Inference from Selectively Disclosed Data

Ying Gao*

January 22, 2024

Click here for latest version.

Abstract

We consider the disclosure problem of a sender with a large dataset of hard evidence. The sender has an incentive to drop observations before submitting the data to the receiver to persuade them to take a favorable action. We predict which observations the sender discloses using a model with a continuum of data, and show that this model approximates the outcomes with large, multi-variable datasets. In the receiver's preferred equilibrium, the sender plays an imitation strategy, under which they submit evidence that imitates the natural distribution under a more desirable target state. As a result, it is enough for an experiment to record data on outcomes that maximally distinguish higher states. A characterization of these strategies shows that senders with little data or a favorable state fully disclose their data, but still suffer from the receiver's skepticism, and therefore are worse-off than they are under full information. On the other hand, senders with large datasets can benefit from voluntary disclosure by dropping observations under low states.

*Department of Economics, MIT. Email: gaoy@mit.edu. I thank Drew Fudenberg and Stephen Morris for their generous advice and guidance, and Isaiah Andrews, Ian Ball, In-Koo Cho, Laura Doval, Wioletta Dzuida, Bob Gibbons, Ellen Muir, Marco Ottaviani, Harry Pei, Alexander Wolitzky, and audiences at the Stony Brook International Conference in Game Theory, Econometric Society Summer School in Game Theory, European Summer Meetings of the Econometric Society, MIT Organizational Economics Seminar, and MIT Theory Lunch for helpful comments.

1 Introduction

Many decisions – including technology adoption, regulatory approval, and research grantmaking – are based on self-disclosed data. The datasets used can often be very large, on the order of tens of thousands of trials for drug approval, and often hundreds of thousands of datapoints about locations and sales in merger cases. In and of themselves, big datasets may paint an accurate picture of reality, but the sender can disclose them strategically: it is easier to verify that the data submitted are real than that they are complete, and even in the presence of mandatory disclosure rules, deciding which observations are admissible to include in the dataset is largely at the sender's discretion.

We want to understand the role data play in strategic communication between the sender and the decision-maker when receivers have uncertainty about the underlying dataset from which the sender extracted the submitted data. We consider the case of a sender with state-independent motives to persuade the receiver towards a particular action, and a receiver who observes a dataset the sender discloses, but interprets it with partial skepticism that the data are incomplete. Equilibrium play between the sender and receiver involves the sender submitting data as "proof" that the receiver should take a favorable action, and the receiver evaluating how persuasive the proof is depending on how likely it is sent by a sender with less persuasive data who has trimmed some discouraging observations. This can be modeled under the framework of an evidence game in which senders that have access to datasets with weakly more observations of each outcome can always mimic senders with fewer observations. A special case, in which senders either have or do not have access to a single data point, with probability known to the receiver, is already well-understood (Dye 1985), and demonstrates that senders can manipulate the receiver by disclosing nothing when the evidence is sufficiently poor.

Our primary innovation is to characterize disclosure in the opposite extreme, when datasets contain many observations. We propose a continuous-data model of the asymptotic distribution over potential datasets of the sender that depends on two things: the true state of the world that generates the data, and a random variable that describes the amount of data the sender collects. The continuum assumption captures the fact that empirical distributions are approximately deterministic in the limit with large numbers, and allows us to eliminate uncertainty over the randomness of draws, which makes the model more tractable than directly modeling large, finite N. Instead, we show that the outcome we characterize in the continuous model describes the limit outcome of communication in finite-data games as $N \to \infty$.

In addition to an extensive list of observations, a second characteristic feature of "big" data is a large outcome space. This motivates the novel use of a framework that encompasses general statistical settings, including those in which outcome and state spaces are large and the relationship between them complex. In particular, we place essentially no restrictions on the state-contingent data distribution. In general, unlike in a "good news-bad news" model of data, the ranking of states and the shape of their data-generating distributions endogenously affects the interpretation of different outcomes, and context determines whether data take on a positive or negative connotation.

Indeed, our first main result, Prop. 1, is a sufficiency result that says that in the receiver-optimal partial pooling equilibrium outcome, the state-contingent experimental outcome distribution affects the information transmitted only through a handful of key features: what matters are the observations of outcomes with the greatest likelihood ratio under a better vs. a worse state. Strikingly, since the distribution of data that distinguishes one state from another depends solely on the relative probabilities of likelihood ratio-maximizing outcomes, a receiver who wants to distinguish a relatively small number of states with many observations of high-dimensional data can do just as well restricting the dataset to only retain information about these outcomes. When state-contingent distributions of experimental outcomes satisfy the monotone likelihood ratio property (MLRP), we return to the case of only one "good news" outcome, and distinguishing it from other outcomes is sufficient to support receiver-optimal communication.

Our second result, Theorem 2, characterizes an "imitation" equilibrium implementation of the receiver-optimal equilibrium outcome, in which senders always show the receiver a dataset that can correspond to a naturally-generated dataset, so that on path, the receiver always places positive probability on the event that the sender is sending all their data. However, the receiver also infers from some datasets that the sender has with positive probability observed data corresponding to a different state than the revealed data suggest, but has dropped observations in order to imitate a more favorable distribution. When MLRP fails, it is important for the imitating sender to send a large-enough mass of realizations of a certain outcome, but not too much. The resulting outcome benefits senders under low states with more data at the expense of senders with less data in high states, since the former pool with the latter. The extent of pooling depends on the receiver's uncertainty about the sender's data collection capabilities: the greater the variance in the receiver's belief about how much data the sender starts out with, the more senders can profitably imitate other senders, with outcomes converging to the full-information one as uncertainty vanishes.

The other contribution is an algorithm, in section 3.1, to construct the limit game equilibrium outcomes that follows a top-down logic: senders with more data receive weakly greater payoffs, and we can construct the payoff frontiers of the continuous payoff function by specifying the burden of proof, or how much data of a given state's distribution a sender needs, to induce a particular belief in the receiver. The algorithm is applicable to any number of states and to any datasets with finite support, and we illustrate it with representative 2 and 3 state examples.

1.1 Related literature

Strategic disclosure has been studied since the work of Grossman (1981) and Milgrom (1981), which showed that full disclosure is the unique outcome when receivers know that the sender wishes to prove the value of a good is high using verifiable information that they could choose to disclose. The assumption that receivers know the sender is informed is crucial to this benchmark, as Dye (1985) and Jung and Kwon (1988) show. They consider a case in which the sender has access to a single, real-valued piece of evidence with interior probability $p \in (0, 1)$, and shows that only senders for whom the evidence exceeds a threshold will choose to disclose it, with the rest withholding it in order to pool with those senders who lack evidence altogether. Shin (1994, 2003) shows that in the case where senders have an uncertain endowment of good news and bad news, the fact that senders withhold bad enough evidence implies a "sanitation equilibrium", in which all bad news is disposed of.

We extend these results by considering evidence structures with large, multidimensonal datasets. In our data-based setting, evidence is neither exogenously good nor bad, but the receiver draws inferences statistically, based on knowledge of the relationship between relevant state-parameters and the distributions of data they generate. The setting we consider encompasses the settings above and captures a special case of more abstract evidence games of the type considered by Green and Laffont (1986), Hart et al. (2017), and Glazer and Rubinstein (2006). The main focus in those settings has been on receiver-optimal mechanisms to induce beneficial disclosures from the sender. Hart et al. (2017) in particular is foundational to our equilibrium selection criterion. Their observation that the optimal deterministic mechanism, the receiver-optimal equilibrium, and the unique truth-leaning equilibrium all yield the same outcome generalizes straightforwardly to our setting.[1]

Rappoport (2022) and Jiang (2022) use an iterative algorithm to solve for truth-leaning equilibrium outcomes in finite evidence games, but it is computationally demanding to use it in games with large type spaces, and therefore infeasible to directly compute the large-N limit of outcomes in finite-data games. Our approach is instead to use a continuous-data approximation to solve for asymptotic outcomes without explicitly computing outcomes of finite-data games, and to show that it is exactly the big-data limit outcome.

Only one other paper that we know of, by Dzuida (2011), uses a continuous measure of evidence to solve for communication with verifiable evidence. Differently from our model, hers considers evidence with a continuum of states but a simple "good news-bad news" outcome structure, and assumes there is a positive probability of a behavioral, honest type of the sender. The existence of

[1]The optimal mechanism equivalence result has been noted by others, including Glazer and Rubinstein (2006), Sher (2011), and Ben-Porath et al. (2019), who show that the fact that commitment is not necessary for the optimum is robust to other settings, in particular with binary actions and multiple senders with type-dependent preferences.

the honest type, along with the assumption of continuity in outcomes, selects the most plausible equilibrium in a similar fashion to truth-leaning. The honest type also drives one of the paper's key observations: providing interior amounts of negative evidence can be optimal in otherwise sanitation-like equilibria, because honest types will send some negative evidence. The sender sends intermediate amounts of evidence of some outcomes in our model as well, even though we lack honest types. In particular, with more than two states or two outcomes, there is no single "good" outcome, and we show that sending interior amounts of some outcomes may rule in just the right set of rational types, while sending either none or as much as possible of an outcome might both be strictly worse.

We also relate to a broader literature about the optimal collection and disclosure of evidence, that considers costly (Migrow and Severinov (2022)) and dynamic evidence acquisition (Felgenhauer and Schulte 2014, Henry and Ottaviani 2019), sender-optimal disclosure mechanisms (Haghtalab et al. 2022), and discretionary disclosure after test or information design (Shiskin 2022, Dasgupta et al. 2022). Several papers use a restricted notion of evidence but are also explicitly concerned with the effect of allowing sample selection: Fishman and Haggerty (1990) and Di Tillio et al. (2021) study the case in which only a subset of observations are disclosed, and give conditions under which it is better that an informant have discretion over which data are selected.

Finally, there is a small literature of empirical findings about common patterns of voluntary disclosure. Some work in econometrics by Simonsohn et al. (2014), Andrews and Kasy (2019) and others studies the bias that arises from a range of exogenous patterns of selective reporting, and describes inference procedures that correct for it. In addition, a small body of experimental work studies how subjects disclose evidence when incentivized to persuade receivers in the lab. Jin et al. (2021) finds evidence that receivers' inferences are often biased by accounting insufficiently for the sender's nondisclosure, and Li and Schipper (2020) shows that senders are also biased towards naive, truthful behavior. Both suggest that these behaviors are consistent with an initial lack of higher-order sophistication that is remedied, to some extent, by experience. On the other hand, Osun and Ozbay (2021) suggest that in a binary-type evidence game, senders' disclosure policies and receivers' commitment policies differ from those predicted by Hart et al. (2017) in a direction consistent with a positive cost of lying, which is absent from monetary payoffs but may be inherent to the subjects' preferences.

2 Model

States and payoffs. There is a sender (S), who wishes to communicate to a receiver (R) about an unknown state of the world, $\theta \in \Theta = \{\theta_1, \dots, \theta_J\}$. The sender and receiver share a common prior $\beta_0(\cdot)$ over Θ. We assume that the receiver takes an action $a_r \in \mathbb{R}$ and that $\theta_1, \dots, \theta_J$ are real numbers ordered with $\theta_j \leq \theta_{j+1}$, representing the optimal action for the receiver under each state, if it was known with certainty. The sender's payoff is simply (a monotone function of) a_r;[2] in short, regardless of their type, they want to induce the receiver to take the highest action possible.

Finally, we assume the receiver has an expected utility that is differentiable and single-peaked at the action that matches their expectation of the value of θ, that is, that for any belief $\beta \in \Delta\Theta$, the receiver's expected payoff $\mathbb{E}_\beta[u_r(a)]$ is single-peaked at $a_r(\beta) = \mathbb{E}_\beta[\theta]$.[3] We work with the sender's indirect utility as a function of the receiver's beliefs, which induces them to maximize the receiver's posterior expectation of θ:

$$u_s(\beta) = \mathbb{E}_\beta[\theta]. \tag{1}$$

For example, when the receiver is a policymaker, states can represent the true optimal policy. While the policymaker might be uncertain, they wish to enact a policy that matches the optimal policy in expectation, while the sender wishes for them to take as high an action as possible.

Evidence. The private information of the sender comes in the form of hard evidence about the state of the world. In particular, the sender has access to a dataset of observations drawn from a finite set of outcomes, $\mathcal{D} = \{1, \dots, D\}$. The underlying data-generating distribution is state-contingent: under state θ_j, the observations are i.i.d. draws from distribution f_j.

We model the amount of data the sender has access to as a mass, $\mu \in [0, 1]$, that represents the fraction of total potential data that the sender can access, and has a continuous distribution, g, that is state-independent[4], supported on $[0, 1]$ with $g(1) = 0$, and infinitely left-differentiable[5]. The continuum assumption models big datasets in which the large number of draws essentially removes all uncertainty about the impact of randomly realized outcomes on the sender's dataset: conditional on state θ_j, the empirical distribution of data the sender observes is certain to be f_j, and μ does not affect the distribution of their evidence, only the amount of it. In other words, with probability 1, a sender with a mass μ of data under state θ_j observes the dataset $t = \mu f_j$. Any nonzero measure of data fully informs the sender of the state, and the set of possible complete datasets and types of the sender is $\mathcal{T} = [0, 1] \times \Theta$.

The receiver, on the other hand, is uninformed about how much data the sender has. Their

[2]Because the receiver will always play a pure strategy, the sender's problem is unchanged if their payoffs are rescaled through a monotone mapping.

[3]The assumption of single-peakedness is necessary to identify the receiver-optimal equilibrium and the receiver-optimal mechanism.

[4]For simplicity of exposition, we focus on the case in which their belief about μ conditional on θ is given by a probability density g that is independent of θ, although most results hold identically for cases in which the distribution of μ is state-specific.

[5]The assumption that g has a vanishing right tail ensures that it is continuous on $\mathbb{R}^+$ while being supported on $[0, 1]$, and simplifies the equilibrium construction: specifically, it ensures that the equilibrium payoffs are continuous in μ.

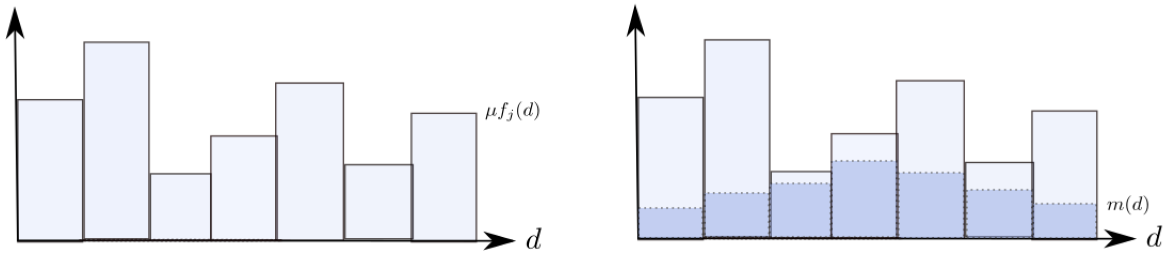


Figure 1: A feasible type and a feasible message.

prior belief about the sender's type is given by the density

$$q(\mu f_j) = \beta_0(\theta_j) g(\mu). \tag{2}$$

Messaging and inference. Senders can choose a subset of observations from their dataset to submit to the receiver. We assume total flexibility in the choice of subset:

Assumption 1. *The sender can send any message $m \in \mathcal{M} = [0, 1] \times \Delta\mathcal{D}$ that is a* subset *of their dataset ($m \tilde{\subseteq} \mu f_j$), where*

$$m \tilde{\subseteq} \mu f_j \;\; \Leftrightarrow \;\; m(d) \leq \mu f_j(d) \;\; \forall d \in \mathcal{D}.$$

That is, a sender can drop an arbitrary mass of observations from their data, and then show the remaining ones to the receiver. By dropping observations, they can arbitrarily alter the relative frequencies of each outcome in the submitted dataset in order to imitate any distribution. However, this is costly in that it reduces the size of the submitted dataset, which is observable.

We have that $\mathcal{M} \supset \mathcal{T}$: the message space contains the set of all possible complete datasets, but also a D-dimensional set of other datasets that could be disclosed to the receiver after excluding part of their dataset. For any set of messages M, define the upper set $U(M)$ to be the set of types that can send a message in M, and for any set of types T, define the lower set $L(T)$ as the set of messages that some $t \in T$ can send.

Call the disclosure game with these parameters $\mathcal{G}(\Theta, \mathcal{D}, \beta_0, \{f_j\}_{j=1}^J, G)$. Upon observing the sender's message, the receiver updates their belief about the sender's type to $q(t|m)$, and then forms a new belief about the state,

$$\beta(\theta_j|m) = \frac{\sum_{j=1}^J \int_{\mu=0}^1 q(\mu f_j|m)\theta_j}{\sum_{j=1}^J \int_{\mu=0}^1 q(\mu f_j|m)}. \tag{3}$$

2.1 Equilibrium

The sender plays a messaging strategy $\sigma^* : \mathcal{T} \to \Delta\mathcal{M}$, knowing which the receiver infers the content of message they receive. As usual, the equilibrium we consider will be a Perfect Bayesian Equilibrium (Fudenberg and Tirole 1970), that is, $\beta^*(\cdot|m)$ must be consistent with the sender's strategy σ^*, and the sender must optimize, so $\sigma^*(m|t) > 0$ only if $m \in \arg\max_{m' \in L(t)} \mathbb{E}_{\beta(\cdot|m')}[\theta]$.

Call the map from types to payoffs, $u_{\sigma^*}(t)$, the *outcome* of the equilibrium.[6] In the perfectly separating outcome, the sender obtains a payoff of θ. As in Milgrom (1981), Grossman (1981), and Dye (1985), when g is a degenerate distribution such that μ is known to the receiver, then all attempts to mislead the receiver unravel, and the fully separating outcome obtains in every PBE. When g and all f_j have full support, there is partial pooling in every PBE. However, PBE are often not unique, and in this case, there may be multiple β^*, differing on off-path messages, that are consistent with σ^*, and the game generically has multiple, non-payoff-equivalent PBE outcomes. Any message that can be played by some type of sender under a state $\theta_j \geq \mathbb{E}_{\beta_0}[\theta]$ is played on-path in some PBE.

Intuition suggests that the game is fundamentally one of imitation: senders tailor their data to increase the receiver's belief that the state is a higher one, and they can only do so by imitating the datasets submitted by higher-state types, who themselves may be imitating others or trying to distinguish themselves as well as possible from lower-state types. One way to imitate a higher-state type of sender is to try to prove you have all the data that they would, and no more – that is, to imitate their complete dataset. We define an *imitation equilibrium* to capture the idea that sender masquerades as other type by imitating their full datasets.

Definition 2.1. *(σ^*, β^*) is an imitation equilibrium if it is an equilibrium, and under σ^*,*

a. Every on-path message is in $\mathcal{T}$,

b. Type μf_j plays $m \neq \mu f_j$ if and only if $\theta_j < \max_{m' \in L(t)} \mathbb{E}_{\beta^*(\cdot|m')}[\theta]$, and otherwise reports their full dataset.

In other words, with an imitation messaging strategy every type of the sender either fully reveals their data or imitates another type's full dataset, and they only consider the latter if it could give them a better payoff than letting the receiver be fully informed of the state.

Why do we focus on these equilibria? Imitation equilibria are *truth-leaning*, as first defined by Hart et al. (2017) in the context of general evidence games with finite types. The idea applies identically in this setting. Formally, given a base game $\mathcal{G}$, for $\epsilon = (\epsilon_t, \epsilon_{t|t})_{t \in \mathcal{T}}$, let a game $\mathcal{G}_\epsilon$ be the game with an identical type set and type distribution, but with two differences. First, type t's

[6] This is a departure from the usual definition of an outcome of an extensive-form game, but consistent with the definition in Hart et al. (2017) and Rappoport (2022). It describes the action the receiver plays after communicating with each type, and so describes the consequences of communication in the game.

payoffs to playing t are perturbed by ϵ_t, so that t's payoff to playing t is $\mathbb{E}_{\beta(\cdot|t)}[\theta] + \epsilon_t$. Secondly, type t plays t with at least probability $\epsilon_{t|t}$ – i.e. with probability $\epsilon_{t|t}$ a sender with dataset t is a commitment type that plays their full dataset regardless of whether doing so is optimal, while with probability $1 - \epsilon_{t|t}$ type t is strategic. A truth-leaning equilibrium is an equilibrium of the base game that can be obtained as a limit of equilibria of ϵ-perturbed games as $\epsilon \to 0$.

While truth-leaning equilibrium strategies capture a sender's slight bias towards truth-telling, the truth-leaning equilibrium outcome has desirable properties in its own right. When the receiver's expected payoffs are single-peaked in their action, the truth-leaning equilibrium outcome is also receiver-optimal, and is the outcome of the optimal mechanism when the receiver can commit to a single action as a response to each message. This is well-known in the finite case studied by Hart et al. (2017), and continues to be true in the continuous model that we study. It is also the only equilibrium outcome robust to a slightly stronger version of a credible announcement (Matthews et al. 1991). We say that under equilibrium σ^* a collection T of types of the sender can benefit from an *inclusive* credible announcement if there is a set of messages such that T is comprised of every type that 1) finds some message in the set feasible and 2) weakly benefits from the receiver updating that their type is in T from the prior, relative to the receiver's equilibrium inference; and there is at least one type in T that strictly benefits.[7] Robustness to such announcements means that the equilibrium survives even if senders are able to override the receiver's beliefs by proposing sensible reinterpretations of messages, and can coordinate to do so; it rules out, for example, play that is "stuck" in a bad equilibrium due to immalleable off-path beliefs. For a deeper discussion of these refinements, see Hart et al. (2017) and Appendix B.

