbf{h})^{\mathrm{tr}}+\frac{1}{2}\frac{\mathbf{e}_k-\mathbf{e}_l}{\sqrt{N}}\nabla^2 u_s(\mathbf{h})\frac{\mathbf{e}_k^{\mathrm{tr}}-\mathbf{e}_l^{\mathrm{tr}}}{\sqrt{N}}+O\left(\frac{1}{N\sqrt{N}}\right)\right)\right]$$

We thus have

$$\frac{1}{N}\mathbb{E}^N\left[\sigma^N(\mathbf{h})\left(\frac{h_k}{f_k}(\mathbf{e}_k-\mathbf{e}_l)\nabla u_s(\mathbf{h})^{\mathrm{tr}}-\frac{1}{2}(\mathbf{e}_k-\mathbf{e}_l)\nabla^2 u_s(\mathbf{h})(\mathbf{e}_k^{\mathrm{tr}}-\mathbf{e}_l^{\mathrm{tr}})\right)\right]+O\left(\frac{1}{N\sqrt{N}}\right)\geq$$
$$\frac{1}{N}\mathbb{E}^N\left[\sigma^N(\mathbf{h})\left(\frac{h_l}{f_l}(\mathbf{e}_k-\mathbf{e}_l)\nabla u_s(\mathbf{h})^{\mathrm{tr}}+\frac{1}{2}(\mathbf{e}_k-\mathbf{e}_l)\nabla^2 u_s(\mathbf{h})(\mathbf{e}_k^{\mathrm{tr}}-\mathbf{e}_l^{\mathrm{tr}})\right)\right]+O\left(\frac{1}{N\sqrt{N}}\right)$$

Hence, as N converges to infinity, the above multiplied by N converges to

$$\mathbb{E}\left[\sigma(\mathbf{h})\left(\frac{h_k}{f_k}-\frac{h_l}{f_l}\right)\left(\frac{\partial u_s(\mathbf{h})}{\partial h_k}-\frac{\partial u_s(\mathbf{h})}{\partial h_l}\right)\right]\geq$$
$$\mathbb{E}\left[\sigma(\mathbf{h})\left(\frac{\partial^2 u_s(\mathbf{h})}{\partial h_k^2}+\frac{\partial^2 u_s(\mathbf{h})}{\partial h_l^2}-2\frac{\partial^2 u_s(\mathbf{h})}{\partial h_l\partial h_k}\right)\right]$$

The only if proof follows closely that of Theorem 1. Specifically, we use the power bound from Assumption 1 and apply Lemma ??? to show the uniform convergence used in proof of Theorem 2. This concludes the proof. □

D.6 Proof of Proposition 4

Proof. If we apply the characterization result of Proposition 2 to this setup, it implies that a mechanism $\sigma(\mathbf{h})$ is ICL if and only if it satisfies

$$\mathbb{E}\left[\sigma(\mathbf{h})\left(\mathbf{t}_s\cdot\mathbf{h}\frac{h_k}{f_k}-t_{s,k}\right)\right]=U,\forall k \tag{48}$$
$$\mathbb{E}\left[\sigma(\mathbf{h})\frac{h_k}{f_k}\right]\geq\mathbb{E}\left[\sigma(\mathbf{h})\frac{h_l}{f_l}\right],\forall k>l \tag{49}$$

The fact that $\sum_k f_k t_{s,k}=0$ implies that $U=0$.

The rest of the proof closely follows that of Proposition 1. Namely, if σ is ICL, i.e., satisfies (48) and (49), then by multiplying (48) by $f_k t_{s,k}$ and summing over k, we arrive at

$$\mathbb{E}\left[\sigma(\mathbf{h})\left(\omega_s^2-1\right)\right]=0 \tag{50}$$

where $\omega_s = \mathbf{t}_s \cdot \mathbf{h}$. Similarly, by repeating this for $f_k t_{r,k}$, we arrive at

$$\mathbb{E}\left[\sigma\left(\mathbf{h}\right)\left(\omega_s\omega_r - \gamma\right)\right] = 0 \tag{51}$$

Finally, note that properties of the normal distribution implies that

$$\mathbb{E}\left[h_k|\omega_s,\omega_r\right] = \begin{bmatrix} \omega_s & \omega_r \end{bmatrix} \begin{bmatrix} \mathbb{E}\left[\omega_s^2\right] & \mathbb{E}\left[\omega_s\omega_r\right] \\ \mathbb{E}\left[\omega_s\omega_r\right] & \mathbb{E}\left[\omega_r^2\right] \end{bmatrix}^{-1} \begin{bmatrix} \mathbb{E}\left[\omega_s h_k\right] \\ \mathbb{E}\left[\omega_r h_k\right] \end{bmatrix}$$