Claim 1. *Imitation equilibrium messaging strategies are the truth-leaning equilibrium messaging strategies of $\mathcal{G}$. Imitation equilibrium outcomes are:*

- *Receiver-optimal among equilibria and deterministic mechanisms;*
- *The unique inclusive announcement-proof equilibrium outcomes.*

2.2 Examples

2.2.1 A 2-state prediction problem

A sender wishes to provide evidence to prove the quality of a prediction algorithm that aims to classify whether a future event (e.g., rain) is likely or unlikely. The quality of the algorithm is either high or low ($\Theta = \{\theta_L, \theta_H\}$), with $\theta_L = 0$ and $\theta_H = 1$.[8] Suppose that there are 4 possible outcomes,

[7] The departure from the usual credible announcement is that the collection of types making the announcment must also contain all types who are indifferent between participating in the announcement and their equilibrium payoff.

[8] We can think of the high-quality algorithm as being able to more accurately make the right call when it will not rain, while the low-quality algorithm often predicts rain even when it will not rain.

$\mathcal{D} = \{1, 2, 3, 4\}$ with the following distribution of outcomes per state:

j	$f_j(1)$	$f_j(2)$	$f_j(3)$	$f_j(4)$
H	0.6	0.2	0.1	0.1
L	0.4	0.2	0.3	0.1

Table 1: The generating distribution of outcomes under states θ_H and θ_L.
Outcome 1 is *(predict no rain, no rain)*, outcome 2 is *(predict rain, rain)*, outcome 3 is *(predict rain, no rain)*, and outcome 4 is *(predict no rain, rain)*.

Imitation implies that every on-path message either contains data distributed like f_H or like f_L, which the receiver can interpret as a claim that "the state is θ_H" or "the state is θ_L", respectively. But since the sender strictly prefers the receiver to believe the state is H with higher probability, there is no reason to imitate f_L. Indeed, part 2.1(b) of the definition of an imitation equilibrium ensures that the only on-path messages take the form μf_H, since no posterior belief of the receiver is worse for the sender than full certainty that $\theta = \theta_L$.

Additionally, the sender chooses an amount of data to send, which the receiver can interpret as an amount of support to back up their claim. The sender's true dataset determines whether they are able to submit more or less data that fits the distribution, and it is optimal for the receiver distinguish them along this margin to encourage partial separation. When evidence is generated as in Table 1, the sender can send $m = \mu f_H$ if and only if the true data are $\mu' f_H$ with $\mu' \geq \mu$, or $\mu' f_L$ with $\mu' \geq \frac{3}{2}\mu$. The *distinguishability factor* of $\frac{3}{2}$ reflects the relative advantage to a sender under θ_H of imitating f_H, and comes from the fact that in order to be able to submit enough observations of outcome 1 to imitate μf_H, a sender under θ_L must start with $\frac{3}{2}$ as much data.

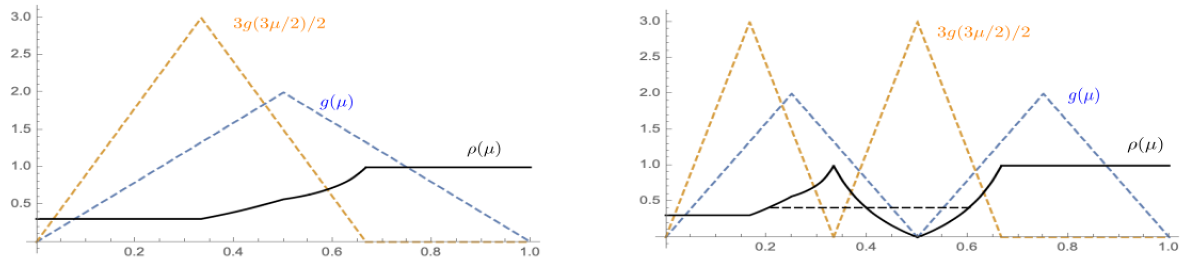
As a naive first guess, suppose that the sender's strategy is to always send the maximum possible amount of data that is distributed like f_H.

$$m_{max}(\mu f_H) = \mu f_H, \quad m_{max}(\mu f_L) = \frac{2}{3}\mu f_L. \tag{4}$$

Consider the uniform prior $\beta_0(\theta_H) = \frac{1}{2}$ and a data-mass distribution that is "triangular",

$$g(\mu) = 2 - 4|x - 1/2|.$$

The receiver's inference upon receiving a message $m_{max} = \mu f_H$, plotted by the solid line in Figure



(a) Posterior beliefs when g is triangular (b) Posterior beliefs when g is double-triangular

Figure 2: Inferences from message $m = \mu f_H$ in the binary-state example.

2(a), is

$$\rho_{\max}(\mu) = \begin{cases} 1, & \mu \geq 2/3 \\ \frac{4-4\mu}{10-13\mu}, & \mu \in [1/2, 2/3) \\ \frac{4\mu}{6-5\mu}, & \mu \in [1/3, 1/2) \\ \frac{4}{13}, & \mu < 1/3. \end{cases}$$

To visualize how the receiver constructs the posterior inference, observe that the density of senders who send a message μf_H for a μ for whom the true state is θ_H and θ_M are $g(\mu)$ and $\frac{3g(3\mu/2)}{2}$, which are plotted as two dotted lines. Their ratio is the likelihood ratio of the high vs. the low state given message μf_H.

Observation 1. *$\rho_{\max}$ depends only on β_0, g, and the distinguishability factor.*

In other words, the distinguishability of f_H from f_L is a sufficient statistic for both distributions that captures their implications for inferences under the naive strategy. In fact, we can verify that the naive messaging strategy in eq. 4 supports an equilibrium, under the assumption that any off-path messages feasible for some low-state type of the sender are evidence of the low state. More generally, the naive strategy is the unique imitation equilibrium strategy whenever it induces monotone inferences from the receiver.

In some cases, $\rho_{max}(\mu)$ is nonmonotone, such as when μ takes the "double triangular" distribution

$$g(\mu) = \begin{cases} 2 - 8|x - 1/4|, & x \in [0, 1/2] \\ 2 - 8|x - 3/4|, & x \in (1/2, 1]. \end{cases}$$

If all types of the sender send the maximal mass of data imitating f_H, then the message $1/2 f_H$ makes the receiver more pessimistic than the message $1/3 f_H$, and incentive compatibility fails because a sender who was to send the former would choose to send the latter instead. This is easily fixed, however, if, within a pooling interval, all types of the sender still imitate f_H, but send less than the maximal mass. The dashed line in Figure 2(b) shows that the receiver's inferences given all messages in an interval can be equalized this way, so that the unique equilibrium inference is instead an ironed version of $\rho_{\max}(\mu)$.[9]

Remark. *In this example, data on outcome* 1 *is the limiting factor that restricts the state-*θ_L *sender from providing more data to support "the state is* θ_H*". Data on the remaining outcomes does not matter. There are several things one could do with the remaining outcomes that would not change the results of disclosure:*

1. *Merge them, i.e., design an experiment that does not distinguish between outcomes* 2, 3, *and* 4, *and let the sender self-disclose raw data generated from the simplified experiment.*
2. *Delete them, i.e. allow the sender to report only observations of outcome* 1, *while leaving no option to input instances of outcomes* 2, 3, *and* 4.

Finding the determinants of distinguishability therefore points out ways to lighten the burden of data transmitted while retaining the most essential information.

2.2.2 A 3-state extension

Now suppose there is a 3rd possible quality of the prediction model, represented by state θ_M. The medium-quality model yields a different distribution of predictions; to summarize, the distributions of the same 4 outcomes under all states are given by Table 2.[10]

	$f_j(1)$	$f_j(2)$	$f_j(3)$	$f_j(4)$
H	0.6	0.2	0.1	0.1
M	0.4	0.25	0.3	0.05
L	0.4	0.2	0.3	0.1

Table 2: Data-generating distributions under states θ_H, θ_M and θ_L.

Consider first the problem of a sender who knows that $\theta = \theta_L$. There are now 2 distributions that they can imitate: f_M and f_H. On the other hand, a sender for whom $\theta = \theta_M$ may wish to

[9] The ironing process can be described as follows. If all types that would send $m = \mu f_H$ for some $\mu \in [\underline{\mu}, \bar{\mu}]$ were pooled, the receiver's inference given the pool would be

$$p(\underline{\mu}, \bar{\mu}) = \frac{\int_{\underline{\mu}}^{\bar{\mu}} g(\mu) d\mu}{\int_{\underline{\mu}}^{\bar{\mu}} \left(g(\mu) + \frac{3g(3\mu/2)}{2} \right) d\mu}.$$

Given some μ^* at which $\rho_{\max}(\mu)$ is decreasing, we can find $\underline{\mu} < \mu^* < \bar{\mu}$, such that either $\rho_{\max}(\mu)$ is increasing at both $\underline{\mu}$ and $\bar{\mu}$, and $\rho(\underline{\mu}) = \rho(\bar{\mu}) = p(\underline{\mu}, \bar{\mu})$; or $\underline{\mu} = 0$ and $\rho_{\max}(\bar{\mu})$ is increasing at $\bar{\mu}$ with $\rho(\bar{\mu}) = p(\underline{\mu}, \bar{\mu})$; or, $\bar{\mu} = 0$ and $\rho_{\max}(\underline{\mu})$ is increasing at $\underline{\mu}$ with $\rho(\underline{\mu}) = p(\underline{\mu}, \bar{\mu})$. There is a pair $(\underline{\mu}, \bar{\mu})$ satisfying these criteria that are closest to μ^*, and they are the endpoints of the ironing interval.

[10] In the example weather-prediction application, the state-M algorithm is better than the state-H algorithm at calling the presence of rain, but worse at identifying when it will not rain. It is correct less often than the state-H algorithm, but more often than the state-L algorithm.

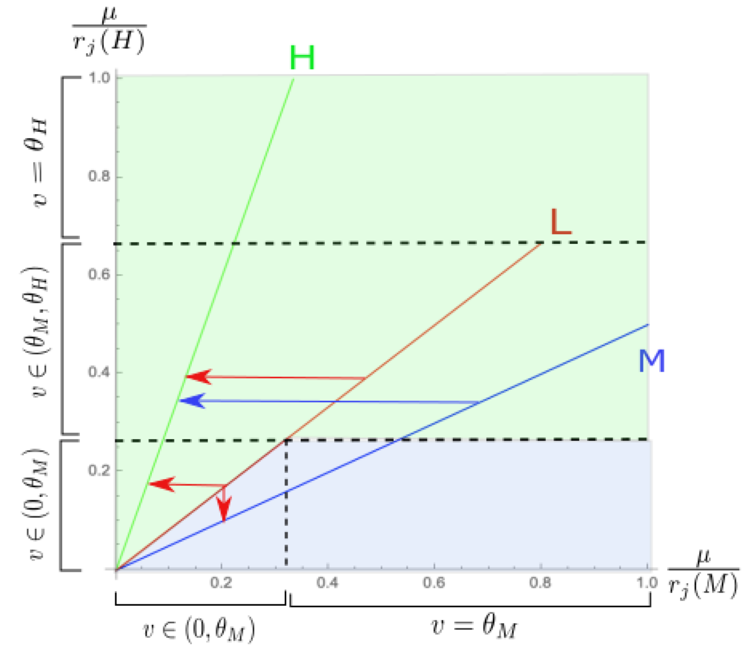


Figure 3: Under states θ_L, θ_H, and θ_M, the sender either imitates f_H (in green region) or f_M (in blue region), or mixes (boundary). The arrows indicate datasets that selected types imitate, and show that some types mix (bottom left).

imitate is f_H, but never f_L. It takes at least $\frac{5}{4}\mu f_L$ and $\frac{3}{2}\mu f_L$ to imitate μf_M and μf_H, respectively, and $2\mu f_M$ to imitate μf_H. We can now keep track of three distinguishability factors, $r_L(M) = \frac{5}{4}$, $r_L(H) = \frac{3}{2}$, and $r_M(H) = 2$.

Relative to the binary-state case, solving for the equilibrium when $|\Theta| \geq 3$ involves an extra step: understanding which state a sender will choose to target in imitation. Nevertheless, construction can proceed from the top down. First observe that types μf_H with $\mu > \frac{1}{r_L(H)}$ can separate and obtain a payoff of θ_H. We then ask which types of senders obtain a payoff $v \in (\theta_M, \theta_H)$. For this restricted set of payoff frontiers, it suffices to consider imitating f_H only, since no message imitating f_M can yield a payoff greater than θ_M. Similarly to the binary-state case, in this regime the receiver can conjecture that the sender "imitates as much of f_H as possible", and restore monotonicity if needed by ironing. For payoff frontiers corresponding to $v < \theta_M$, one of two things is possible. If the state is θ_M and the sender has enough data to separate from all other types that cannot obtain $v > \theta_M$ by imitating f_H, then they play their full dataset and separate. Otherwise, unless the state is θ_L, the sender plays their full dataset, but their full dataset is imitated by some type for whom the state is low, and who plays a strategy that mixes between imitating f_H and f_L. Figure 3 summarizes how the three distinguishability factors $r_L(M), r_L(H)$, and $r_M(H)$ determine the equilibrium: it projects all types onto a space that summarizes how imitable f_H and f_M are, as the vertical and horizontal dimensions, and shows their imitation strategies and payoffs in equilibrium.

A novelty of the payoff structure with 3, and indeed more, states is that the sender will separate and fully inform the receiver of the state only if they possess an intermediate amount of data – ignoring the best and worst state, under θ_M there is a temptation to drop evidence with too much data, and an inability to distinguish oneself from imitators when too little data is acquired. The generality of multiple states also has other implications.

Observation 2. *The multi-state case has features that do not occur when* $|\Theta| = 2$ *or* $|\mathcal{D}| = 2$*:*

- *Sending observations of multiple outcomes may be necessary, and an interior mass of observations of some outcomes may be strictly optimal.*
- *Keeping* μ *constant, the sender can receive greater payoffs under a lower state.*

As an example of the first point, consider type μf_L imitating $\frac{2}{3}\mu f_H$ by sending a mass $\frac{1}{15}\mu$ of observations of outcome 4. Sending a greater mass would rule out the type $\frac{2}{3}\mu f_H$ that it wants to imitate, but sending less would rule in types like $(\frac{2}{15} - \epsilon)\mu f_M$, which would worsen the receiver's inference from the message. To demonstrate the second point, observe that the type μf_L obtains a greater payoff than the type μf_M when $\mu = 1$, because the former can imitate $\frac{2}{3}\mu f_H$, while the latter can only imitate $\frac{1}{2}\mu f_H$. In this case, the true state determines a sender's welfare not so much through its value as through the relative advantage it confers in matching *better* states on observables.

3 Construction and characterization

This section characterizes the imitation equilibrium, constructs it, and shows that it is essentially unique. The imitation equilibrium is distinguished among equilibria by the fact that in it, worse types imitate better types (condition 2.1b). This is directly reflected in the structure of the receiver's beliefs once they receive an on-path message m: the best case for any message is that the receiver takes it literally to be the sender's full dataset, while any skepticism that this is true negatively affects their inferences. Any off-path dataset $m \in \mathcal{T}$ might as well be taken literally,

$$q^*(\cdot|m) = \mathbb{1}_m \text{ for all off path } m \in \mathcal{T}, \tag{5}$$

and is off path not because the receiver's inferences are "artificially depressed" but because imitating some other dataset is strictly preferred for the type $t = m$. Therefore, the sender benefits from selective disclosure if and only if they lie – there are no imitation equilibria that increase the payoff of truthful senders relative to their payoff when the receiver is fully informed. On the other hand, truthful senders can suffer – since other senders can dishonestly imitate them, the receiver can be skeptical of their dataset even if they tell the truth.

In addition, for any dataset *not* resembling some raw dataset, $m \notin \mathcal{T}$, there are off-path beliefs

$$q^*(\cdot|m) = q^*(t| \underset{t' \supseteq m,\ t' \in \mathcal{T}}{\arg\min}\ \mathbb{E}_{\beta(\cdot|\sigma^*(t'))}[\theta]) \text{ for all } m \in \mathcal{M} \setminus \mathcal{T}, \tag{6}$$

and given these beliefs, senders never benefit from playing a dataset that the receiver knows for sure to be incomplete. Because of this, an observer of the interaction between senders and receivers would not be able to tell if senders are strategically omitting data simply by looking at the distributions of the published data – some prior about how much data the sender ought to have is necessary to know if observations are being dropped.

We have established that, in an imitation equilibrium, a sender's ability to positively influence the receiver depends on the extent to which they can imitate another state. In turn, this depends on the mass of their own dataset, μ, and the extent to which f_k can be distinguished from f_j, which is given by

$$r_j(k) = \max_{d \in \mathcal{D}} \frac{f_k(d)}{f_j(d)}.$$

This distinguishability factor $r_j(k)$ is a measure of the comparative advantage to a sender under state θ_k to reporting a dataset distributed like f_k, relative to a sender under state θ_j.[11] It can be interpreted to mean that "under state θ_j, a sender would need $r_j(k)$ times as much data to imitate μf_k than under θ_k". A sharp feature of the continuum model is that pairwise distinguishability comparisons fully suffice to summarize the impact of the shape of generating distributions $\{f_j\}_{j=1}^J$ on the imitation equilibrium outcome.

Proposition 1 (Sufficiency)**.** *Two games $\mathcal{G}$ and $\mathcal{G}'$ must yield the same outcome if they share the same state space Θ and priors β_0 and G, and for all j and k,*

$$\max_{d \in \mathcal{D}} \frac{f_k(d)}{f_j(d)} \equiv r_j(k) = r'_j(k) \equiv \max_{d' \in \mathcal{D}'} \frac{f'_k(d')}{f'_j(d')}.$$

In other words, even if $\mathcal{D}$ is very large, $\{f_j\}_{j=1}^J$ only affect the menu of possible beneficial manipulations through a select set of summary statistics, which are each supported by a single point in $\mathcal{D}$. We will delay discussion of the comparative statics of distinguishability, as well as their implications for optimal experimental design, to section 5. In the present section, we leverage these factors to complete our characterization of the imitation equilibrium. All equilibrium outcomes can be described by a vector-valued function $\hat{\mathbf{u}}(\mu) = (\hat{u}_j(\mu))_{j=1}^J$, with $\hat{\mu}_j(\mu) = u_\sigma(\mu f_j)$. The imitation equilibrium outcome has an even simpler description: each sender's messaging problem can be simplified down to the choice of a weakly better state to imitate, $k \in \{j, \ldots, J\}$, and an amount μ

[11] Equivalently, we can consider its inverse, $\frac{1}{r_j(k)}$, an *imitability* factor that describes how easily f_k is imitated under state θ_j.

of that state's distribution to send – "as much as possible" is always weakly optimal, though, as with ironing in example 2.2.1, may not be the only strategy played in equilibrium. Since $\hat{u}$ describes the payoff under every state to every μ, its inverse $\hat{\boldsymbol{\mu}}$, defined as

$$\hat{\mu}_j(u) = \min\{\mu : \hat{u}_j(\mu) \geq u\},$$

describes a *burden of proof* in order to achieve payoff u, and what is necessary is that a type t can provide at least a measure $\hat{\mu}_k(u)$ of distribution f_k, where $\theta_k \geq u$. Crucially, fixing the pairwise distinguishability factors, optimality of the sender's imitation strategy amounts to saying that a sender that achieves payoff u via imitation is either truthful with $\theta_j \geq u$ or imitates another state $\theta_k > u$ in the set

$$A_j(\mu) = \left\{\theta_k : k \in \arg\max_{k>j} \hat{u}_k\left(\frac{\mu}{r_j(k)}\right)\right\}.$$

Theorem 2 (Existence and uniqueness)**.** *There there exists an essentially*[12] *unique imitation equilibrium, implemented by a vector-valued burden of proof function* $\hat{\boldsymbol{\mu}} : [0, \theta_J] \to \mathbb{R}^J$ *with outcome* $\hat{\mathbf{u}}$ *such that*

1. $\hat{u}_j(\mu)$ *is continuous and (weakly) increasing in* μ *for all* j*.*
2. $\sigma^*(\mu f_j)$ *is supported on* $\left\{\mu' f_k : \mu' = \hat{\mu}_k\left(\hat{u}_k\left(\frac{\mu}{r_j(k)}\right)\right) \text{ and } \theta_k \in A_j(\mu)\right\}$.

3.1 Construction of the equilibrium

In the Appendix, we give the details of the step-by-step construction of σ^* in general. But to capture the main idea, consider a minimal setup that illustrates the forces at play. Suppose we face the problem of constructing $\hat{\boldsymbol{\mu}}(v)$ for $v \in [\theta_{J-1}, \theta_J]$, assuming that $\hat{\boldsymbol{\mu}}(\theta_{J-1})$ is known. Fig. 4 shows that a typical type space can be projected onto 2 dimensions: one dimension describes the ability of each type to imitate f_J, given by $\frac{\mu}{r_j(J)}$, and the other dimension describes their ability to imitate f_{J-1}, given by $\frac{\mu}{r_j(J-1)}$. We can plot senders with all possible amounts of data under a given state as a ray when we describe the type space this way. Since any $v > \theta_{J-1}$ is obtained through imitating one of these two types, this description is sufficient to determine the imitation strategies used to obtain this subset of responses from the receiver.

The burden-of-proof vector lies in the same space and describes two simple things: which of the two states each type imitates, and what the highest action is that they can induce the receiver to take by doing so. A couple of observations allow us to identify the unique continuation of $\hat{\boldsymbol{\mu}}(v)$ at and to the left of any v^* whenever $\hat{\boldsymbol{\mu}}(v)$ is already known for all $v > v^*$.

Taking the higher payoff frontiers to be fixed, focus on the set of types unable to meet any

[12] β^* is uniquely determined, and σ^* is uniquely determined up to payoff-irrelevant mixing probabilities.

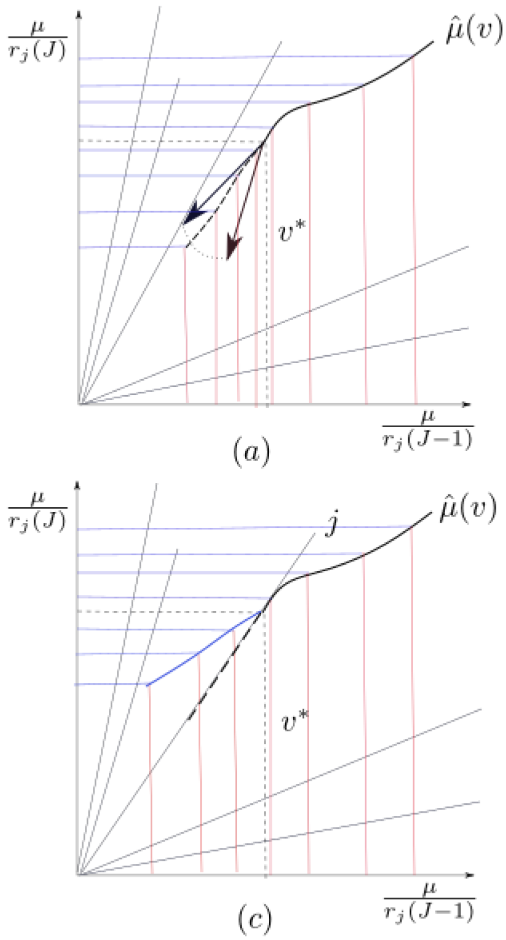


Figure 4: In equilibrium, $\hat{\boldsymbol{\mu}}(v)$ equalizes payoffs to imitating each state $\theta_k > v$. Rays represent types in $\mathcal{T}$ and red and blue lines represent payoff frontiers to those imitating f_{J-1} and f_J, respectively.

component of $\hat{\boldsymbol{\mu}}(v)$ for any $v > v^*$. There may exist within this set a self-separating set of positive measure that can pool with each other to induce action v^*. Fig. 4(b) shows that if so, $\hat{\boldsymbol{\mu}}(v)$ is discontinuous at v^*, since the equilibrium construction then immediately pools these types and assigns them all a payoff of v^*. Otherwise, $\hat{\boldsymbol{\mu}}(v^*) = \lim_{\epsilon \to 0} \hat{\boldsymbol{\mu}}(v^* + \epsilon)$.

The key fact is that given $\hat{\boldsymbol{\mu}}(v^*)$, it is always possible to exactly specify $\hat{\boldsymbol{\mu}}(v)$ for v in some, possibly small, nonempty interval $(v^* - \Delta, v^*)$. Consider first the case in which all types in the v^*-payoff frontier strictly prefer to imitate either f_J or f_{J-1}. When $\hat{\mu}_j(v^*)f_j$ imitates distribution f_J, then all types μf_j with μ close to $\hat{\mu}_j(v^*)$ behave likewise, and the same is true for those imitating distribution f_{J-1}. In other words, the payoff frontiers are locally determined because imitation strategies are fixed, up the amount of data submitted. Panel (a) of Fig. 4 shows that $\hat{\boldsymbol{\mu}}(v)$ then follows along the path of equivalent payoffs from imitating either state, and is continuous, due to the continuity of g.

A second possibility is that for some j, type $\hat{\mu}_j(v^*)f_j$ may indeed be indifferent between imitating f_J and f_{J-1}, and mixes between the two with interior probability. Locally, for μ close to $\mu_j(v^*)$, the types μf_j must also be indifferent, and so for a set of values $v \approx v^*$, $\hat{\boldsymbol{\mu}}(v)$ coincides with the set of types under state θ_j that achieve the corresponding payoff. For all state-θ_j senders obtaining a payoff in this range, the mixed strategy played equalizes payoffs to imitating each of the two highest states. Fig. 4(c) shows that if $\sigma(\frac{\mu}{r_j(J)} f_J | \mu f_j)$ increases too quickly, this fails to hold, since then payoffs to imitating θ_J decrease quickly relative to those to imitating θ_{J-1}, and (d) shows that payoffs to imitating θ_J decrease too quickly in the opposite case. There is, then, a unique continuation of the mixed strategy that respects the restriction on $\hat{\boldsymbol{\mu}}$, and it is continuous due to the continuity of g.

When there are more than 2 candidate states to imitate, the construction is slightly more complicated in that there may be more than one state under which types are indifferent across distributions to imitate, and a given type may be indifferent between imitating more than 2 different states. Nevertheless, the idea is the same. It is always possible to construct an interval of frontiers and their associated equilibrium strategies, given knowledge of higher-payoff frontiers. The construction technique then proceeds interval-by-interval, where we note that each interval formed in a step of the process is nonempty but may be small: it may be necessary to switch from handling the problem as in the first case to handling it as in the second case, and vice versa, multiple times as the algorithm proceeds to successively lower payoff frontiers.

3.2 A separation theorem

Let us return briefly to the matter of why the imitation outcome stands out from other equilibrium outcomes. It turns out that, although we can construct the imitation equilibrium payoff frontiers iteratively, we can also characterize them each individually, and independently of the remainder of the equilibrium. Put simply, imitation equilibrium payoff frontiers universally divide the type space into a greater-value upper region and a lesser-value lower region, and they are the only frontiers to do so.

We start with some definitions.

Definition 3.1. *An* upper pool *of payoff frontier* $\hat{\boldsymbol{\mu}}(v)$ *is a set*

$$\bar{T} = U(\hat{\boldsymbol{\mu}}(v)) \setminus U(M)$$

for some collection of messages M*.*

Definition 3.2. *A* lower pool *of of payoff frontier* $\hat{\boldsymbol{\mu}}(v)$ *is a set*

$$\underline{T} = U(M) \setminus U(\hat{\boldsymbol{\mu}}(v))$$

for some collection of messages M*.*

An upper pool consists of all types above the payoff frontier $\hat{\boldsymbol{\mu}}(v)$ but below some other frontier, while a lower pool consists of types below it but above another frontier.

We define the pooled value of any set of types, $u_{pool}(T)$, to be the receiver's expectation of the state given that the sender's type is in the set T, and state the separation theorem:

Theorem 3 (Separation)**.** *For any nonempty upper pool* $\bar{T}$ *and lower pool* $\underline{T}$ *of* $\hat{\boldsymbol{\mu}}(v)$*,*

$$u_{pool}(\bar{T}) \geq v > u_{pool}(\underline{T}).$$

In other words, upper pools are weakly improving and lower pools are strictly worsening — for any subset of $\mathcal{T}$ that is bounded by two frontiers and contains $\hat{\boldsymbol{\mu}}(v)$, the value of the part above $\hat{\boldsymbol{\mu}}(v)$ is at least v, while the value of the part below is less than v.[13]

The fact that upper pools are improving is a consequence of the conditions of imitation equilibria: the property holds because in each group of senders who send the same message under σ^*, only those with worse-than-average values can be truncated by excluding $U(M)$. On the other hand, the equilibrium we construct has worsening lower pools because in it, any potentially self-separating pool of senders below $\lim_{\epsilon \to 0} \hat{\boldsymbol{\mu}}(v + \epsilon)$ that achieves a value of at least v must lie above the frontier $\hat{\boldsymbol{\mu}}(v)$.

These properties guarantee uniqueness of the imitation equilibrium outcome if we use them to compare outcomes under σ^* and another PBE, σ. If the outcome under σ differs from that under

[13]The former inequality is weak and the latter strict because we have defined the imitation payoff frontiers such that, when there are multiple types μf_k that all achieve v, $\hat{\mu}_k(v)$ is the lowest such μ.

σ^*, then worsening lower pools under σ^* imply that there is a frontier with a worsening upper pool under σ. Moreover, the only frontiers in $\mathcal{T}$ that satisfy either property are the frontiers of σ^*. Given any prospective frontier and its associated payoff, checking either of these properties in isolation is enough to verify that it shows up in the imitation equilibrium, and may in some cases be easier than constructing the entire imitation equilibrium outcome.

The separation theorem is a general result — it also applies to finite evidence games, where it is related to the "downward biased" characterization of Rappoport (2022). In all these cases, worsening lower pools rules out credible inclusive announcements, and improving upper pools turns out to imply that no other equilibrium is credible inclusive announcement-proof.

4 Comparative statics

All imitation outcomes share some concrete features. Here, we present comparative statics of the sender's reports in μ, of the sender's welfare with respect to the receiver's prior belief about θ and μ, and of separation as $Var[\mu] \to 0$. We begin with a corollary to Theorem 2.

Corollary (to Thm. 2)**.** *Under σ^*, there are thresholds $z_j^* > z_j^{**}$ for each state such that:*

- *Whenever the sender's type is μf_j with $\mu > z_j^*$, the sender masquerades as a higher type, and receives a payoff $\hat{u}_j(\mu) > \theta_j$.*
- *Whenever $\mu \in (z_j^{**}, z_j^*]$, the sender is honest and the receiver knows it upon receiving the data: $\hat{u}_j(\mu) = \theta_j$.*
- *Whenever $\mu \leq z_j^{**}$, the sender is honest, but the receiver believes they are a worse type with positive probability, and $\hat{u}_j(\mu) < \theta_j$.*

We can think of senders with $\mu > z_j^*$ as high-data senders, with enough data to benefit from manipulating their data against the receiver's uncertainty about their data endowment. The costs of voluntary disclosure are borne by low-data senders, those with $\mu < z_j^{**}$, who the receiver is skeptical of even when they are truthful. These thresholds vary by j, and in particular, $z_1^* = 0$ and $z_J^* = 1$. However, they need not be monotone in j.

The potential presence of an intermediate, full-information interval between disjoint upper and lower partial-pooling intervals when we fix θ and vary μ is a novel feature of these equilibria that occurs when there are multiple imitated states with different distinguishing outcomes. It is a consequence of the fact that it requires a strictly greater amount of data to benefit from imitating a different state than it does to send one's full dataset and discourage all imitators. The structure of pooling and separation contrasts with strategies in binary-state models of voluntary disclosure, or in models with ordered outcomes. In those cases, full separation only occurs at the very top, that is, for types with a maximal state and a maximal amount of evidence (see, for example, Dye

(1985) and Dzuida (2011)). We show that this doesn't have to be true in general: although they remain able to separate, types with the most evidence are often more tempted to pool with others.

Partial pooling occurs in the receiver-optimal equilibria in our model because uncertainty about μ makes it impossible for lower-data, higher-state senders to separate from higher-data, lower-state senders. In the absence of uncertainty about μ (that is, if μ is commonly known to the sender and the receiver) the disclosure game is a case of the games studied by Grossman (1981) and Milgrom (1981), in which unraveling occurs. The distribution of μ in our model, while assumed to be nondegenerate, can be arbitrarily close to a point mass, and outcomes converge to the full-information outcome as the receiver's uncertainty about μ vanishes.

Claim 2. *As $Var[\mu] \to 0$, we have $Pr(|\hat{u}_j(\mu) - \theta_j| > \epsilon) \to 0$ for all j.*

This should be unsurprising: when the receiver knows μ quite well, any dataset with fewer-than-expected observations is quite suspicious and is heavily discounted, and this limits the returns to omitting data.

4.1 Complementarity with public information

Next, we show that public information can complement voluntary disclosure. As a preliminary, observe that the receiver is worse off given uncertainty about either the payoff-relevant state, or its relation to the experiment. Suppose that there are any two games $\mathcal{G}(\Theta, \mathcal{D}, \beta_0, \{f_j\}_{j=1}^J, G)$ and $\mathcal{G}'(\Theta', \mathcal{D}, \beta_0', \{f_{j'}'\}_{j'=1}^J, G)$ that have the same outcome space $\mathcal{D}$ and data-mass distribution G,[14] but describe different sets of possible states. If the receiver knows that the true state might be in Θ or Θ', but is uncertain of which and has prior α, $1-\alpha$, respectively, of the likelihood of each case, then they can be modeled as playing a third game, $\mathcal{G}^{uc}$, in which the set of states is $\Theta^{uc} = \Theta \bigcup \Theta'$ with each state retaining its data-generating distribution. Their prior over Θ^{uc} is given by

$$\beta_0^{uc}(\theta_j) = \alpha\beta_0(\theta_j) \ \text{ for } \theta_j \in \Theta, \qquad \beta_0^{uc}(\theta_{j'}') = (1-\alpha)\beta_0(\theta_{j'}') \ \text{ for } \theta_{j'}' \in \Theta'.$$

Claim 3. *The receiver's expected payoff in $\mathcal{G}^{uc}$ conditional on $\theta \in \Theta$ is less than their expected payoff in $\mathcal{G}$, and strictly so if there is a type of the sender with state in Θ for which the outcomes differ in the two games.*

Greater ex-ante uncertainty, in any of a number of dimensions, is generally worse for the receiver. When the receiver suffers from not knowing whether the state is in Θ or Θ', it is because a state $\theta_{j'} \in \Theta'$ differs from a state $\theta_j \in \Theta$ for one or more of the following reasons:

1. Its numerical value is different;

[14] We prove the claim that follows in a more general setup, in which the the games need not have the same data-mass distribution and the distribution of μ can be state-contingent, that is, the two games have state-contingent data mass distributions $\{G_j\}_{j=1}^J$, $G_{j'}'\}_{j=1}^{J'}$, respectively.

2. The prior likelihood that the receiver assigns to it is different, $\beta_0(\theta_j) \neq \beta_0'(\theta_{j'})$;

3. The distribution of data it generates is different, $f_j \neq f_{j'}'$.

Applying Claim 3 to the first case shows that the receiver is worse off when they understand the set of possible states of the world corresponding to different distributions of experimental outcomes, but are uncertain about the optimal action to take even with full knowledge of the distribution of the data. In the second case, we see that having an incorrect prior about the state of the world cannot benefit the receiver in expectation.

The third case is perhaps the most interesting: it captures a receiver's uncertainty about the distribution of outcomes conditional on the true, payoff-relevant state. We illustrate how this could play out in the context of our previous example (Ex. 2.2.1), by expanding the state space to include not only $\Theta = \{\theta_H, \theta_L\}$, but also $\Theta' = \{\theta_{H'}, \theta_{L'}\}$, where θ_H' and θ_L' are analogous to θ_H and θ_L, respectively, except the baseline rate at which an out-of-the-box prediction model accurately predicts no rain is elevated (0.45 instead of 0.4). Suppose that this difference does not affect the receiver's optimal action, so $\theta_{H'} = 1$ and $\theta_{L'} = 0$.

j	$f_j(1)$	$f_j(2)$	$f_j(3)$	$f_j(4)$
H	0.6	0.2	0.1	0.1
L	0.4	0.2	0.3	0.1
H'	0.65	0.2	0.05	0.1
L'	0.45	0.2	0.25	0.1

Table 3: The generating distribution of outcomes under states θ_H, θ_L, $\theta_{H'}$, and $\theta_{L'}$.
Recall outcome 1 is *(predict no rain, no rain)*, outcome 2 is *(predict rain, rain)*, outcome 3 is *(predict rain, no rain)*, and outcome 4 is *(predict no rain, rain).*

It is almost immediate to see that the receiver must make different inferences if they place 50% probability on the event that the state is in Θ', relative to certainty that the state is in Θ – for instance, they can no longer know that the true state is θ_H if the sender sends $m = \mu f_H$ for $\mu \in (2/3, 3/4)$, because a sender with mass $\mu' \in (8/9, 1)$ of data distributed like $f_{L'}$ could also have sent such a message. Since the equilibrium outcomes for the sender are not the same, we know that the receiver's uncertainty about the baseline accuracy of prediction models makes them strictly worse off.

Strategic disclosure can also impair the receiver's ability to extract information from the sender that would make up for a lack of ex-ante knowledge. In a world with full disclosure, this could be achieved in the above example by adding a "control" arm of the experiment that provides statistics on the baseline performance of an out-of-the-box prediction model for comparison with the performance of the sender's preferred model. However, the same is not true under voluntary disclosure, because the sender can also omit data about the performance of the out-of-the-box model.[15]

Put differently, the receiver can't be hurt by observing a signal ϕ that tells them whether the game is $\mathcal{G}$ or $\mathcal{G}'$, and this is true of any finite public signal. A signal that is not informative about θ per se may still be valuable alongside voluntarily disclosed data if it informs the receiver about how to interpret the data, and the signal need not be fully informative about any aspect of the experiment to yield a strict benefit.

Claim 4. *Let ϕ be a finite-valued public signal that is only informative about $\mathcal{D}$ and $\{f_j\}_{j=1}^{J}$. It strictly benefits the receiver for some prior β_0 if and only if two distinct realizations $\hat{\phi}$ and $\hat{\phi}'$ induce games $\mathcal{G}$ and $\mathcal{G}'$ such that* $\arg\max_d \frac{f_k(d)}{f_j(d)} \neq \arg\max_d \frac{f_k'(d)}{f_j'(d)}$ *for some $k > j$.*

These kinds of signals complement disclosed data by making the receiver less susceptible to manipulations of their auxiliary beliefs through data omission. They might pertain to any jointly estimated covariates. In addition to base rates, voluntarily disclosed experimental data is more useful in the presence of trustworthy information about the space of underlying (i.e., not cherry-picked) outcomes, the likelihood of randomization to treatment or control (in the case of an RCT or an A/B test), or the composition of the trial population (when experimental outcomes are heterogeneous depending on group membership).

4.2 Impact of beliefs on the sender

When the receiver's belief about the ex-ante probability of a given state θ_j increases relative to others, the receiver's skepticism weakly increases for all messages that yield a higher payoff to the sender than full certainty of that state. The reverse is true of all messages that yield a lower payoff than θ_j. An increase in the probability of θ_j therefore "pulls" the receiver's action towards θ_j given any message, which has the consequence of decreasing ex-post payoffs for all types of the sender that would originally have achieved $\hat{u}_j(\mu) \geq \theta_j$, and increasing them if originally, $\hat{u}_j(\mu) \leq \theta_j$.

To formalize this, let $\mathcal{G}$ be a disclosure game with prior β_0 about θ and $\mathcal{G}'$ be a game that is identical except for the prior β_0' which differs from β_0, with $\beta_0'(\theta_j) > \beta_0(\theta_j)$ and $\frac{\beta_0(\theta_k)}{\beta_0(\theta_{k'})} = \frac{\beta_0'(\theta_k)}{\beta_0'(\theta_{k'})}$ for all other k, k'.

Claim 5. *Suppose that $\hat{\mathbf{u}}, \hat{\mathbf{u}}'$ are imitation equilibrium outcomes of $\mathcal{G}$ and $\mathcal{G}'$, respectively. Then $\hat{u}_{j'}(\mu) \geq \hat{u}_{j'}(\mu)$ whenever $\hat{u}_{j'}(\mu) \geq \theta_j$, and $\hat{u}_{j'}(\mu) \leq \hat{u}_{j'}(\mu)$ whenever $\hat{u}_{j'}(\mu) \leq \theta_j$.*

Lastly, we point out that MLRP shifts in the receiver's beliefs have a monotone impact on the

[15]Specifically, the experiment that randomizes a day's weather to be predicted either with the standard prediction model or the new one with 50-50 odds, which tracks additional outcomes 5, 6, 7, and 8 which are analogous to outcomes 1, 2, 3, and 4 in the event the standard algorithm is used instead of the new one, does not change outcomes relative to the experiment in Table 3.

sender's welfare. Simply put, every type of the sender benefits from a monotone likelihood ratio shift in the receiver's belief about the state, and suffers from a monotone likelihood ratio shift in their belief about the data-mass distribution. Intuitively, an upwards shift in the prior distribution in θ makes the receiver more willing to believe a claim that the state is high, and we show this formally in Appendix B. Such a shift unambiguously benefits the sender both conditional their realized dataset, and ex-ante. On the other hand, when the receiver expects μ to be greater, they are more *skeptical*: they infer a greater likelihood that a given message may have been selected from a larger dataset.[16]

Claim 6. *If two disclosure games $\mathcal{G}$ and $\mathcal{G}'$ are identical except for priors $\beta_0 \leq_{MLRP} \beta_0'$ and $g \geq_{MLRP} g'$, then*

$$\hat{u}_j(\mu) \leq \hat{u}_j'(\mu) \quad \forall \mu f_j \in \mathcal{T}.$$

5 Experimental design

Our results highlight that the quality of the information the receiver obtains depends on how the data-generating process distinguishes states. This section focuses on interventions that aim to maximize distinguishability, and proposes a framework for optimally designing experiments to allow the receiver to extract payoff-relevant information from the sender through voluntary disclosure. In our model, an *experiment* is the data-generating process that provides the sender with their raw dataset, and is captured by a tuple $\mathcal{E} = (\mathcal{D}, \{f_j\}_{j=1}^J)$ consisting of the space of reported outcomes and the generating distribution of data over them. We assume that the remaining primitives of the game – state space, payoffs, and priors – are fixed, and consider the effect of varying the experiment that the sender observes.

A key fact is that whenever an experiment makes states pairwise more distinguishable, the receiver's welfare improves. Intuitively, increasing distinguishability allows higher-state types to separate themselves more effectively from lower-state types who would imitate them. The resulting equilibrium does not better separate every type from every other type – indeed there are types that would play different messages under one experiment that would play the same message in the other, in both directions – but, given the receiver's single-peaked expected utility, the more distinguishing experiment always makes the receiver better able to target the optimal action.[17]

[16] Rappoport (2022)'s result can be used to show that the latter holds in finite-data games viewed as an instance of an abstract evidence game, and a similar argument shows that this is directly true in the continuum.

[17] The proof that distinguishability improves payoffs uses the fact that a mechanism designer that takes a sender's submitted dataset as a report is weakly more constrained by a sender's ability to deviate to sending a false dataset if the experiment has poor distinguishability. If the receiver does not have single-peaked preferences, then the imitation equilibrium outcome and the outcome of the optimal mechanism do not necessarily coincide, and increasing distinguishability may force the receiver to take a higher action after observing a message that few low-state types can imitate, when they would instead like to commit to responding to it with a lower action.

Proposition 4 (Improvement)**.** *Suppose two experiments $\mathcal{E}$ and $\mathcal{E}'$ yield imitation equilibrium actions a and a', respectively.*

- *If $r'_j(k) \geq r_j(k)$ for all $k > j$, then $\mathbb{E}_{a,\theta}[u_r(a)] \leq \mathbb{E}_{a',\theta}[u_r(a')]$.*
- *If, in addition, $r'_j(k) > r_j(k)$ for some j, k such that there is some μf_j imitating $\hat{\mu} f_k$ under the imitation equilibrium with experiment $\mathcal{E}$, then $\mathbb{E}_{a,\theta}[u_r(a)] < \mathbb{E}_{a',\theta}[u_r(a')]$.*

In fact, by making a state arbitrarily distinguishable from others, we can guarantee that a sender under that state elicits at least their full-information action with high probability: the imitation equilibrium guarantees that . In the limit as all states become highly distinguishable, the receiver also approximately attains their full information payoff. On the other hand, if all states are negligibly distinguishable, the receiver learns essentially nothing, even as the sender is fully informed.