where

$$\begin{aligned}
\mathbb{E}\left[\omega_s^2\right] &= \sum_{k,l} \mathbb{E}\left[t_{s,k}t_{s,l}h_k h_l\right] = \sum_{k,l} t_{s,k}t_{s,l}f_k\left(\mathbf{1}\left[k=l\right] - f_l\right) \\
&= \sum_k f_k t_{s,k}^2 = 1 \\
\mathbb{E}\left[\omega_r^2\right] &= 1, \mathbb{E}\left[\omega_r\omega_s\right] = \sum_k f_k t_{r,k}t_{s,k} = \gamma \\
\mathbb{E}\left[\omega_s h_k\right] &= f_k t_{s,k}, \mathbb{E}\left[\omega_r h_k\right] = f_k t_{r,k}
\end{aligned}$$

and hence

$$\begin{aligned}
\mathbb{E}\left[h_k|\omega_s,\omega_r\right] &= f_k \begin{bmatrix} \omega_s & \omega_r \end{bmatrix} \begin{bmatrix} 1 & \gamma \\ \gamma & 1 \end{bmatrix}^{-1} \begin{bmatrix} t_{s,k} \\ t_{r,k} \end{bmatrix} \\
&= \frac{f_k}{1-\gamma^2} \begin{bmatrix} \omega_s & \omega_r \end{bmatrix} \begin{bmatrix} 1 & -\gamma \\ -\gamma & 1 \end{bmatrix} \begin{bmatrix} t_{s,k} \\ t_{r,k} \end{bmatrix} \\
&= \frac{f_k}{1-\gamma^2}\left[\left(t_{s,k} - \gamma t_{r,k}\right)\omega_s + \left(t_{r,k} - \gamma t_{s,k}\right)\omega_r\right]
\end{aligned}$$

Replacing the above in the (49) leads to

$$\mathbb{E}\left[\sigma\left(\mathbf{h}\right)\left(\left(t_{s,k} - \gamma t_{r,k}\right)\omega_s + \left(t_{r,k} - \gamma t_{s,k}\right)\omega_r\right)\right] : \text{increasing in } k \tag{52}$$

Hence, similar to the proof of Proposition 1, we can focus on a relaxed problem where instead of ICL we impose (50), (51), and (52). The solution of the relaxed problem should then satisfy the conditions provided in proposition 4. Finally, we must show that the solution of the relaxed problem is indeed ICL. This follows steps similar to those in proof of Proposition 1. □

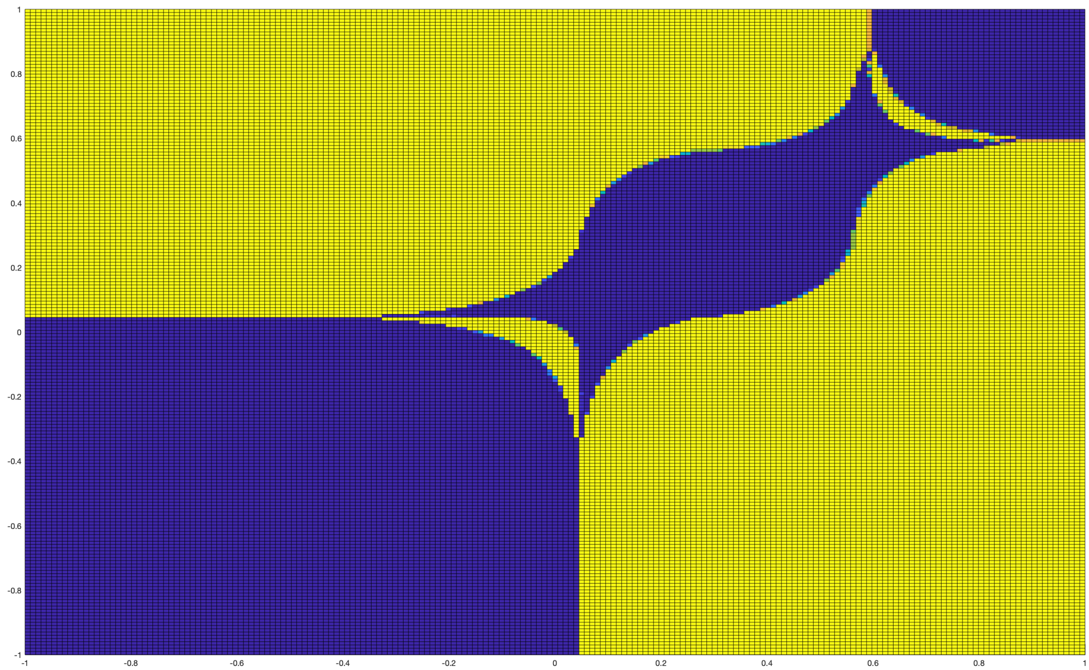

Figure 5: Numerical simulation of the optimal mechanism

E Numerical Solution of Small Economy with Large Bias

The problem of finding the optimal mechanism, (P), is a linear program. Here we plot the numerical solution of this linear program when $N = 2$, $b = 0.6\sqrt{2}$, $s_i \sim U[-1, 1]$, and $r = 0$. For the numerical solution, we discretize $[-1, 1]$ into 200 subintervals and solve for optimal σ for each square generated by all such subintervals, i.e., we solve for a vector of 40000 values. We use the `linprog` function in MATLAB to solve this linear program. Figure 5 depicts the solution. The yellow area represents the values for which $\sigma = 1$ and the blue area is associated with $\sigma = 0$.