Claim 7. *With very high and very low distinguishability, outcomes approach those under full information and no information, respectively: for all $\epsilon, \delta > 0$,*

- *There exists $\underline{R} < \infty$ such that if $r_j(k) > \underline{R}$ for all $j < k$, then $Pr(|\hat{u}_k(\mu) - \theta_k| > \epsilon) < \delta$*
- *There exists $\bar{R} > 1$ such that if $r_j(k) < \bar{R}$ for all $j < k$, then $Pr(|\hat{u}_k(\mu) - \mathbb{E}_{\beta_0}[\theta]| > \epsilon) < \delta$*

where the likelihood is taken over realizations of μ.[18]

One way to better distinguish two states is to undertake a more detailed experiment. Without changing the experimental technology – that is, the underlying likelihood of events under different states – a researcher could investigate and record a more detailed set of outcomes in order to obtain finer data. To formalize this, suppose there is an existing outcome space $\mathcal{D}$, and consider a notion of a more elaborate outcome space $\mathcal{D}'$ that the researcher can obtain by splintering an existing outcome into multiple sub-outcomes to track.

Definition 5.1. *If there are two experiments $\mathcal{E} = (\mathcal{D}, \{f_j\}_{j=1}^J)$ and $\mathcal{E}' = (\mathcal{D}', \{f'_j\}_{j=1}^J)$ and a partition $\mathcal{P} = \{P_d\}_{d \in \mathcal{D}}$ of $\mathcal{D}'$ such that*

$$\sum_{d' \in P_d} f'_j(d') = f_j(d)$$

for all d in $\mathcal{D}$, then $\mathcal{E}'$ splinters *the outcome space of $\mathcal{E}$ and $\mathcal{E}$* merges *the outcome space of $\mathcal{E}'$.*

Immediately, we observe that for all θ_j and θ_k,

$$\max_{d' \in P_d} \frac{f'_k(d')}{f'_j(d')} \geq \frac{f_k(d)}{f_j(d)},$$

[18] Despite the fact that $\bar{R} < \underline{R}$, we use this notation because $\bar{R}$ is an upper bound and $\underline{R}$ is a lower bound on the distinguishability sufficient for each case.

and so $r_j(k) \geq r'_j(k)$ whenever $\mathcal{D}'$ splinters $\mathcal{D}$.

Claim 8. *Splintering the outcome space weakly improves the receiver's expected payoff.*

In some cases, there is a most elaborate possible experiment $\mathcal{E}^*$, i.e., one that is a splintering of every other possible experiment. Suppose that costs and constraints on gathering, storing, and transmitting data are negligible. Then it is optimal for a designer who acts on behalf of the receiver to choose the most elaborate possible experiment. If instead the sender chooses the experiment, then the receiver should, if possible, incentivize the sender to choose the most detailed experiment by committing to accept nothing else. Since they follow a simple rule of thumb, these recommendations don't require detailed knowledge of the true data-generating process, and would be easy for even an uninformed designer to implement.

On the other hand, in practice there is often no binding limit to the number of ways that an experiment can be refined and complicated, at ever increasing cost. Despite the fact it never hurts, further splintering a dataset does not always strictly improve distinguishability. With precise information about the data-generating process, Proposition 1 allows us to identify instances when it is without loss to the receiver to merge outcomes relative to $\mathcal{E}^*$.

Proposition 5 (Merging)**.** *Suppose that* $\mathcal{E}^* = (\mathcal{D}^*, \{f_j^*\}_{j=1}^J)$*. Let* $S^* = \bigcup_{j<k} \arg\max_d \frac{f_k^*(d)}{f_j^*(d)}$*. Then merging all outcomes in* $\mathcal{D}^* \setminus S^*$ *does not change the imitation equilibrium outcome.*

The set S^* consists of all outcomes that maximally distinguish one state from another. Mmerging other outcomes is without loss because S^* is sufficient to maximize every distinguishability factor in $\{r_j(k)\}_{j<k}$. We are left, generically, with a minimal experiment that suffices to reveal as much payoff-relevant information as possible to the receiver robustly over all possible priors.

Claim 9 (Minimality)**.** *Fix* u_r*,* Θ*, and* g*, suppose that* $\mathcal{E}$ *is obtained from* $\mathcal{E}^*$ *by merging* S^* *and that* $\mathcal{E}'$ *merges some outcomes in* $\mathcal{E}$*, and suppose that* $\arg\max_d \frac{f_k^*(d)}{f_j^*(d)}$ *is unique for all* $j < k$*.*

Then there exists β_0 *such that the receiver is strictly better off with* $\mathcal{E}$ *than with* $\mathcal{E}'$*.*

This simplification of the experiment can be quite drastic, and in some familiar cases, including the case of a binary state space or an outcome space ordered by the monotone likelihood ratio property (MLRP), S^* is a singleton with only one "good news" outcome that maximally distinguishes higher states from lower ones, while all other outcomes in $\mathcal{D}^*$ can be merged and essentially ignored.[19] Formally, we say that $\mathcal{D}^*$ satisfies MLRP with respect to $\{f_j^*\}_{j=1}^J$ if, for any $j < k$ and

[19] The imitation equilibrium in these cases has the same outcome as a sanitation equilibrium (Shin (2003)) in which the sender only reports observations of the outcome in S^*, and omits all others; however, it differs in that imitating senders generally report a positive mass of observations of these outcomes anyways, with no impact on the receiver's inferences.

$d < d'$,

$$\frac{f_k^*(d')}{f_j^*(d')} > \frac{f_k^*(d)}{f_j^*(d)}.$$

It is straightforward to see that MLRP implies that S^* comprises of the single maximal element in $\mathcal{D}^*$.

Even when $J > 2$, some degree of dimensionality reduction is often possible, especially if $J << |\mathcal{D}^*|$. In general, $|S^*| \leq \frac{J(J-1)}{2}$. The 3-state example 2.2.2 gives an instance in which this bound is tight because the maximizer, $\arg\max_d \frac{f_k(d)}{f_j(d)}$, is unique for all pairs $j < k$.

Corollary (to Prop. 5)**.** *The minimal optimal experiment tracks at most $\frac{J(J-1)}{2} + 1$ outcomes, and furthermore, if $\mathcal{D}^*$ satisfies MLRP with respect to $\{f_j^*\}_{j=1}^J$, then a binary outcome space suffices.*

6 Relationship to finite data

In the big picture, the purpose of modeling communication in this stylized, continuous-data disclosure game is to understand how senders will volunteer data in real-world disclosure settings, in which datasets are always finite. The comparative statics of section 4 and the experimental design implications of the previous section depend on the fact that datasets are well-described by μ and f_j, which is exactly true only in the continuum, but nearly true with large N in such a way that those results approximately carry over. This section makes precise the finite-data settings that we aim to approximate, and describes how the continuous-data model captures their regularities in the limit.

We model a sender who has access to a finite dataset of n i.i.d. observations drawn from $\mathcal{D}$ according to the state-contingent distribution f_j. The size of the sender's dataset is upper-bounded by N, but the sender may have access to $n < N$ observations as well, and the receiver is uninformed about how much data the sender has. Nature's sequence of moves in drawing the sender's dataset is: 1) draw the state, θ_j, according to prior β_0; 2) draw the number of observations, n, from distribution G_N; 3) for each of the n datapoints, draw their realized value i.i.d. from f_θ. Call the disclosure game with these parameters $\mathcal{G}_N(\Theta, \mathcal{D}, \beta_0, \{f_j\}_{j=1}^J, G)$. The data mass distributions $G_N(\cdot)$ capture the receiver's uncertainty about how much raw evidence the sender has, prior to selecting observations to reveal: for example, there may be uncertainty about the number of total trials in an experiment, or the number of trials out of N attempted that survived the entire trial period.

The sender's dataset is the empirical probability mass function $t = \frac{1}{N}(t_1, \ldots, t_D)$, where t_d is the number of observations of outcome d and $n(t) = \sum_{d=1}^D t_d$ is the number of observations they get. They are able to send any subset of their dataset as a message to the receiver, where

$$m \tilde{\subseteq} t \;\Leftrightarrow\; m_d \leq t_d \; \forall d \in \mathcal{D}.$$

In summary, the type space is $\mathcal{T}_N = \bigcup_{n=0}^{N} \mathcal{D}^n$, with type distribution

$$q_N(t) = \frac{n(t)!}{\Pi_{d=1}^{D} t_d!} \sum_{j'} \beta_0(\theta_{j'}) g_N(n(t)) \Pi_{d=1}^{D} f_{j'}(d)^{t_d},$$

and the message space $\mathcal{M}_N$ is identical to the space of types.

When datasets are finite, the sender's dataset does not perfectly inform them about the state: when f_j all have full support, any state is possible after observing any dataset. The likelihood of θ_j given that the raw dataset is t is

$$\pi_N(\theta_j|t) = \frac{\beta_0(\theta_j) g_N(n(t)) \Pi_{d=1}^{D} f_j(d)^{t_d}}{\sum_{j'} \beta_0(\theta_{j'}) g_N(n(t)) \Pi_{d=1}^{D} f_{j'}(d)^{t_d}},$$

and so, when the receiver observes a message and updates their belief about the sender's type to $q_N(t|m)$, their posterior about the state updates to

$$\beta(\theta_j|m) = \frac{\sum_{t \in \mathcal{T}_N} q_N(t|m) \pi_N(\theta_j|t)}{\sum_{t \in \mathcal{T}_N} q_N(t|m)}. \tag{7}$$

We highlight that the distribution of datasets in the finite-data setting converges to the distribution of datasets in a continuous-data model. In particular, $g(\mu)$ represents the likelihood of obtaining a fraction μ of total potential data under state j, and analogously, $\frac{n}{N}$ is the fraction of total data available to the sender in the finite-data game. We can study a sequence of games such that as N increases, $NG_N(\frac{n}{N}) \rightarrow_{unif.} G(\mu)$, and note that if so, the type distributions also converge uniformly: $q_N \rightarrow_{unif.} q$.

Definition 6.1. $\mathcal{G}(\Theta, \mathcal{D}, \beta_0, \{f_j\}_{j=1}^{J}, G)$ *is the* limit game *for a sequence of finite-data games* $\{\mathcal{G}_N(\Theta, \mathcal{D}, \beta_0, \{f_j\}_{j=1}^{J}, G_N)\}_{N=1}^{\infty}$ *if* $NG_N(\frac{n}{N}) \rightarrow_{unif.} G(\mu)$.

Despite the fact that the type distributions converge, the type space $\mathcal{T}_N$ is drastically different from $\mathcal{T}$: in particular, $\mathcal{T}_N \sim \mathcal{M}_N$ and both approximately span a D-dimensional space of datasets for large N, while $\mathcal{T}$ is only 2-dimensional, as every dataset is described by μ and θ. While datasets far away from $\mathcal{T}$, that have distributions unlike the data-generating distribution in any state, become vanishingly unlikely as N grows large, they are never impossible except in the limit; this is why the continuum model is much easier to work with.

It remains possible to describe an imitation equilibrium and a truth-leaning equilibrium in the finite-data setting. The finite-data model is a special case of the evidence model in Hart et al. (2017) and Rappoport (2022). The former shows that truth-leaning equilibria exist and are unique and receiver-optimal in the finite-type setting, and also that they are always outcome-equivalent to imitation equilibria, although it does not guarantee that the strategies are equivalent. The latter includes an iterative algorithm to compute these equilibria; the number of steps is, however, exponential in $|\mathcal{T}_N|$, and as far as we can tell, there is no obvious way to obtain a significantly more efficient closed-form solution.

We can instead establish that the imitation equilibrium of the continuous-data model gives a perfect approximation to the limit outcome of communication in truth-leaning equilibria of finite-data games as $\mathcal{G}_N$ converge.[20] To make the comparison, the notion of an outcome should be extended across type spaces. There is a global data space $[0,1] \times \Delta\mathcal{D}$, invariant to N, that contains $\mathcal{T}_1, \ldots$ and $\mathcal{T}$ as long as they all share a space of observations. Recall that $u_{\sigma^*}(t)$, the outcome of the game for type t, is their payoff from the best feasible message given equilibrium beliefs. If $t \in [0,1] \times \Delta\mathcal{D}$, it need not also be in the literal type set for the outcome to be well-defined, since we can already infer whether t can feasibly send a message from the subset relation on $[0,1] \times \Delta\mathcal{D}$. The outcome to the hypothetical type can be understood as a thought experiment: "if the receiver believes we are playing a game with equilibrium σ^* or σ^*_N, and my dataset is t, what is the best payoff I can attain, even if t is inconsistent with the receiver's perceived game?"

Definition 6.2. *A sequence of equilibria $(\sigma_1, \sigma_2, \ldots)$ of games $\{\mathcal{G}_N(\Theta, \mathcal{D}, \beta, \{f_j\}_{j=1}^J, G_N)\}_{N=1}^\infty$ has outcomes that* converge *to the outcome of an equilibrium σ of the limit infinite-data game $\mathcal{G}(\Theta, \mathcal{D}, \beta, \{f_j\}_{j=1}^J, G)$ if the payoffs $u_{\sigma^*_N}(t)$ converge uniformly to $u_{\sigma^*}(t)$ over types in $\mathcal{T}$.*

Theorem 6. *If $\mathcal{G}$ is the limit game for finite-data games $\mathcal{G}_1, \mathcal{G}_2, \ldots$ with $N = 1, 2, \ldots$ respectively, then the truth-leaning equilibrium outcomes in $\mathcal{G}_1, \mathcal{G}_2, \ldots$ converge to the imitation equilibrium outcome of $\mathcal{G}$.*

Outcome convergence shows that it's reasonable to use the limit game to describe the distribution of actions the receiver takes after the sender discloses a large dataset, as well as the mapping from the truth to the receiver's inferences. At a high level, the proof follows from the convergence of type distributions $\mathcal{T}_N$ to $\mathcal{T}$, and from the separation theorem, which holds as well in truth-leaning equilibria of finite-data games. Appendix E gives the formal argument and shows that the limit equivalence result partially extends to strategies, in addition to outcomes.

In addition, outcome convergence shows that previous sections' results on comparative statics and experimental design hold approximately for large finite datasets. When the number of obser-

[20]We state the definition of convergence and the theorem below in terms of $N = 1, 2, \ldots$ rather than an arbitrary sequence of dataset sizes $N_1, N_2, \ldots$ only for the sake of notational brevity. The theorem applies just as well to any sequence of games $\{\mathcal{G}_{N_i}\}_{i=1}^\infty$ of increasing dataset size with uniformly convergent data distributions, since any such sequence is a subsequence of a convergent sequence of games $\{\mathcal{G}_N\}_{N=1}^\infty$.

vations is finite, splintering the data always leads to a strict improvement in the receiver's welfare, even when the outcome space already contains S^* and thus distinguishes the states as well as possible. However, in this case, the magnitude of the improvement vanishes and is negligible for large N. While merging non-distinguishing outcomes is only sharply optimal in the continuum, the convergence result guarantees us that it remains an actionable recommendation, yielding, in practice, nearly-optimal information to the receiver with minimally cumbersome datasets.

7 Conclusion

Inference under selective disclosure depends on an understanding of how data are generated, and how senders report – or omit – it. This paper underscores the simplicity of a receiver-optimal equilibrium reporting strategy for the sender: claim a possibly inflated state, and provide a large-enough body of evidence that supports it by mimicking the distribution of data it implies. Given their behavior, voluntary disclosure benefits precisely those senders who engage in strategic omission – those with a large amount of data or a low state – and worsens outcomes for those who are imitated – those who have less data, or a more desirable state to imitate.

Even when datasets are large enough to guarantee the sender is fully informed, strategic omission and uncertainty about exactly how much data the sender had mean that only muddled information will reach the receiver. However, in the absence of direct ways to monitor the sender's true data, a planner can nevertheless design the underlying experiment to elicit more informative disclosures. For large-enough datasets, the quality of evidence an experiment provides to the receiver is contingent only on how well its most informative outcomes distinguish one state from another state. In many cases, this sharp fact of big data gives the designer a way to restrict to a lower-dimensional dataset, without loss of efficiency.

This work suggests several compelling directions for future research. One concerns voluntary disclosure with endogenous, costly data acquisition. For example, we would like to know how a sender might acquire data in order to persuade through voluntary disclosures. We conjecture that, if having more data benefits senders strategically regardless of its informational value, then data might be systematically over-collected precisely when it is cheap and plentiful. Another direction would consider the incentive effects of voluntary disclosure for agents: for example, some high-value ideas may see little investment because it is hard to distinguish success in realizing those ideas from success in other objectives based on evidence. Finally, there is a line of work that brings these questions closer to the ground by considering how they manifest in the structured data-generating processes commonly assumed in statistical or econometric models, for instance in logit and probit treatment effects models, or in linear models with Gaussian error. Doing so could help guide practitioners on what strategic omission looks like in those settings, and when it most undermines the value of research.

References

Isaiah Andrews and Maximillian Kasy. Identification of and correction for publication bias. *American Economic Review*, 109(8):2766–2794, 2019.

Elchanan Ben-Porath, Eddie Dekel, and Barton L. Lipman. Mechanisms with evidence: Commitment and robustness. *Econometrica*, 87(2):529–566, 2019.

Sulagna Dasgupta, Ilia Krasikov, and Rohit Lamba. Hard information design. *Working paper*, 2022.

Alfredo Di Tillio, Marco Ottaviani, and Peter N. Sorenson. Strategic sample selection. *Econometrica*, 89(2):911–953, 2021.

Ronald A. Dye. Disclosure of nonproprietary information. *Journal of Accounting Research*, 23(1): 123–145, 1985.

Wioletta Dzuida. Strategic argumentation. *Journal of Economic Theory*, 146(4), 2011.

Mike Felgenhauer and Elisabeth Schulte. Strategic private experimentation. *American Economic Journal: Microeconomics*, 6(4):74–105, 2014.

Michael J. Fishman and Kathleen M. Haggerty. Investment and information acquisition. *Quarterly Journal of Economics*, 105(2):427–444, 1990.

Drew Fudenberg and Jean Tirole. *Game Theory*. MIT Press, 1970.

Jacob Glazer and Ariel Rubinstein. A study in the pragmatics of persuasion: A game theoretical approach. *Theoretical Economics*, (1):395–410, 2006.

Jerry Green and Jean-Jacques Laffont. Partially verifiable information and mechanism design. *Review of Economic Studies*, 53(3):447–456, 1986.

Sanford J. Grossman. The informational role of warranties and private disclosure about product quality. *The Journal of Law and Economics*, 24(3):461, 1981.

Nika Haghtalab, Nicole Immorlica, Brendan Lucier, Markus Mobius, and Divyarthi Mohan. Persuading with anecdotes. *NBER working paper*, 2022.

Sergiu Hart, Ilan Kremer, and Motty Perry. Evidence games: Truth and commitment. *American Economic Review*, 107(3):690–713, 2017.

Emeric Henry and Marco Ottaviani. Research and the approval process: The organization of persuasion. *American Economic Review*, 109(2):911–955, 2019.

Shaofei Jiang. Disclosure games with large evidence spaces. *Working paper*, 2022.

Ginger Zhe Jin, Michael Luca, and Daniel Martin. Is no news (perceived as) bad news? an experimental investigation of information disclosure. *AEJ: Microeconomics*, 13(2):141–173, 2021.

Woon-Oh Jung and Young K. Kwon. Disclosure when the market is unsure of information endowment of managers. *Journal of Accounting Research*, 26(1):123–145, 1988.

Ying Xue Li and Burkhard Schipper. Strategic reasoning in persuasion games: An experiment. *Games and Economic Behavior*, 121:329–367, 2020.

Steven Matthews, Masahiro Okuno-Fujiwara, and Andrew Postlewaite. Refining cheap-talk equilibria. *Journal of Economic Theory*, 55(2):247–273, 1991.

Dimitri Migrow and Sergei Severinov. Investment and information acquisition. *AEJ: Microeconomics*, 14(3):480–529, 2022.

Paul R. Milgrom. Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, 12(2):380–391, 1981.

Elif Osun and Erkut Ozbay. Evidence games: Lying aversion and commitment. *Working paper*, 2021.

Daniel Rappoport. Evidence and skepticism in verifiable disclosure games. *Working paper*, 2022.

Itai Sher. Credibility and determinism in a game of persuasion. *Games and Economic Behavior*, 71(2):409–419, 2011.

Hyun Song Shin. News management and the value of firms. *RAND Journal of Economics*, 21(1): 58–71, 1994.

Hyun Song Shin. Disclosures and asset returns. *Econometrica*, 71(1):105–133, 2003.

Denis Shiskin. Evidence acquisition and voluntary disclosure. *Working paper*, 2022.

Uri Simonsohn, Leif D. Nelson, and Joseph P. Simmons. p-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, 2014.

A Construction of the imitation equilibrium

We will prove that Theorem 2 holds in a more general case with potentially state-contingent, rather than state-independent, data-mass distributions. Describe a game in this general setting by $\mathcal{G}(\Theta, \mathcal{D}, \beta_0, \{f_j\}_{j=1}^J, d\{G^j\}_{j=1}^J)$ where G^j describes the distribution of μ under state j. The model we describe in the main text corresponds to the case in which $G^j = G$ for all j. We also note that a restricted burden of proof function suffices to capture the strategic essence of the equilibrium.

Theorem 7. *Suppose that $g^1, \ldots, g^J$, the densities of μ under states $\theta_1, \ldots, \theta_J$, respectively, are continuous on $\mathbb{R}$ and supported on $[0,1]$. There exists a unique imitation equilibrium outcome, implemented by a (restricted) vector-valued burden of proof function $\tilde{\boldsymbol{\mu}}(u) : [0, \theta_J] \to \mathbb{R}^{J - \max_{\theta_j < u} j}$ with inverse $\tilde{u}_k(\mu)$ such that*

1. *$\tilde{u}_j(\mu)$ is continuous and (weakly) increasing in μ for all j.*
2. *$\sigma^*(\mu f_j)$ is supported on $\{\mu' f_k : \mu' = \tilde{\mu}_k(\tilde{u}_k(\mu / r_j(k)))$ and $\theta_k \in A_j(\mu)\}$.*

To outline the argument, we first prove the existence of a imitation equilibrium by construction. Then we prove the separation theorem, which we use to show uniqueness.

Recall that $\hat{u}_k(\mu)$ is the equilibrium payoff to sending the message μf_k.

We construct $\hat{u}_k(\mu)$ that is monotone increasing in μ – this implies that it must be almost-everywhere differentiable. Since it is also continuous, it is completely determined by its derivative over the points at which the derivative exists. To avoid confusion, we focus on the left derivative of $\hat{u}_k$, which we denote by $\hat{u}_k^-$ and, analogously to the top-down construction of the finite-data equilibrium, we construct the payoff function starting from the top down, starting from the frontier $v = \theta_J$.

Recall that $r_j(k) = \max_{d \in \mathcal{D}} \frac{f_k(d)}{f_j(d)}$ is the ratio of the amount of data necessary to imitate a certain amount of f_k under state j to the amount necessary under state k, and

$$A_j(\mu | \tilde{\boldsymbol{\mu}}) = \left\{ \theta_k : k \in \arg\max_{k > j} \hat{u}_k \left(\frac{\mu}{r_j(k)} \right) \right\}.$$

is the set of states that type μf_j finds it weakly optimal to target given $\hat{\boldsymbol{\mu}}$.

The range of $\hat{u}_k(\mu_k)$ is $[0, \theta_k]$ since no type of higher state ever targets state θ_k, so payoffs to targeting θ_k cannot exceed θ_k itself.

Define

$$S(v) = \{\theta_k : \theta_k > v\}$$

to be the set of states under which the receiver optimally takes an action that yields the sender a payoff greater than v. Then $\hat{\mu}_k(v) < \infty$ iff $\theta_k \in S(v)$, and since play is supported on $\{\hat{\mu}_k(u_k(\mu / r_j(k))) f_k : \theta_k \in A_j(\mu | \tilde{\boldsymbol{\mu}})\}$ and $\sigma(\hat{\mu}_k(u) f_k | \hat{\mu}_k(u) f_k) = 1$, $S(v)$ is exactly the set of states that are targeted by some type under σ to obtain a payoff of v.

Given a partially-specified burden of proof vector $\tilde{\boldsymbol{\mu}}(v) = (\hat{\mu}_k(v))_{\theta_k \in S(v)}$, we can fully reconstruct the full vector, as it is the frontier of all types that are just able to meet some component of $\tilde{\boldsymbol{\mu}}(v)$

with no slack, that is, all types $\hat{\mu}_j f_j$ such that

$$r_j(k)\hat{\mu}_j = \tilde{\mu}_k(v) \text{ for some } \theta_k \in S(v), \text{ and } \nexists \theta_{k'} \in S(v) \text{ s.t. } r_j(k')\hat{\mu}_j > \tilde{\mu}_{k'}(v). \tag{8}$$

Given a particular partial burden of proof function $\tilde{\boldsymbol{\mu}}$, the implied frontier for payoff v is $\hat{\boldsymbol{\mu}}(v|\tilde{\boldsymbol{\mu}}) = (\hat{\mu}_1, \dots, \hat{\mu}_{l-1}, \tilde{\mu}_l(v), \dots, \tilde{\mu}_J(v))$ if $S(v) = \{\theta_l, \dots, \hat{\theta}_J\}$ where $\hat{\mu}_1, \dots, \hat{\mu}_{l-1}$ satisfy eq. 8.

Let the set of states under which some type of sender obtains payoff v and finds it weakly optimal to target state θ_k be

$$\tau^{opt}_{\tilde{\boldsymbol{\mu}}}(\theta_k, v) = \{\theta_j : f_k \in A_j(\hat{\mu}_j(v|\tilde{\boldsymbol{\mu}})|\tilde{\boldsymbol{\mu}})$$

and let the set of states such that some type of sender obtains payoff v by targeting a state θ_k with strictly positive probability under σ be

$$\tau^{supp}_{\tilde{\boldsymbol{\mu}}}(\theta_k, v) = \{\theta_j : \hat{\mu}_k(v) f_k \in \text{supp } \sigma(\cdot|\mu f_j) \text{ for some } \mu\}.$$

Of course, $\tau^{supp}_{\tilde{\boldsymbol{\mu}}}(\theta_k, v) \subseteq \tau^{opt}_{\tilde{\boldsymbol{\mu}}}(\theta_k, v)$.

For convenience of notation, we extend the definitions of these set-valued functions to any set of inputs (rather than a single input) by letting the function of the set be the union of the function applied to each individual element of the input set: thus for every set S of states, $\tau^{opt}_{\tilde{\boldsymbol{\mu}}}(S, v) = \bigcup_{\theta_k \in S} \tau^{opt}_{\tilde{\boldsymbol{\mu}}}(\theta_k, v)$ and $\tau^{supp}_{\tilde{\boldsymbol{\mu}}}(S, v) = \bigcup_{\theta_k \in S} \tau^{supp}_{\tilde{\boldsymbol{\mu}}}(\theta_k, v)$, and for every set $\omega \subseteq [0, 1]$, we let $A_j(\omega) = \bigcup_{\mu \in \omega} A_j(\mu|\tilde{\boldsymbol{\mu}})$.

Additionally, we define the expectation of the state under the (receiver's) belief that the the sender is a type that receives v under $\tilde{\boldsymbol{\mu}}$ and finds it weakly optimal to target a state in S as follows.

$$V_{\tilde{\boldsymbol{\mu}}}(S, \tilde{\boldsymbol{\mu}}(v)) = \frac{\sum_{\theta_j \in \tau^{supp}_{\tilde{\boldsymbol{\mu}}}(S,v)} \beta_0(\theta_j)\theta_j g^j(\hat{\mu}_j(v|\tilde{\boldsymbol{\mu}}))\frac{d\hat{\mu}_j(v|\tilde{\boldsymbol{\mu}})}{dv}}{\sum_{\theta_j \in \tau^{supp}_{\tilde{\boldsymbol{\mu}}}(S,v)} \beta_0(\theta_j) g^j(\hat{\mu}_j(v|\tilde{\boldsymbol{\mu}}))\frac{d\hat{\mu}_j(v|\tilde{\boldsymbol{\mu}})}{dv}}.$$

In contrast, the expectation of the state under the receiver's true belief over θ conditional on knowing that the sender has sent some message that yields payoff v and targets a state in S is

$$W_{\tilde{\boldsymbol{\mu}}}(S, v|\sigma) = \frac{\sum_{\theta_j \in \tau^{opt}_{\tilde{\boldsymbol{\mu}}}(S,v)} \beta_0(\theta_j)\theta_j g^j(\hat{\mu}_j(v|\tilde{\boldsymbol{\mu}}))\frac{d\hat{\mu}_j(v|\tilde{\boldsymbol{\mu}})}{dv}\sigma(\{\tilde{\mu}_k f_k\}_{\theta_k \in S}|\hat{\mu}_j(v|\tilde{\boldsymbol{\mu}})f_j)}{\sum_{\theta_j \in \tau^{opt}_{\tilde{\boldsymbol{\mu}}}(S,v)} \beta_0(\theta_j) g^j(\hat{\mu}_j(v|\tilde{\boldsymbol{\mu}}))\frac{d\hat{\mu}_j(v|\tilde{\boldsymbol{\mu}})}{dv}\sigma(\{\tilde{\mu}_k f_k\}_{\theta_k \in S}|\hat{\mu}_j(v|\tilde{\boldsymbol{\mu}})f_j)} = v. \tag{9}$$

For any partial strategy $\hat{\sigma}$ that gives mixing probabilities between the messages $\tilde{\mu}_k(v)\mathbf{f}_k$, the payoff $W_{\tilde{\boldsymbol{\mu}}}(S, v|\hat{\sigma}(v))$ is always weakly greater than $V_{\tilde{\boldsymbol{\mu}}}(S, v)$. The two are equal exactly when all types obtaining payoff v that find it weakly optimal to target a state in M do so with probability 1.

Fix a frontier $\hat{\mu}(v)$, where $\theta_{l-1} < v \leq \theta_l$. It will be useful to define an undirected graph $H(v)$ on $S(v)$ by adding an edge between θ_k and $\theta_{k'}$ if and only if $\tau^{opt}_{\tilde{\boldsymbol{\mu}}}(\theta_k, v) \bigcap \tau^{opt}_{\tilde{\boldsymbol{\mu}}}(\theta_{k'}, v) \neq \emptyset$, that is, if there is some type that finds it optimal to target either state θ_k or state $\theta_{k'}$, and is indifferent between the two. Let C be the collection of connected components of $H(v)$.

We use the following algorithm to partition $S(v)$ at a given frontier $\tilde{\boldsymbol{\mu}}(v)$.

Algorithm: This algorithm calculates the payoffs to targeting a state in $S(v)$ at frontier $\tilde{\boldsymbol{\mu}}(v)$ when all types that do not obtain higher payoffs than v and who can target some $\tilde{\mu}_k(v)f_k, \theta_k \in S(v)$ target the highest-payoff of these messages among those that they can, and assigns states θ_k to the same partition element if, across them, $\tilde{\mu}_k(v)f_k$ must result in the same payoff, and for α close to 1, $\alpha\tilde{\mu}_k(v)f_k$ must also result in the same payoff, so that for states under which types at the frontier are indifferent between such messages, they remain so for nearby frontiers.

First, note that if σ is such that, when there is a collection of states $\Sigma \subseteq S(v)$ such that, over an interval of payoffs, there always exists between any 2 states in Σ a path of other states in Σ such that there are types that mix with interior probability between any two successive states, then for all $\theta_k, \theta_{k'} \in \Sigma$,

$$\frac{r_j(k)}{r_j(k')} = \frac{\tilde{\mu}_k(u)}{\tilde{\mu}_{k'}(u)} = \frac{\frac{d\tilde{\mu}_k(u)}{du}}{\frac{d\tilde{\mu}_{k'}(u)}{du}} \quad \left(= \frac{\frac{d\hat{u}_{k'}(\tilde{\mu}_{k'}(u))}{d\mu}}{\frac{d\hat{u}_k(\tilde{\mu}_k(u))}{d\mu}} \right) \tag{10}$$

for all u in the interval of payoffs and for all j that target some state in Σ at the frontier $\tilde{\boldsymbol{\mu}}(u)$.

We define

$$\Delta_n(\Sigma, \hat{\alpha}) = \frac{d^n}{d\alpha^n} \frac{\sum_{\theta_j \in \tau^{supp}_{\tilde{\boldsymbol{\mu}}}(\Sigma, v)} \beta_0(\theta_j)\theta_j g\left(\alpha\hat{\mu}_j(v|\tilde{\boldsymbol{\mu}})\right)\hat{\mu}_j(v|\tilde{\boldsymbol{\mu}})}{\sum_{\theta_j \in \tau^{supp}_{\tilde{\boldsymbol{\mu}}}(\Sigma, v)} \beta_0(\theta_j) g\left(\alpha\hat{\mu}_j(v|\tilde{\boldsymbol{\mu}})\right)\hat{\mu}_j(v|\tilde{\boldsymbol{\mu}})} \Bigg|_{\alpha=\hat{\alpha}}.$$

This is equal to the nth derivative of the payoff to the set of senders in states that target a state in Σ with positive probability at frontier $\tilde{\boldsymbol{\mu}}(v)$, that have an amount $\hat{\alpha}\hat{\mu}_j(v|\tilde{\boldsymbol{\mu}})$ of data, when we assume that eq. 10 holds over Σ.

Start with a collection of assigned partition elements, $\mathcal{A}_0 = \emptyset$, and a collection of sets of unassigned states, $\mathcal{C}_0 = C$. Given $\mathcal{A}_n$ and $\mathcal{C}_n$, initialize $\mathcal{A}_{n+1} = \mathcal{C}_{n+1} = \emptyset$, and, taking each set $S \in \mathcal{C}_n$ sequentially, proceed as follows:

1. Take all subsets $\Sigma \subseteq S$ and calculate $\Delta_0(\Sigma, 1)$. Tiebreak any with the same value by $\Delta_1(\Sigma, 1), \Delta_2(\Sigma, 1), \ldots$, successively, and take the largest subset Σ that is maximal. Label it with $\tau^{supp}_{\tilde{\boldsymbol{\mu}}}(\Sigma, v)$, and add it to $\mathcal{A}_{n+1}$.

   Note that this implies that $\frac{du_k(\tilde{\mu}_k(v))}{d\mu_k} = \frac{\Delta_1(\Sigma,1)}{\tilde{\mu}_k(v)}$ when equation 10 holds for $\theta_k, \theta_{k'} \in \Sigma$ over $[v - \epsilon, v]$, $\epsilon > 0$.

2. Take $S \setminus \Sigma$, and let $C(S)$ be the collection of connected components of the graph on S constructed analogously to $H(v)$. Add $C(S)$ to C_{n+1} (i.e. augment C_{n+1} as the union of itself and $C(S)$).

3. Repeat on $\mathcal{A}_{n+1}$ and $\mathcal{C}_{n+1}$ until $\mathcal{C}_{n+1} = \emptyset$.

Putatively, if senders of types $\alpha\hat{\mu}_j(v|\tilde{\boldsymbol{\mu}})f_j$ for some $\theta_j \in \tau_{\tilde{\boldsymbol{\mu}}}^{supp}(v)$ pooled with each other, then payoffs are equal to

$$\hat{u}_k(\alpha\tilde{\mu}_k(v)) = v_\Sigma(\alpha, \tilde{\boldsymbol{\mu}}(v)) \equiv \left. \frac{\sum_{\theta_j \in \tau_{\tilde{\boldsymbol{\mu}}}^{supp}(\Sigma,v)} \beta_0(\theta_j)\theta_j g\left(\alpha\hat{\mu}_j(v|\tilde{\boldsymbol{\mu}})\right)\hat{\mu}_j(v|\tilde{\boldsymbol{\mu}})}{\sum_{\theta_j \in \tau_{\tilde{\boldsymbol{\mu}}}^{supp}(\Sigma,v)} \beta_0(\theta_j) g\left(\alpha\hat{\mu}_j(v|\tilde{\boldsymbol{\mu}})\right)\hat{\mu}_j(v|\tilde{\boldsymbol{\mu}})} \right|_{\alpha=\hat{\alpha}},$$

which is continuous in α because g is continuous everywhere in $(0, 1]$.[21] The burden-of-proof function for $\underline{v} \leq v$ is then given by

$$\mu_\Sigma^{put}(\underline{v}) \equiv \{v_\Sigma^{-1}(\underline{v}, \tilde{\boldsymbol{\mu}}(v))\tilde{\mu}_k(v)f_k\}_{\theta_k \in S(v)},$$

where $v_\Sigma^{-1}(\underline{v}, \tilde{\boldsymbol{\mu}}(v))$ is the inverse of $v_\Sigma(\cdot, \tilde{\boldsymbol{\mu}}(v))$.

The reason that a partition element is a subset of targetable states in which all messages must achieve the same payoff at the is that, since Σ is a maximal highest-value subset over those that do not already have a higher value, it is either partitionable into smaller subsets, each of which also achieves the same value, or not; but in either case, in each minimal subset that achieves the maximal value, there is a path of messages between any two messages in the subset such that, in the targeting strategy, some type mixes with strictly positive probability between any two adjoining messages. The reason for this is that, for any smaller subset $\Sigma' \subset \hat{\Sigma}$, we have that $V_{\tilde{\boldsymbol{\mu}}}(\Sigma', \tilde{\boldsymbol{\mu}}(v)) < V_{\tilde{\boldsymbol{\mu}}}(\hat{\Sigma}, \tilde{\boldsymbol{\mu}}(v))$ if Σ is a minimal subset that achieves the maximal value. Since the expectation of the state conditional on knowing the message played is in $\hat{\Sigma}$ is at least $V_{\tilde{\boldsymbol{\mu}}}(\hat{\Sigma}, \tilde{\boldsymbol{\mu}}(v))$, there must be some message that yields payoff at least $V_{\tilde{\boldsymbol{\mu}}}(\Sigma, \tilde{\boldsymbol{\mu}}(v))$. But since there is no message, and indeed no proper subset of messages in $\hat{\Sigma}$ that achieve payoff $V_{\tilde{\boldsymbol{\mu}}}(\Sigma, \tilde{\boldsymbol{\mu}}(v))$ if all types that can play one of them do, it must be that for any subset, there is a type that can play some message in the subset but plays a message outside the subset with positive probability.

The reason the same holds true in frontiers to the left of $\tilde{\boldsymbol{\mu}}(v)$ is that, if $\Delta_0(\Sigma, 1)$ is uniquely maximal, then $\Delta_0(\Sigma, \alpha)$ is still greater than $\Delta_0(\Sigma', 1)$ for any Σ' and α sufficiently close to 1. So, in any state under which senders target a state in Σ at $\tilde{\boldsymbol{\mu}}(v)$, it remains optimal for them to do so for α close to 1, assuming the putative payoffs above. In addition, the putative payoffs are feasible, because every subset of Σ has lower value. If tiebroken by Δ_1, Δ_2, and so on, then although $\Delta_0(\Sigma, 1)$ is not uniquely maximal, Σ does maximize $\Delta(\cdot, 1)$ immediately to the left of $\tilde{\boldsymbol{\mu}}(v)$.

[21] It is important that $g(1) = 0$, since this ensures that g^j is continuous at $\mu = 1$.

We will use the partition constructed by the algorithm to construct the equilibrium in chunks. For consistency, we want the following condition:

Condition 1. The value of each partition element constructed using the algorithm is the same, and is equal to v.

Under this condition, there is a partial strategy $\hat{\sigma}$ on each partition element such that $W_{\tilde{\boldsymbol{\mu}}}(\theta_k, v|\hat{\sigma}) = v$ for all states θ_k in the partition element, and furthermore, there is no partial strategy on a subset of messages in that partition element such that all messages in the subset result in the same payoff that is greater than v.

If Condition 1 holds at $\tilde{\boldsymbol{\mu}}(v)$ and Σ is the partition constructed using the algorithm at $\tilde{\boldsymbol{\mu}}(v)$, then there exists some $\epsilon > 0$ such that, for all $\underline{v} \in [v-\epsilon, v]$, Condition 1 holds for the frontier $\{v_\Sigma^{-1}(\underline{v}, \tilde{\boldsymbol{\mu}}(v))\tilde{\mu}_k(v) f_k\}_{\theta_k \in S(v)}$. To show this, observe the following claim, which follows directly from statement of the condition and from continuity of $v_\Sigma(\alpha, \tilde{\boldsymbol{\mu}}(v))$:

Claim 10. *Let the set of types that target a state in Σ and achieve a payoff of $\underline{v}$ under μ_Σ^{put} be $\tau_\Sigma^{put}(\underline{v})$.*

If Condition 1 holds at $\tilde{\boldsymbol{\mu}}(v)$, then if there exists no $v' \in (\underline{v}, v]$ such that either

1. *There is a type $t \in \tau_\Sigma^{put}(v')$ such that t can imitate a higher-value state, i.e. there exists partition element such that Σ' $v_{\Sigma'}^{-1}(v'', \tilde{\boldsymbol{\mu}}(v))\tilde{\mu}_k(v) f_k \tilde{\subseteq} t$ for some $v'' > v'$*
2. *There is a partition element Σ with a subset $\Sigma' \subseteq \Sigma$ such that $v_{\Sigma'}(v_\Sigma^{-1}(v', \tilde{\boldsymbol{\mu}}(v)), \tilde{\boldsymbol{\mu}}(v)) > v'$,*

then Condition 1 continues to hold at $\hat{\underline{v}}$.

Note that, because for any partition element $\Sigma' \neq \Sigma$ either $v_{\Sigma'}^{-1}(v, \tilde{\boldsymbol{\mu}}(v))\tilde{\mu}_k(v) f_k \tilde{\not\subseteq} t$, or $\Delta_n(\Sigma', 1) < \Delta_n(\Sigma, 1)$ for some n such that $\Delta_i(\Sigma', 1) = \Delta_i(\Sigma, 1)$ for all $i < n$, the continuity of $v_\Sigma(\alpha, \tilde{\boldsymbol{\mu}}(v))$ implies that for $\underline{v}$ close to v (1) cannot not hold. Again by continuity, (2) cannot hold for $\underline{v}$ close to v because for all $\Sigma' \subseteq \Sigma$, $v_{\Sigma'}(v_\Sigma^{-1}(v, \tilde{\boldsymbol{\mu}}(v)), \tilde{\boldsymbol{\mu}}(v)) \leq v$ and $\Delta_n(\Sigma', 1) < \Delta_n(\Sigma, 1)$ for some n such that $\Delta_i(\Sigma', 1) = \Delta_i(\Sigma, 1)$ for all $i < n$.

We will use this to construct the equilibrium in segments over which Condition 1 holds, and re-construct partitions using the algorithm in at most countably many points at which either (1) or (2) holds. For every reasonable example we can think of, the number of such points (and thus steps in the construction) is not just countable, but finite.

Now we turn to constructing larger pooling sets when there is a positive-measure set of types that can achieve the frontier payoff. Given that types support their play on $\{\tilde{\mu}_k(u_k(\mu/r_j(k)))f_k : \theta_k \in A_j(\mu|\tilde{\boldsymbol{\mu}})\}$, and $\hat{u}_k(\mu_k)$ is increasing, all types capable of sending a message in $\{\hat{\mu}_j(v)\mathbf{f}_j\}_{j=1}^J$ achieve a payoff of at least v. We define the set of types that are incapable of sending a message in $\{\hat{\mu}_j(v)\mathbf{f}_j\}_{j=1}^J$, but capable of sending a message in set M, as $T(v, M)$. We will denote the payoff to the sender of the receiver knowing they are one of a set of types that has positive probability measure under the receiver's prior as $U(T)$, and in particular,

$$U(T(v,M)) = \frac{\sum_{j=1}^{J}\beta_0(\theta_j)\theta_j \max(\max_{k\geq l}(G^j(\frac{\tilde{\mu}_k(v)}{r_j(k)})) - \min\{G^j(\mu) : \exists m \in M \text{ s.t. } m\tilde{\subseteq}\mu f_j\}, 0)}{\sum_{j=1}^{J}\beta_0(\theta_j) \max(\max_{k\geq l}(G^j(\frac{\tilde{\mu}_k(v)}{r_j(k)})) - \min\{G^j(\mu) : \exists m \in M \text{ s.t. } m\tilde{\subseteq}\mu f_j\}, 0)}.$$

Note that $\sup_M U(T(v,M)) \geq v$, because $\lim_{\alpha\to 1} U(T(v,\alpha\tilde{\boldsymbol{\mu}})) = v$. If there is a positive-measure type set $T(v,M)$ that achieves the value $\sup_M U(T(v,M))$, then take the largest such set and call it $\hat{T}^{max}_{\tilde{\boldsymbol{\mu}}}(v)$. Then the following hold:

1. If there exists a set $T(v,M)$ that achieves the value $\sup_M U(T(v,M))$, then there is a unique largest set that does so, and so $\hat{T}^{max}_{\tilde{\boldsymbol{\mu}}}(v)$ is well-defined.
2. Whenever $\hat{T}^{max}_{\tilde{\boldsymbol{\mu}}}(v)$ exists, there exist $\mu_l, \dots, \mu_J$ such that $\hat{T}^{\max}_{\tilde{\boldsymbol{\mu}}}(v) = T(v, \{\mu_l f_l, \dots, \mu_J f_J\})$.
3. Whenever $\hat{T}^{max}_{\tilde{\boldsymbol{\mu}}}(v)$ exists, there exists a partial strategy $\hat{\sigma} : \hat{T}^{\max}_{\tilde{\boldsymbol{\mu}}}(v) \to M = \{\mu_l f_l, \dots, \mu_J f_J\}$ such that the payoff to any message $m \in M$ given that senders in $\hat{T}^{\max}_{\tilde{\boldsymbol{\mu}}}(v)$ play according to $\hat{\sigma}$ is $\hat{U}_{\tilde{\boldsymbol{\mu}}}(v)$.

The first point follows from the fact that, unless the union of two such sets yields payoff at least $\hat{U}_{\tilde{\boldsymbol{\mu}}}(v)$, then their intersection – which corresponds to the pool of types implemented by a different message set – yields strictly greater payoff. To see the 2nd point, simply take μ_k to be the minimum amount of data distributed f_k such that the dataset still contains a message in M, for each $k \geq l$, and note that the resulting set of types is a subset of $T(v,M)$ that has a smaller mass of types θ_j, $j < l$ but the same mass of types θ_k, $k \geq l$. Since $U(T(v,M)) \geq v \geq \theta_{l-1}$, this can only improve the payoff to the pool. The last point comes from the fact that, if $\hat{T}^{\max}_{\tilde{\boldsymbol{\mu}}}(v)$ is a maximum-payoff pool, then for each subset $S \subseteq M$, the payoff to the pool implemented by S is no greater than $U(\hat{T}^{\max}_{\tilde{\boldsymbol{\mu}}}(v))$, which is sufficient to ensure that $\hat{\sigma}$ exists. In addition, $U(T(v,M))$ is absolutely continuous with respect to every component of $\tilde{\boldsymbol{\mu}}(v)$ and each μ_k.

Lemma 8. *If $\hat{T}^{\max}_{\tilde{\boldsymbol{\mu}}}(v)$ exists, then Condition 1 is satisfied by the burden of proof vector $M = \{\mu_l f_l, \dots, \mu_J f_J\}$ such that $\hat{T}^{\max}_{\tilde{\boldsymbol{\mu}}}(v) = T(v,M)$.*

Proof. Suppose not; then one of two cases is true:

1. There is a collection of states $\Sigma \subset S(v)$ such that $V_{\tilde{\boldsymbol{\mu}}}(\Sigma, M) > v$.

   Then, since $V_{\tilde{\boldsymbol{\mu}}}(\Sigma, \alpha(\mu_k f_k)_{k=l}^J)$ is continuous in α, there is $\underline{\alpha} < 1$ such that $V_{\tilde{\boldsymbol{\mu}}}(\Sigma, \alpha(\mu_k f_k)_{k=l}^J) > v$ for all $\alpha \in [\underline{\alpha}, 1]$. Consider an alternative type set, $T(M_{\underline{\alpha},\Sigma}, v)$ where $M_{\underline{\alpha},\Sigma}$ includes the messages $\mu_k f_k$ for $\theta_k \in S(v) \setminus \Sigma$, and the messages $\underline{\alpha}\mu_k f_k$ for $\theta_k \in \Sigma$.

   For $\underline{\alpha}$ small enough, the set of types in $T(M_{\underline{\alpha},\Sigma}, v) \setminus T(M, v)$ includes exactly those in frontiers $(\alpha M)^1_{\alpha=\underline{\alpha}}$ that find it weakly optimal to target a state in Σ. So, the expectation of the state given that the sender's type is in $T(M_{\underline{\alpha},\Sigma}, v) \setminus T(M, v)$ exceeds v, and so $T(M_{\underline{\alpha},\Sigma}, v)$ is higher-payoff than $T(M, v)$, contradicting that $T(M, v) = \hat{T}^{\max}_{\tilde{\boldsymbol{\mu}}}(v)$.

2. There is a element of the partition, $\Sigma' \subset S(v)$, such that $V_{\tilde{\boldsymbol{\mu}}}(\Sigma', M) > v$.

   Then WLOG let Σ' be the lowest-value element of the partition. Similarly to the above, since $V_{\tilde{\boldsymbol{\mu}}}(\Sigma, \alpha(\mu_k f_k)_{k=l}^J)$ is continuous in α, there is $\bar{\alpha} > 1$ such that $V_{\tilde{\boldsymbol{\mu}}}(\Sigma, \alpha(\mu_k f_k)_{k=l}^J) < v$ for all $\alpha \in [1, \bar{\alpha}]$. Consider an alternative type set, $T(M_{\bar{\alpha},\Sigma'}, v)$ where $M_{\bar{\alpha},\Sigma'}$ includes the messages $\mu_k f_k$ for $\theta_k \in S(v) \setminus \Sigma'$, and the messages $\bar{\alpha}\mu_k f_k$ for $\theta_k \in \Sigma'$.

   For $\bar{\alpha}$ small enough, the set of types in $T(M, v) \setminus T(M_{\bar{\alpha}}, \Sigma')$ includes exactly those in frontiers $(\alpha M)_{\alpha=1}^{\bar{\alpha}}$ that find it weakly optimal to target a state in Σ. Then the expectation of the state given that the sender's type is in $T(M, v) \setminus T(M_{\bar{\alpha}}, \Sigma')$ is less than v, so the expectation given that the type is in $T(M_{\bar{\alpha}}, \Sigma')$ exceeds v, contradicting that $T(M, v) = \hat{T}^{\max}_{\tilde{\boldsymbol{\mu}}}(v)$.

Since neither case is possible, M, taken as the payoff frontier corresponding to v, must satisfy Condition 1. □

The iterative algorithm to construct the equilibrium of 2 starts from the highest-potential-payoff senders and creates payoff frontiers that satisfy Condition 1. It proceeds as follows:

1. Start with $l = J$ and $\tilde{\mu}_J(\theta_J) = 1$.
2. For each l, construct frontiers $\tilde{\mu}_k(v)$ as follows:
   (a) Start at $v = \theta_l$ and burden-of-proof vector $\tilde{\boldsymbol{\mu}}(\theta_l)$, as constructed from the previous step. For all $v > \theta_l$, let $\tilde{\boldsymbol{\mu}}(v)$ be as already constructed. Define

   $$\check{\mu}_l(\theta_l) = \max\{\mu : \exists j < l \text{ s.t. } \hat{\mu}_j[\tilde{\boldsymbol{\mu}}(\theta_l)] \geq \mu\},$$

   and rewrite $\tilde{\boldsymbol{\mu}}(\theta_l) = (\check{\mu}_l(\theta_l), \tilde{\mu}_{l+1}(\theta_l), \ldots, \tilde{\mu}_J(\theta_l))$. Proceed as below to rewrite $\tilde{\boldsymbol{\mu}}(v)$ for $v < \theta_l$:

   (b) Fix $S = \{\theta_k\}_{k=l}^J$. Given the frontier $\tilde{\boldsymbol{\mu}}(v)$, check if $\hat{T}^{max}_{\tilde{\boldsymbol{\mu}}}(v)$ exists, and if so, find $M = \{\mu_l f_l, \ldots, \mu_J f_J\}$ that implements $\hat{T}^{max}_{\tilde{\boldsymbol{\mu}}}(v)$ and rewrite $\tilde{\boldsymbol{\mu}}(v) = M$.

   (c) At $\tilde{\boldsymbol{\mu}}(v)$, using the algorithm, partition S into subsets of states, and calculate $v_\Sigma(\alpha, \tilde{\boldsymbol{\mu}}(v))$ for all $\alpha \in [0, 1]$ for each subset. Take the lowest-value frontier, $\tilde{\boldsymbol{\mu}}(v')$, under putative payoffs $v_\Sigma(\alpha, \tilde{\boldsymbol{\mu}}(v))$ such that the conditions of Claim 10 are satisfied and such that $\hat{T}^{max}_{\tilde{\boldsymbol{\mu}}}(v'')$ does not exist for any $v'' \in (v', v]$, and assign strategies according to Algorithm 2 between $\tilde{\boldsymbol{\mu}}(v)$ and the new frontier $\tilde{\boldsymbol{\mu}}(v')$.

   (d) Set $v = v'$ and set $\tilde{\boldsymbol{\mu}}(v')$ as the new frontier, and repeat the above 2 steps until $v' = 0$.
3. Repeat the above steps for each l in descending order until $l = 1$, and fix the resulting $\tilde{\boldsymbol{\mu}}$.

The existence of an imitation equilibrium, and the monotonicity of $\hat{u}_k$, follow directly from this construction. Continuity of $\hat{u}_k$ also follows from this construction. The value of $\hat{u}_k$ is defined on series of closed intervals on each of which it is continuous – $v_\Sigma(\alpha, \mu)$ is continuous in α, and $\hat{u}_k(\mu)$ is constant for $\mu f_k \in T(v, M)$. Together, these cover the domain of u_k, that is, $[0, 1]$, and they overlap only at their endpoints, at which they coincide.

B Properties of σ^*

Here, we prove our main results about the structure of the imitation equilibrium. We begin by showing the separation theorem holds, which we use to show that u_{σ^*} is unique. We then proceed to give proofs of other results on selection (from section 2) and on comparative statics (from 4), many of which rely on it.

B.1 Proof of separation theorem and uniqueness

First, we prove the separation theorem. It has 2 parts, which we will prove as lemmas. We start by proving that upper pools are improving:

Lemma 9. *If M is a collection of messages and $\{\hat{\mu}_j(v) f_j\}_{j=1}^J$ is the frontier of types achieving a payoff of at least v under σ^*, where $\theta_i < v \leq \theta_{i+1}$, then*

$$\mathbb{E}_q[\theta | t \in U(\{\underline{\mu}_j f_j\}_{j=1}^J) \setminus U(M)] \geq v$$

whenever $U(\{\hat{\mu}_j f_j\}_{j=1}^J) \setminus U(M)$ is nonempty.

Proof of Lemma. Denote $T(v, M) = U(\{\hat{\mu}_j f_j\}_{j=1}^J) \setminus U(M)$. Let $(\bar{\mu}_1, \ldots, \bar{\mu}_i; \bar{\mu}_{i+1}, \ldots, \bar{\mu}_J)$ be the minimum masses of data distributed like $f_1, \ldots, f_i; f_{i+1}, \ldots, f_J$, respectively, necessary to send some message in M. Then

$$\mathbb{E}_q[\theta | t \in T(v, M)] = \frac{\sum_{j=1}^J \beta_0(\theta_j) \theta_j (G^j(\bar{\mu}_j) - G^j(\hat{\mu}_j))}{\sum_{j=1}^J \beta_0(\theta_j)(G^j(\bar{\mu}_j) - G^j(\hat{\mu}_j))}.$$

If $(\bar{\mu}_{i+1}, \ldots, \bar{\mu}_J) \leq (\hat{\mu}_{i+1}, \ldots, \hat{\mu}_J)$ pointwise, then $T(v, M)$ is empty. Otherwise, let the states $j_1, \ldots, j_A$ be the maximal set such that $(\bar{\mu}_{j_1}, \ldots, \bar{\mu}_{j_A}) > (\hat{\mu}_{j_1}, \ldots, \hat{\mu}_{j_A})$ pointwise. Call the set of types that send $\mu' f_{j_a}$ with positive probability under σ^* by $\tau_{\sigma^*}^{supp}(\mu' f_{j_a})$, and let $\theta(t)$ refer to the state corresponding to the distribution of dataset t. Denote by $\hat{\sigma}_v$ the partial strategy, restricting to types in $T(v, M)$, where those types play as they do in σ^*, and assume that the receiver knows the sender is in $T(v, M)$ and playing according to this strategy.

Let ϕ_{σ^*} be a joint density over types and messages induced by σ^*, so that for type $t = \mu f_j$ and message $m = \tilde{\mu} f_{\tilde{j}_a}$, we can define

$$\phi(t, m) = g^j(\mu r_j(\tilde{j}_a))\sigma^*(m|t)\beta_0(\theta_j) r_{\theta_j}(j')$$

to be the density on the event that the sender is type t and plays message m, when t plays m with positive probability. In the case when payoffs under u_{σ^*} are strictly increasing at $\mu' f_{j_a}$, each sender who plays $\mu' f_{j_a}$ is randomizing between at most a finite number of messages in their mixed strategy, one corresponding to each state that is weakly optimal for them to imitate. Thus, they play each message in the support of their strategy with strictly positive probability, rather than randomizing with some density over a continuum of messages; ϕ therefore fully captures the distribution of play for senders playing $\mu' f_{j_a}$.

When payoffs are strictly increasing at $\mu' f_{j_a}$, we know that for every $\mu' f_{j_a}$ that is in $T(v, M)$ and is on-path in σ^*, the receiver's inference when they know the sender's type is in $T(v, M)$ in addition to knowing they played message $\mu' f_{j_a}$ is weakly better than if they only know $\mu' f_{j_a}$ was the message played. Formally,

$$\begin{aligned}
\mathbb{E}_q[\theta|\mu' f_{j_a}] &= \frac{\sum_{t \in \tau_{\sigma^*}^{supp}(\mu' f_{j_a}) \bigcap T(v,M)} \theta(t)\phi(t, \mu' f_{j_a}))}{\sum_{t \in \tau_{\sigma^*}^{supp}(\mu' f_{j_a}) \bigcap T(v,M)} \phi(t, \mu' f_{j_a})} \\
&\geq \frac{\sum_{t \in \tau_{\sigma^*}^{supp}(\mu' f_{j_a})} \theta(t)\phi(t, \mu' f_{j_a})}{\sum_{t \in \tau_{\sigma^*}^{supp}(\mu' f_{j_a}))} \phi(t, \mu' f_{j_a})} \\
&\geq v
\end{aligned} \tag{11}$$

where the first inequality comes from the fact that $\theta_{j_a} \geq v > \theta(t)$ whenever $\theta(t) \neq \theta_{j_a}$, and $\mu' f_{j_a} \in T(v, M)$ only if all types that play it under σ^* are also in $T(v, M)$.

Since, of course, payoffs under u_{σ^*} may not be strictly increasing at every $\mu' f_{j_a}$ in $T(v, M)$, we have to separately consider the case in which they are constant, i.e. the case where there are positive-measure pools T of senders achieving the same payoff $v' > v$ under σ^* with $T \bigcap T(v, M)$ nonempty. Then let M' be the set of messages that implements the pool, and

$$\mathbb{E}_{\hat{\sigma}_v}[\theta|m \in M'] = \mathbb{E}_{\hat{\sigma}_v}[\theta|t \in T \bigcap T(v, M)].$$

The value of $T \setminus T(v, M)$ is equal to the value of $T \bigcap U(M)$, which is no more than v' since $T = \hat{T}_{\hat{\boldsymbol{\mu}}}^{max}(v')$, and so it contains no subsets of higher value. Therefore, $\mathbb{E}_{\hat{\sigma}_v}[v(\theta)|t \in T \bigcap T(v, M)] \geq v' \geq v$.

Then, taking the total expectation over both cases, the expectation of θ given that the sender's type is in $T(v, M)$ is a weighted average of $\mathbb{E}_{\hat{\sigma}_v}[\theta|\mu' f_{j_a}]$ over on-path messages $\mu' f_{j_a}$ in $T(v, M)$ in which the payoff is strictly decreasing; and the value over positive-measure sets of equal payoff.

We have shown that each component is no less than v, and so the weighted average is also at least v. □ □

Next we prove that the imitation equilibrium we construct has worsening lower pools. This is relatively simple.

Lemma 10. *If M is a collection of messages and $\{\hat{\mu}_j(v)f_j\}_{j=1}^J$ is the frontier of types achieving a payoff of at least v under σ^*, where $\theta_i < v \le \theta_{i+1}$, then*

$$\mathbb{E}_q[\theta | t \in U(M) \setminus U(\{\hat{\mu}_j f_j\}_{j=1}^J)] < v$$

whenever $U(M) \setminus U(\{\hat{\mu}_j f_j\}_{j=1}^J)$ is nonempty.

Proof. If there was a payoff frontier $\hat{\boldsymbol{\mu}}(v)$ that had a nonempty, weakly improving lower pool lower-bounded by messages M, then there is a frontier $\hat{\boldsymbol{\mu}}(w) \neq M$ for some $w \ge v$ such that

$$u_{pool}(U(M) \setminus U(\hat{\boldsymbol{\mu}}(w))) = w.$$

The construction algorithm rules this out, because if indeed the payoff frontiers above $\hat{\boldsymbol{\mu}}(w)$ are correctly constructed, then it would next set $\hat{\boldsymbol{\mu}}(w) = M$. □

Finally, we show that the constructed equilibrium outcome is the only imitation equilibrium outcome, and thus that the imitation equilibrium outcome is unique.

Proof. Let the constructed equilibrium be σ^*, and let σ be an alternative equilibrium, with a different outcome. We aim to show that σ^* does not have improving upper pools, and therefore cannot be an imitation equilibrium.

To see this, let M represent the frontier of messages that are used to achieve payoff v in σ. Worsening lower pools under σ^* imply that $u_{pool}(U(M) \setminus U(\hat{\boldsymbol{\mu}}(v))) \le v$, implying that M has a worsening upper pool. Since M is a payoff frontier of σ, the alternative equilibrium σ does not have improving upper pools, and is therefore not an imitation equilibrium. □

B.2 Proof of results in Section 2

Next, we formally show that the imitation equilibrium satisfies the 3 selection criteria that we discuss in Section 2: credible inclusive announcement-proofness, truth-leaning, and receiver-optimality.

First we discuss a way in which the imitation equilibrium outcome arises from optimal behavior for the sender. The concept of optimality we use, *inclusive* announcement-proofness, refines PBE by requiring that there is no self-separating set of sender types who could weakly improve their payoffs by announcing a strategy that uses some set of messages differently than they are used in the baseline equilibrium.

Definition B.1. *Given an outcome u_{σ^*}, a set of types T has a* credible inclusive announcement *that they will play a strategy $\hat{\sigma}_M$ supported over message set M for payoff v if*

- $\hat{\sigma}_M : M \times T \to \mathbb{R}$ *is such that* $\sum_{t\in T} \hat{\sigma}_M(m|t) = 1$ *for all* $m \in M$, $\sum_{m\in M} \hat{\sigma}_M(m|t) = 1$ *for all* $t \in T$, *and* $\mathbb{E}_{\beta_{\hat{\sigma}_M}(\cdot|m)}[\theta] = v$ *for all* $m \in M$.
- $T = \{t \in U(M) : u_{\sigma^*}(t)\}$, *and there is some* $t \in T$ *with* $u_\sigma(t) < v$.

A very closely-related notion, that we take the name from, is the idea of a credible announcement, from Matthews et al. (1991). There is, however, a subtle difference, which is that in a credible announcement, $T = \{t \in U(M) : u_{\sigma^*}(t) \leq v\} \bigcup S$ where $S \subseteq \{t \in U(M) : u_{\sigma^*}(t) = v\}$. Thus what we use is an "inclusive" notion of a credible announcement in that the set of announcing types must include all who weakly prefer to participate; it is stronger to claim there exists an credible inclusive announcement than that there exists a credible announcement, and correspondingly, inclusive announcement-proofness is weaker than announcement-proofness. In fact, there may exist no announcement-proof equilibrium at all in the game we study, while there always exists exactly one inclusive announcement-proof equilibrium outcome.

Claim 11. *In $\mathcal{G}$, the unique inclusive announcement-proof equilibrium outcome is the imitation equilibrium outcome.*

Proof. For any equilibrium σ with a different outcome than the imitation-equilibrium outcome σ^*, there is some v such that the v-payoff frontier under σ differs from that under σ^*, and such that some types that achieve a payoff of v or greater under σ^* achieve a payoff no more than v under σ. Lemma 9 ensures that when all such types pool, the expected value of the state is at least v. Then, from the continuity of $\hat{u}_j(\mu)$, there exists some $v' < v$ such that when the set of all types that achieve a payoff of at least v' under σ^*, but a payoff of no more than v under σ, is pooled, the expected value of the state is exactly v. Starting from equilibrium σ, this set of types has a credible inclusive announcement that yields a payoff of v to each type, and so σ is not inclusive announcement-proof.

On the other hand, any credible announcement relative to baseline equilibrium σ^* requires the existence of some v and set of messages M such that there exists a pool of types

$$T = \{t \in U(M) : u_{\sigma^*}(t) \leq v\}$$

such that $\mathbb{E}[\theta|t \in T] = v$, with at least one type $t' \in T$ such that $u_{\sigma^*}(t') < v$. Since T contains all types $t \tilde{\supset} t'$ with $u_{\sigma^*}(t) \leq v$, we know T is a set of positive measure. The construction algorithm for σ^*, however, rules out the presence of any such set T, since if all frontiers for payoffs in $(v, \theta_J]$ are correctly constructed, then all types in T must be pooled under σ^* and must obtain a payoff of v exactly. $\square$

We prove that truth-leaning equilibria and imitation equilibria coincide in $\mathcal{G}$, that the imitation equilibrium outcome is unique, and that it is the optimal outcome of communication under commitment for the receiver.

Claim 12. *Every imitation equilibrium of $\mathcal{G}$ is a truth-leaning equilibrium of $\mathcal{G}$.*

Proof. We take the 2 perturbations separately. First, perturb the likelihood of honest commitment types by a sequence with $\epsilon^k_{t|t} = \epsilon^k \to 0$. There exists an equilibrium $u_{\sigma^*_{\epsilon^k}}$ of $\mathcal{G}^{\epsilon^k}$ in which strategies of non-commitment types are identical to the imitation equilibrium strategies in a game $\tilde{\mathcal{G}}^{\epsilon^k}$ under which

$$q(\mu f_j) = \begin{cases} \frac{\beta_0(\theta_j)(g(\mu)-\epsilon^k)}{1-\epsilon^k \sum_i \beta_i(1-G^i(\hat{\mu}_i(\theta_i))}, & \mu \geq \hat{\mu}_j(\theta_j) \\ \frac{\beta_0(\theta_i)g(\mu)}{1-\epsilon^k \sum_i \beta_i(1-G^i(\hat{\mu}_j(\theta_i))}, & \mu < \hat{\mu}_j(\theta_j). \end{cases}$$

Under the metric induced by the L2 norm, the set of equilibrium strategies is compact, and payoffs in $\tilde{\mathcal{G}}^{\epsilon}$ are continuous in ϵ, so the limit point as $k \to \infty$ of the imitation equilibria of $\tilde{\mathcal{G}}^{\epsilon^k}$ must also be an equilibrium of $\mathcal{G}$. It is easy to verify that it must also satisfy the conditions in 2.1, so it is the imitation equilibrium of $\mathcal{G}$.

Now, for fixed ϵ_k, consider in addition the perturbation of payoffs by an additional payoff bump ν to a truthful report. When $\nu < \min_{j,k} |\theta_k - \theta_j|$, there exists an equilibrium $\sigma^*_{\epsilon^k,\nu}$ that is identical to the equilibrium $u_{\sigma^*_{\epsilon^k}}$ specified above, except for types μf_j with $u_{\sigma^*_{\epsilon^k}} \in (\theta_j, \theta_j + \nu)$, who instead play the truth with positive probability. In particular, for a given message $\mu' f_k$ that yields a payoff in $(\theta_j, \theta_j + \nu)$ and is played by μf_j under $\sigma^*_{\epsilon^k}$, the probability that it is played by μf_j in the equilibrium of the further-perturbed game is 0 if the expected state over types playing $\mu' f_k$ for whom the state is not θ_j is no greater than $\theta_j + \nu$, and otherwise, the probability that μf_j plays μf_k is exactly such that the payoff to playing μf_k is $\theta_j + \nu$, so that μf_j is indifferent between playing message $\mu' f_k$ and revealing all their data. As $\nu \to 0$, the set of affected types shrinks towards a measure-0 set, and so these equilibria converge to $u_{\sigma^*_{\epsilon^k}}$ as $\nu \to 0$.

Finally, given the equilibria $\{\sigma^*_{\epsilon^k,\nu^j}\}$ for $\epsilon^k \to 0$, $\nu^j \to 0$, diagonalize by taking, for every k, some j_k such that $||\sigma^*_{\epsilon^k,\nu^{j_k}} - \sigma^*_{\epsilon^k}|| < \frac{1}{k}$, and observe that then the sequence of perturbations $(\epsilon_{t|t} = \epsilon^k \forall t, \epsilon_t = \nu^{j_k} \forall t)_{k=1}^{\infty}$ yields equilibria that converge to σ^*. $\square$

Claim 13. *Every truth-leaning equilibrium in $\mathcal{G}$ is an imitation equilibrium of $\mathcal{G}$.*

Proof. Let us break down the definition of the imitation equilibrium into 3 parts:

Observation. (σ^*, β^*) is an imitation equilibrium if it is an equilibrium, and under σ^*,

a. Every on-path message is in $\mathcal{T}$,

b. Type $t = \mu f_j$'s dataset is off-path if $\theta_j < \max_{m \subseteq t} \mathbb{E}_{\beta_{\sigma^*}(\cdot|m)}[\theta]$

c. Type $t = \mu f_j$ is always truthful if $\theta_j \geq \max_{m \subseteq t} \mathbb{E}_{\beta_{\sigma^*}(\cdot|m)}[\theta]$.

For part a), note that if a message m is on-path in σ, then there exists K_1 such that for all $k > K_1$, m is on-path in $\sigma^*_{\epsilon_k}$. For every k, however, all on-path messages are in $\mathcal{T}$, since if m is on-path and $m \notin \mathcal{T}$, then there is a type $t = \mu f_j$ with $\theta_j > u_{\sigma^*_{\epsilon_k}}(m)$ that plays m, and t itself is not played as a message on path by any non-commitment types. But then $\mathbb{E}_{\beta_{\sigma^*_{\epsilon^k}}(\cdot|t)}[\theta] = \mathbb{E}_{\pi(\cdot|t)}[\theta] \geq \mathbb{E}_{\beta_{\sigma^*_{\epsilon_k}}(\cdot|m)}[\theta]$, leading to a contradiction. Hence, all on-path m must be in $\mathcal{T}$.

To prove that a truth-leaning equilibrium messaging strategy satisfies c), suppose there is t such that $\mathbb{E}_{\pi(\cdot|t)}[\theta] > \max_{m \tilde{\subseteq} t} \mathbb{E}_{\beta_\sigma(\cdot|m)}[\theta]$ but $\sigma(t|t) < 1$.

. We will show that there is no sequence of perturbations $\{\epsilon^k_t, \epsilon^k_{t|t}\}_{k=1}^\infty \to 0$ such that equilibria of the associated perturbed games $\mathcal{G}^k$ converge to σ. Start by supposing for the sake of contradiction that there is. First, we know t must be on path in σ. If σ^k is an equilibrium of game $\mathcal{G}^k$ with $\epsilon^k_t > 0$, there cannot $t' \neq t$ such that $\sigma^k(t|t') > 0$, otherwise $\mathbb{E}_{\beta_{\sigma^k}(\cdot|t)}[\theta] \geq \max_{t' \tilde{\subset} t} \mathbb{E}_{\beta_{\sigma^k}(\cdot|t')}[\theta]$ and so $\mathbb{E}_{\beta_{\sigma^k}(\cdot|t)}[\theta] + \epsilon^k_t > \max_{t' \tilde{\subset} t} \mathbb{E}_{\beta_{\sigma^k}(\cdot|t')}[\theta]$ and we would have to have $\sigma^k(t|t) = 1$. Then, likewise, in the limit σ, we must have $\sigma(t|t') = 0$ for all t'. Since t is on-path in σ, it must be that $\sigma(t|t) \in (0, 1)$.

Take a type $t'' \neq t$ such that $\sigma(t''|t) > 0$. We know that there exists K such that for all $k > K$, $\sigma^k(t''|t) > 0$ as well. Then whenever $k > K$, $\mathbb{E}_{\pi(\cdot|t)} + \epsilon^k_t = \mathbb{E}_{\beta_{\sigma^k}(\cdot|t'')}[\theta]$. Because $\sigma^k \to \sigma$, we have that

$$\lim_{k\to\infty} \mathbb{E}_{\beta_{\sigma^k}(\cdot|t'')}[\theta] = \mathbb{E}_{\beta_\sigma(\cdot|t'')}[\theta] = \max_{m \tilde{\subseteq} t} \mathbb{E}_{\beta_\sigma(\cdot|m)}[\theta].$$

But this contradicts that $\mathbb{E}_{\pi(\cdot|t)}[\theta] > \max_{m \tilde{\subseteq} t} \mathbb{E}_{\beta_\sigma(\cdot|m)}[\theta]$ and

$$\lim_{k\to\infty} \mathbb{E}_{\beta_{\sigma^k}(\cdot|t'')}[\theta] = \lim_{k\to\infty} \mathbb{E}_{\pi(\cdot|t)} + \epsilon^k_t = \mathbb{E}_{\pi(\cdot|t)}.$$

To show that b) holds, note that for any k, if t is on-path and played by some $t' \neq t$, then $\sigma^k(t|t) = 1$. By c), $\mathbb{E}_{\pi(\cdot|t')} \leq \mathbb{E}_{\beta_{\sigma^k}(\cdot|t)}$, but if t also plays t and $\mathbb{E}_{\pi(\cdot|t)} < \mathbb{E}_{\beta_{\sigma^k}(\cdot|t)}$, then the receiver cannot Bayesian. On the other hand, if t is on-path and only t plays t, then we must have $\mathbb{E}_{\pi(\cdot|t)} = \mathbb{E}_{\beta_{\sigma^k}(\cdot|t)}$. □ □

Finally, closely following the idea in Hart et al. (2017), we show that the imitation equilibrium outcome is the outcome of the optimal deterministic mechanism, that is, the best outcome the receiver can achieve when they can commit to a pure action as a response to the message the sender sends. The revelation principle shows that it suffices to look at direct mechanisms, in which the sender truthfully reports their type and the receiver commits to a deterministic response to the sender's reported type.

A mechanism under which type t elicits the action $a(t)$ is implementable if it satisfies IC:

$$t \tilde{\subseteq} t' \;\Rightarrow\; a(t') \geq a(t). \tag{IC}$$

Claim 14. *The imitation equilibrium outcome is the optimal outcome for the receiver under commitment to deterministic actions.*

To prove this claim, first define $T_{\mu f_k}$ be the set of types that imitate μf_k under σ^*, including μf_k itself. We start with a lemma.

Lemma 11. *There always exists an imitation equilibrium σ^* such that $T_{\mu f_k}$ is finite for every $\mu f_k \in \mathcal{T}$.*

Proof of Lemma 11. First, for any imitation equilibrium, if $\{t : u_{\sigma^*}(t) = u_{\sigma^*}(\mu f_k)\}$ is a measure-0 set, since then it is necessarily true that at most one type under each state lies in the same payoff frontier as μf_k under σ^*, and thus at most one type under each state imitates it.

Now consider the case in which there is a positive-measure set of senders who achieve the payoff $u^* = u_{\sigma^*}(\mu f_k)$, where we have $\theta_l \leq u_{\sigma^*}(\mu f_k) < \theta_{l+1}$. We know that there exists a way to divide the types by which state they imitate, and with what probability, given by sets $S_{l+1}, \ldots, S_J$ and any imitation equilibrium σ^*, such that

$$\frac{\sum_{t \in S_j} \theta(t) q(t) \int_{\hat{\mu}_j(u^*)}^{\inf_{v>u^*} \hat{\mu}_j(v)} \sigma^*(\mu f_j | t) d\mu}{\sum_{t \in S_j} q(t) \int_{\hat{\mu}_j(u^*)}^{\inf_{v>u^*} \hat{\mu}_j(v)} \sigma^*(\mu f_j | t) d\mu} = u^*$$

and for all $\mu^* \in (\hat{\mu}_j(u^*), \inf_{v>u^*} \hat{\mu}_j(v))$,

$$\frac{\sum_{t \in S_j : t \tilde{\supseteq} \mu^*} \theta(t) q(t) \int_{\hat{\mu}_j(u^*)}^{\inf_{v>u^*} \hat{\mu}_j(v)} \sigma^*(\mu f_j | t) d\mu}{\sum_{t \in S_j : t \tilde{\supset} \mu^*} q(t) \int_{\hat{\mu}_j(u^*)}^{\inf_{v>u^*} \hat{\mu}_j(v)} \sigma^*(\mu f_j | t) d\mu} \leq u^*.$$

But it is always feasible to reorder the imitation strategy to construct σ^{**} such that $S_{l+1}, \ldots, S_J$ are unchanged, but if $\mu_1 f_j$ imitates $\mu_1' f_i$ and $\mu_2 f_j$ imitates $\mu_2' f_i$, with $\mu_1 > \mu_2$, then $\mu_1' > \mu_2'$ also. That is, conditional on imitating the same state, higher-data senders always imitate types with more data under σ^{**}. Then any type is imitated by either a single type or an interval of types under any other state; the latter is ruled out by the fact that it would result in a payoff no more than θ_l to the message. Once again, since there is a finite set of states, this ensures that each type is imitated by at most a finite set of other types. □

Proof of Claim 14. Suppose A to be the subset of types in $\mathcal{T}$ that are imitated under the imitation equilibrium σ^*, and suppose that σ^* is an imitation equilibrium in which each type is imitated by a finite set of other types, which exists by the previous lemma. Given $\mu f_j \in A$, let $T_{\mu f_j}$ be the set of types that play μf_j under σ^*, including μf_j itself. Define a distribution over $T_{\mu f_j}$,

$$q_{\mu f_j}(t) = \frac{q(t)\sigma(\mu f_j|t)}{\sum_{t \in T_{\mu f_j}} q(t)\sigma(\mu f_j|t)},$$

which is the probability of type t conditional on the message μf_j.

Call the optimal direct mechanism a^*, that responds with the action $a^*(t)$ after receiving the report t. It must satisfy IC across any subset of types, $T \subseteq \mathcal{T}$, but let us consider instead w, the solution to a relaxed local problem where we impose that IC must hold only between $t, t' \in T_{\mu f_j}$ when types are distributed according to $q_{\mu f_j}$. We will show that for all $t \in T_{\mu f_j}$, we have $w(t) = \mathbb{E}_{q_{\mu f_j}}[\theta]$, and that taking this solution across all $\mu f_j \in A$ assigns a response for the receiver to all $t \in T$ while preserving global IC, and therefore gives the optimal direct mechanism.

We know that $w(\mu f_j) \leq w(t)$ for all $t \in T_{\mu f_j}$. Let $S_{\mu f_j} = \{t \in T_{\mu f_j} : w(t) = w(\mu f_j)\}$. First, note that if $S_{\mu f_j} = T_{\mu f_j}$, then we optimally have $w(t) = \mathbb{E}_{q_{\mu f_j}}[\theta]$ for all $t \in T$. This leaves us to rule out that $w(t) \neq w(t')$ for some $t, t' \in T_{\mu f_j}$.

We rule out that $w(\mu f_j) \geq \mathbb{E}_{q_{\mu f_j}}[\theta]$ and $w(t) \neq w(\mu f_j)$ for some $t \in T_{\mu f_j}$, due to the fact that the receiver can then improve their payoff while preserving IC by instead responding to every type with $w(\mu f_j)$. Next, we rule out that $w(\mu f_j) < \mathbb{E}_{q_{\mu f_j}}[\theta]$ and $w(t) \neq v(\mu f_j)$ for some $t \in T_{\mu f_j}$, since then it is possible to instead respond to every t such that $w(t) = w(\mu f_j)$ with $\min_{t \in T_{\mu f_j} \setminus S} w(t)$, and, by single-peakedness of the receiver's payoff function, this improves the receiver's payoff.

This suffices to show that w corresponds exactly to the outcome of the imitation equilibrium for all $t \in T_{\mu f_j}$, regardless of the choice of $\mu f_j \in A$. As w optimizes the receiver's payoff under a weaker set of IC constraints than a^*, we know that the imitation equilibrium outcome is at least as good as a^* for the receiver; the reverse statement is immediate since every equilibrium outcome is implementable with commitment, and so the two are identical. □ □

Corollary (to Claim 14)**.** *The imitation equilibrium outcome is the receiver-optimal equilibrium outcome.*

Proof. In every equilibrium σ, the receiver has a unique best response to each message, given by the action

$$a_r(\beta(\cdot|m))) = \mathbb{E}_{\beta(\cdot|m)}[\theta].$$

Any type of the sender therefore has an optimal feasible message to send that results in a unique optimal action that they can induce the receiver to take given the receiver's inference function. Any equilibrium outcome can therefore be implemented by the receiver through a direct mechanism that responds to every type with a deterministic message, and so there is no equilibrium that increases the receiver's payoff relative to the optimal outcome of a deterministic mechanism that is equivalent to the imitation equilibrium outcome. □ □

B.3 Proof of results in section 4

B.3.1 Convergence to full-information outcome as $Var(g) \to 0$

Proof of Claim 2. We show that given any infinite sequence of games with data-mass distributions $g_1, g_2, \ldots$ on $[0,1]$ with a fixed mean and variances $Var_1, Var_2, \ldots \to 0$, that are identical in the set of states and their ex-ante distribution, the payoff to a sender conditional on the state converges in probability to their full-information payoff.

In order to do so, we show that for any δ and ϵ, there exists L such that for all $l \geq L$, the distribution g_l is such that $Pr[u_{\sigma^*}(\mu, \theta_k) < \theta_k - \delta] < \epsilon$ under every state.

Define the mean of μ to be $\bar{\mu}$, and

$$B = \max_{j \neq k} \frac{1}{r_j(k)}$$

so that for any two states j and k, the difference between the amount of the state-k distribution that the mean type under state k has and the amount the mean type under state j has is $\bar{\mu}(1-B)$.

Suppose that the variance of μ under density g_L is less than $\Delta^2\epsilon^2$, where $\Delta > 0$ is an arbitrary parameter. Then there can be at most a probability ϵ^2 that the state is k and the sender has less than $\bar{\mu} - \Delta$ data distributed like f_k. A sender under state j has more than $\frac{\bar{\mu}-\Delta}{B}$ data with probability no more than $\frac{\Delta^2\epsilon^2B^2}{(\bar{\mu}(1-B)-\Delta)^2}$.

Recall that whenever $u_{\sigma^*}(\mu, \theta) < \theta$, the type with dataset μf_θ is truthful in equilibrium. So, if under state θ_k we have $Pr[u_{\sigma^*}(\mu, \theta) < \theta - \delta] \geq \epsilon$, then the type with $\mu = G^{-1}(\epsilon)$ must obtain payoff less than $\theta - \delta$, and so must all types with less data, and all such types must be truthful. But the total mass of all types *not* in state k that can pool with types with $\mu \in [G^{-1}(\epsilon^2), G^{-1}(\epsilon)]$ cannot exceed

$$(J-1)(1-\beta_0(\theta_k))\frac{\Delta^2\epsilon^2B^2}{(\bar{\mu}(1-B)-\Delta)^2}$$

and so the payoff to type $G^{-1}(\epsilon)f_k$ cannot be less than

$$\frac{\epsilon(1-\epsilon)\theta_k}{\epsilon(1-\epsilon)+(J-1)(1-\beta_0(\theta_k))\frac{\Delta^2\epsilon^2B^2}{(\bar{\mu}(1-B)-\Delta)^2}}$$

which, for small enough Δ, must be at least $\theta_k - \delta$. Since there is always L large enough that

$Var_L < \epsilon^2\Delta^2$, we are done.

All that remains is to note that, since the ex-ante expected payoff must always be $\mathbb{E}_{\beta_0}[\theta]$, this lower bound on the probability of payoffs less than the full-information payoffs implies a corresponding upper bound on payoffs exceeding the full-information payoffs, and so we obtain convergence of the distribution of payoffs, state-by-state, to those in the outcome where the receiver knows the truth. □

Next we show that the receiver is at least weakly better off knowing more about the state of the world, in a very general sense.

Proof of Claim 3. We can denote the distinguishability factors by

$$r_j(k') = \max_d \frac{f'_{k'}(d)}{f_j(d)}$$

for $\theta_j \in \Theta, \theta_{k'} \in \Theta'$ and

$$r'_{j'}(k) = \max_d \frac{f_k(d)}{f'_{j'}(d)}$$

for $\theta_{j'} \in \Theta', \theta_k \in \Theta$.

The imitation equilibrium outcome in $\mathcal{G}^{uc}$ is the outcome of the optimal determministic mechanism given the same setup. In other words, it is the welfare-maximizing outcome for the receiver that respects the IC constraints,

$$v^{uc}(\mu_1 f_j) \geq v^{uc}(\mu_2 f_j) \;\forall j, \mu_1 > \mu_2. \tag{IC-mon}$$

$$v^{uc}(\mu f_j) \geq v^{uc}(\frac{\mu}{r_j(k)} f_k) \;\forall \mu \in [0,1] \text{ and } \theta_j, \theta_k \in \Theta, \tag{IC-im, Θ}$$

$$v^{uc}(\mu f'_{j'}) \geq v^{uc}(\frac{\mu}{r'_{j'}(k')} f'_{k'}) \;\forall \mu \in [0,1] \text{ and } \theta'_{j'}, \theta'_{k'} \in \Theta', \tag{IC-im, Θ'}$$

$$v^{uc}(\mu f_j) \geq v^{uc}(\frac{\mu}{r_j(k')} f'_{k'}) \;\forall \mu \in [0,1] \text{ and } \theta_j \in \Theta, \theta'_{k'} \in \Theta'; \tag{IC-im, Θ-to-Θ'}$$

$$v^{uc}(\mu f'_{j'}) \geq v^{uc}(\frac{\mu}{r'_{j'}(k)} f_k) \;\forall \mu \in [0,1] \text{ and } \theta'_{j'} \in \Theta', \theta_k \in \Theta. \tag{IC-im, Θ'-to-Θ}$$

Now consider a case when the receiver knows the true state of the world is in Θ vs. a case in which they do not know this (but it is true). In the former case, their welfare is given by the optimal deterministic mechanism subject only to IC-mon and IC-im, Θ. In the latter case, however, the latter 3 incentive compatibility conditions also apply (note that IC-im, Θ' may affect the optimal assignment of values to types in Θ indirectly), and furthermore, the equilibrium gives the same outcome as a mechanism that also optimizes the assignment of values to types in Θ' – neither of these can improve the receiver's outcome conditional on $\theta \in \Theta$, but they can worsen it (i.e. if cross-IC constraints bind). □

B.3.2 Comparative statics of welfare

First, we give a proof of the comparative statics of welfare with respect to $\beta_0(\theta_j)$.

Proof of Claim 5. First, let $\hat{\mu}(v)$ be the frontier of types that attain payoff v under $\mathcal{G}$ and let $\hat{\mu}'(v)$ be the frontier of types that do so under $\mathcal{G}'$. Let q be the distribution of types in $\mathcal{G}$ and q' be the type distribution for $\mathcal{G}'$.

Let $v \geq \theta_j$. Suppose for the sake of contradiction that $U(\hat{\mu}'(v)) \setminus U(\hat{\mu})$ is nonempty. By Lemma 9, in the game $\mathcal{G}'$,

$$\mathbb{E}_{q'}[\theta | t \in U(\hat{\mu}'(v)) \setminus U(\hat{\mu}(v))] \geq v.$$

But we also have $\mathbb{E}_q[\theta | t \in U(\hat{\mu}'(v)) \setminus U(\hat{\mu}(v))] \geq \mathbb{E}_{q'}[\theta | t \in U(\hat{\mu}'(v)) \setminus U(\hat{\mu})]$. The separation theorem then cannot hold for $\hat{\mu}(v)$ and $\hat{\mu}'(v)$ simultaneously.

Similarly, let $v \leq \theta_j$. As with the above, we observe that if $U(\hat{\mu}(v)) \setminus U(\hat{\mu}'(v))$ is nonempty, then

$$\mathbb{E}_q[\theta | t \in U(\hat{\mu}(v)) \setminus U(\hat{\mu}'(v))] \geq v,$$

but since $\mathbb{E}_{q'}[\theta | t \in U(\hat{\mu}(v)) \setminus U(\hat{\mu}'(v)] \geq \mathbb{E}_q[\theta | t \in U(\hat{\mu}(v)) \setminus U(\hat{\mu}'(v))]$, it is likewise impossible for both to satisfy the separation theorem. □

Finally, we give the proof of the effect of an MLRP shift in β_0 or g.

Proof of Claim 6. We treat the two shifts pertaining to β_0 and g separately. Both follow similarly to the above claim from the separation theorem.

We start by showing the first part, that an MLRP upward shift of the state distribution weakly increases payoffs for all types. Suppose that $\mathcal{G}$ and $\mathcal{G}'$ satisfy the assumptions of the claim but only differ in β_0, not g. Let $\hat{\mu}(v)$ be the frontier of types that attain payoff v under $\mathcal{G}$ and let $\hat{\mu}'(v)$ be the frontier of types that do so under $\mathcal{G}'$. If $U(\hat{\mu}(v)) \setminus U(\hat{\mu}'(v))$ is nonempty, then

$$\begin{aligned} \mathbb{E}_{q'}[U(\hat{\mu}(v)) \setminus U(\hat{\mu}'(v))] &= \frac{\sum_j \beta_0'(j)(\hat{\mu}_j'(v) - \hat{\mu}_j(v))\theta_j}{\sum_j \beta_0(j)(\hat{\mu}_j'(v) - \hat{\mu}_j(v))} \\ &> \frac{\sum_j \beta_0(j)(\hat{\mu}_j'(v) - \hat{\mu}_j(v))\theta_j}{\sum_j \beta_0(j)(\hat{\mu}_j'(v) - \hat{\mu}_j(v))} \\ = \mathbb{E}_q[U(\hat{\mu}(v)) \setminus U(\hat{\mu}'(v))] \end{aligned} \tag{12}$$

because, among this same set of datasets, q' places a greater relative likelihood on higher states. But then, once again, the separation theorem cannot simultaneously hold for both frontiers.

To show the second part, that an MLRP downward shift of the receiver's perception of the distribution of μ weakly improves payoffs for all types, assume instead that $\mathcal{G}$ and $\mathcal{G}'$ satisfy the assumptions of the claim but only differ in g, with $g \geq_{MLRP} g'$, and that play follows imitation equilibria σ^* and $\sigma^{*\prime}$ in each, respectively. We know that for every message μf_k sent on-path in $\mathcal{G}$, there is a single high-value type that plays it, μf_k itself, and then a number of worse types that imitate it, all with an original amount of data greater than μ. The MLRP shift implies that the relative probability of every one of the imitators, compared to μf_k itself, is decreased in $\mathcal{G}'$ relative to $\mathcal{G}$. Therefore, if the set of types $T = U(\hat{\mu}(v)) \setminus U(\hat{\mu}'(v))$ is nonempty, then

$$\begin{aligned} \mathbb{E}_{q'}[T] &= \frac{\sum_{m \in T: m \text{ on path in } \mathcal{T}} \mathbb{E}_{q'}[\theta | m, \sigma^*] Pr_{q'}(\sigma^*(t) = m | t \in T)}{q'(T)} \\ &> \frac{\sum_{m \in T: m \text{ on path in } \mathcal{T}} \mathbb{E}_{q}[\theta | m, \sigma^*] Pr_{q}(\sigma^*(t) = m | t \in T))}{q(T)} \\ &= \mathbb{E}_q[U(\hat{\mu}(v)) \setminus U(\hat{\mu}'(v))], \end{aligned} \tag{13}$$

which is once again incompatible with the separation theorem. □

C Results on experimental design

Here, we give proofs of the results in our application to experimental design.

First, we prove the sufficiency theorem, which is almost immediate from the structure of imitation equilibrium discussed in Section 3.

Proof of Prop. 1. Consider an abstract game with type space $\mathcal{T}$ and type distribution q in which we restrict senders to imitating other types, that is, playing messages in $\mathcal{T}$, and only either reporting their own type or those of types for which the state is higher. The type given by state, data-mass pair (θ_j, μ), can imitate the type given by another state, data-mass pair (θ_k, μ') if $k > j$ and

$$\frac{\mu}{r_j(k)} \geq \mu'.$$

We can also identify the imitation equilibrium in this game, which is simply any equilibrium of this game that satisfies Def. 2.1b, and is equivalent to the imitation equilibrium of the unrestricted game. Note that $\{f_j\}_{j=1}^J$ are not directly involved in the set of types that are feasible to imitate, once $\{r_j(k)\}$ are accounted for. We have shown that our game can be abstractly described with only $\{r_j(k)\}_{j<k}$ in the place of $\{f_j\}_{j=1}^J$ while leaving the imitation equilibrium outcome uniquely determined, and so any perturbation of the latter that leaves the former unchanged does not affect

the imitation equilibrium outcome. □

Next, we prove that increasing distinguishability improves the receiver's welfare.

Proposition 12. *Suppose that two games $\mathcal{G}$ and $\mathcal{G}'$ are identical except for their space of outcomes $\mathcal{D}$ and $\mathcal{D}'$ and the generating distributions of data under each state, $\{f_j\}_{j=1}^J$ and $\{f_j'\}_{j=1}^J$, and let σ^* and $\sigma^{*'}$ be their respective imitation equilibria.*

If the $r_j(k) \geq r_j'(k)$ for all j, k, then the receiver's payoff is greater under σ^* than under $\sigma^{*'}$.

Proof of Prop. 4. Let $\mathcal{G}$ and $\mathcal{G}'$ be the implied games, and σ^* and $\sigma^{*'}$ be the imitation equilibria, under experiments $\mathcal{E}$ and $\mathcal{E}'$, respectively.

Under game $\mathcal{G}$, there exists a (pure-strategy) mechanism that implements the outcome of $\sigma^{*'}$. To see this, note that the outcome of $\sigma^{*'}$ is also the outcome of v', the optimal mechanism for the receiver under $\mathcal{G}'$, which respects the IC constraints that can be rewritten as imitation constraints

$$v'(\mu f_j) \geq v'(\frac{\mu}{r_j'(k)} f_k) \;\forall \mu \qquad \text{(IC-im-}j,k,\,\mathcal{G}')$$

for each $j < k$, and monotonicity constraint

$$v'(\mu_1 f_j) \geq v'(\mu_2 f_j) \;\forall j, \mu_1 > \mu_2. \qquad \text{(IC-mon, }\mathcal{G}')$$

On the other hand, in order to be implementable in $\mathcal{G}$, v' need only respect the IC constraints

$$v'(\mu f_j) \geq v'(\frac{\mu}{r_j(k)} f_k) \;\forall \mu \qquad \text{(IC-im-}j,k,\,\mathcal{G})$$

for each $j < k$, as well as the same monotonicity constraint. Each of these imitation constraints is weaker, and so v' remains possible to implement.

Since v' is implementable in $\mathcal{G}$, the outcome of the optimal mechanism, and therefore the imitation equilibrium, in $\mathcal{G}$ gives at least a weak improvement over v' for the receiver.

To show the second part, observe that whenever any type under state j imitates some μf_k in $\mathcal{G}'$, then IC-im-j, k binds in $\mathcal{G}'$. Then relaxing the constraint yields a strict improvement in the outcome of the optimal mechanism, and so the receiver-optimal equilibrium outcome in $\mathcal{G}$ is strictly better than that in $\mathcal{G}'$. □

Along with the improvement theorem, Section 5 establishes some simple benchmarks that in the corners as $r_j(k) \to \infty$ or $r_j(k) \to 1$ for all $j < k$, the outcome approximates the full information and no information outcomes, respectively.

Proof of Claim 7. The first part is quite simple to show: simply note that if $\mu > \frac{1}{r_j(k)}$ for all $j < k$, then the type μf_k is not imitable and therefore is able to at least separate and obtain a payoff of θ_k. Therefore, there are at most $G(\frac{1}{\underline{R}})$ types that obtain a payoff less than their full-information payoff θ_j. Bayes plausibility – the fact that $\mathbb{E}_{\mu,j}[\hat{u}_j(\mu)] = \mathbb{E}_{\beta_0}[\theta]$ then implies that at most a fraction $\frac{[\theta_J - \theta_1]}{\underline{R}\epsilon}$ of types obtain a payoff greater than $\theta_k + \epsilon$. Therefore, any $\underline{R}$ with

$$\delta > G(\frac{1}{\underline{R}}) + \frac{[\theta_J - \theta_1]}{\underline{R}\epsilon}$$

suffices as a uniform lower bound on $r_j(k)$ to ensure that no more than a fraction δ of all types obtain an outcome that differs from their full-information one by more than ϵ.

Next, let us consider when $r_j(k)$ is close to 1. Let $\bar{R} = 1 + \eta$ where η is positive, and let $\bar{b} = \max_\mu g(\mu)$.

Suppose a given type $t = \mu f_j$ obtains payoff v. We have that the average value of types that lie below $\hat{\mu}(v)$, i.e. in $\mathcal{T} \setminus U(\hat{\mu(\mathbf{v})})$, is lower-bounded by

$$\begin{aligned}
\frac{\left[\sum_{j'<v} G((1+\eta)\mu)\theta_{j'}\right] + \left[\sum_{k>v} G(\frac{1}{1+\eta}\mu)\theta_k\right]}{\left[\sum_{j'<v} G((1+\eta)\mu)\right] + \left[\sum_{k>v} G(\frac{1}{1+\eta}\mu)\right]} &\geq \frac{\left[\sum_{j'<v} G((1+\eta)\mu)\theta_{j'}\right] + \left[\sum_{k>v} G((1-\eta)\mu)\theta_k\right]}{\left[\sum_{j'<v} G((1+\eta)\mu)\right] + \left[\sum_{k>v} G((1-\eta)\mu)\right]} \\
&\geq \mathbb{E}_{\beta_0}[\theta] - \frac{b\mu\eta\mathbb{E}_{\beta_0}[\theta] + b(1+\eta)[\theta_J - \theta_1]}{G(\mu)} \\
&\equiv LB(\eta, \mu).
\end{aligned} \tag{14}$$

For the above lower bound, we make use of the fact that $\frac{1}{(1+x)} > (1-x)$.

An upper bound for the average value of types that lie above $\hat{\mu}(v)$, i.e. in $U(\hat{\mu(\mathbf{v})})$, is

$$\begin{aligned}
&\frac{\left[\sum_{j'<v}(1 - G(\frac{1}{1+\eta}\mu))\theta_{j'}\right] + \left[\sum_{k>v}(1 - G((1+\eta)\mu))\theta_k\right]}{\left[\sum_{j'<v}(1 - G(\frac{1}{1+\eta}\mu))\right] + \left[\sum_{k>v}(1 - G((1+\eta)\mu))\right]} \\
&\leq \frac{\left[\sum_{j'<v}(1 - G((1-\eta)\mu))\theta_{j'}\right] + \left[\sum_{k>v}(1 - G((1+\eta)\mu))\theta_k\right]}{\left[\sum_{j'<v}(1 - G((1-\eta)\mu))\right] + \left[\sum_{k>v}(1 - G((1+\eta)\mu))\right]} \\
&\leq \mathbb{E}_{\beta_0}[\theta] + \frac{b\mu\eta\mathbb{E}_{\beta_0}[\theta] + b(1-\eta)[\theta_J - \theta_1]}{1 - G(\mu)} \\
&\equiv UB(\eta, \mu).
\end{aligned} \tag{15}$$

The separation theorem tells us that $LB(\eta, \mu) \leq v \leq UB(\eta, \mu)$. Then for any $\delta > 0$ and $\epsilon > 0$, there is η small enough that for all $\mu \in [G^{-1}(\frac{\delta}{2}), G^{-1}(1 - \frac{\delta}{2})]$, we have $LB(\mu, \eta) > \mathbb{E}_{\beta_0}[\theta] - \epsilon$ and

$UB(\mu, \eta) < \mathbb{E}_{\beta_0}[\theta] + \epsilon$. Then whenever $r_j(k) < 1 + \eta$ for all $j < k$, we have $|\hat{u}_j(\mu) - \mathbb{E}_{\beta_0}[\theta]| \geq \epsilon$ with probability no more than δ. □

Finally, we prove that S^* is a minimal set of outcomes to robustly support optimal communication when β_0 is unknown.

Proof of 9. The assumption that every maximizer of $\frac{f_k(d)}{f_j(d)}$ is unique allows us to assert that merging any $d \in S^*$ with any other outcome strictly decreases $r_j(k)$ for some $j < k$. It is generically satisfied, with respect to uniformly chosen, independent $f_j \in \Delta\mathcal{D}$.

From the improvement proposition, we know that as long as some μf_j imitates some $\mu' f_k$ under experiment $\mathcal{E}^*$, then such a strict decrease in $r_j(k)$ strictly worsens the receiver's payoff. Indeed, we can guarantee that this happens when $\beta_0(\theta_k)$ is very large and $\beta_0(\theta_{j'})$ for $j' \neq j, k$ is very small. To see this, normalize $\theta_1 = 0$, and consider the case in which

$$\beta_0(\theta_k) = 1 - \epsilon, \quad \beta_0(\theta_j) = (1 - \eta)\epsilon, \quad \sum_{j' \neq j,k} \beta_0(\theta_j') = \eta\epsilon.$$

Let $\bar{b}$ be an upper bound on $g(\mu)$ over $\mu \in [0, 1]$, and let $\underline{b}(\xi)$ be a lower bound on $g(\mu)$ over $\mu \in [\frac{1}{r_j(k)}(1 - \xi), \max_{j' \neq k} \frac{1}{r_{j'}(k)}]$. The type with $\mu = 1 - \xi$ and dataset f_j would be able to obtain a payoff of at least

$$B_1(\epsilon, \xi) = \frac{(1 - \epsilon)\underline{b}\theta_k}{(1 - \epsilon)\underline{b}(\xi) + \epsilon\bar{b} \max_{j'} r_{j'}(k)}$$

by imitating $\frac{1}{r_j(k)}$, as this is a lower bound on the receiver's inference after observing $\frac{1}{r_j(k)}(1 - \xi) f_k$ if types under all other states only consider imitating f_k. We have

$$B_1(\epsilon, \xi) \to \theta_k$$

as $\epsilon \to 0$ for any ξ, and in particular, for small-enough ϵ, we have $B_1(\epsilon, \xi) > \frac{\theta_j + \theta_k}{2}$.

In addition, we have a bound on the payoff that type $(1 - \xi) f_j$ could obtain from imitating any combination of other $f_{j'}$ but not f_k, given by

$$B_2(\eta, \xi) = \frac{\eta\bar{b} \max j' \neq k(\theta_{j'}) + \min_{x \leq (1-\xi)} \left[\frac{G(1-\xi) - G(x)}{1 - \xi - x}\right] \theta_j}{\eta\bar{b} + \min_{x \leq (1-\xi)} \left[\frac{G(1-\xi) - G(x)}{1 - \xi - x}\right]}$$

for which we have

$$B_2(\eta, \xi) \to \theta_j$$

as $\eta \to 0$ for any ξ, and in particular, for small-enough η, $B_2(\eta, \xi) < \frac{\theta_j + \theta_k}{2}$.

We have shown that for small enough η and ϵ, the imitation equilibrium must involve $(1-\xi)f_j$ imitating f_k with positive probability. Therefore, a strict decrease in $r_j(k)$ strictly worsens the receiver's welfare under the associated β_0. □

D Properties of truth-leaning equilibria with finite data

First, we establish a few known facts. From Hart et al. (2017), we know that the truth-leaning equilibrium is equivalent (i.e. in outcome) to one that satisfies the properties of imitation. We also know that it is receiver-optimal with and without commitment to pure actions. It remains to prove that it gives the unique inclusive announcement-proof equilibrium outcome, and that it satisfies the separation theorem.

Claim 15. *In any finite-data game $\mathcal{G}_N$, the unique inclusive announcement-proof equilibrium outcome is the truth-leaning equilibrium outcome.*

To show that the truth-leaning equilibrium outcome is inclusive announcement-proof in finite-data games, I construct it, using the algorithm from Rappoport, which I summarize here. In short, the equilibrium is constructed by iteratively choosing a frontier of types such that the set of types "above" the frontier, in the sense of being able to imitate some frontier type, yields as favorable a belief as possible.

Algorithm (Finite N). First, define for any type set T the subset of types $T^+(M) = T \bigcap U(M)$ as the set of types in T that are capable of sending some message in message set M, and define $u_{pool}(T)$ to be the payoff to the sender if the receiver knows only that their type must be in T.

1. Let $T_1 = \mathcal{T}_N$, and find the set of messages $M_1 \subseteq T_1$ that maximizes the payoff to a pool consisting of the set of senders in T_1 who can send at least one message in it:

$$M_1 \in \arg\max_{M \subseteq T_1} u_{pool}(T_1^+(M)).$$

   If there are multiple such pools, then we take their union, which is also such a pool.

2. For $s = 2$ onwards, restrict the set of types to $T_s = T_{s-1} \setminus T_{s-1}^+(M_{s-1})$, and find (the union of)

$$M_s \in \arg\max_{M \in T_s} u_{pool}(T_s^+(M)).$$

3. Continue until $T_s \setminus T_s^+(M_s) = \emptyset$. Given each set M_s, there always exists a mixed strategy profile σ_{pool}^M defined over types in $T_1^+(M)$ such that each message in M yields the same payoff under the receiver's induced beliefs from σ_{pool}^M.[22] Define σ^* by $\sigma^*(m|t) = \hat{\sigma}_{pool}^{M_s}(m|t)$ where

[22] Otherwise, the worst possible payoff to particular message in M over all strategy profiles over M_s is better than the best possible payoff to some other message; then there always exists $M \subset M_s$ such that $T_s^+(M) > T_s^+(M_s)$.

M_s is the pool containing m.

Proof. (Unique credible inclusive announcement-proof outcome). By construction, there is no credible inclusive announcement, since such an announcement would constitute a better set of types than the one constructed at some step of the algorithm; this violates the optimality of the pool of types constructed in each step. No other outcome is immune: if $u_{\sigma^*_{alt}} \neq u_{\sigma^*}$, then there exists a v such that the set of pools achieving a payoff greater than v is identical in $u_{\sigma^*_{alt}}$ and u_{σ^*}, but the pool of types T achieving payoff v under u_{σ^*} is a strict superset of that under $u_{\sigma^*_{alt}}$. Then types in T can make a credible inclusive announcement that they will play as they do in σ^*. □

Next, we show that the separation theorem continues to hold for the truth-leaning equilibrium σ^*_N. To do so, consider the frontier

$$\hat{M}_N(v) = \{t \in \mathcal{T}_N : u_{\sigma^*_N}(t) \geq v,\ \nexists t' \in \mathcal{T}_N \text{ w/ } t' \hat{\subset} t \text{ and } u_{\sigma^*_N}(t') \geq v\}.$$

We can define upper and lower pools completely analogously to the main case:

Definition D.1. *An* upper pool *of payoff frontier $\hat{M}(v)$ is a set*

$$\bar{T} = U(\hat{M}(v)) \setminus U(M)$$

for some collection of messages M.

Definition D.2. *A* lower pool *of of payoff frontier $\hat{M}(v)$ is a set*

$$\underline{T} = U(M) \setminus U(\hat{M}(v))$$

for some collection of messages M.

Claim 16 (Separation for finite N)**.** *For any nonempty upper pool $\bar{T}$ and lower pool $\underline{T}$ of $\hat{M}(v)$,*

$$u_{pool}(\bar{T}) \geq v > u_{pool}(\underline{T}).$$

Proof. It is clear that $v > u_{pool}(\underline{T})$ from the construction algorithm: if not, then when constructing the lowest-payoff pool of senders who obtain value at least v, the algorithm could not have chosen the payoff-maximizing upper pool, as adding $\underline{T}$ would have strictly increased the payoff to the pool.

To show that $u_{pool}(\bar{T}) \geq v$, note that if not, then similarly, if the lowest-payoff pool of senders who obtain value at least v is T_v, then we could instead take

$$T'_v = T_v \bigcup U(M)$$

at the step that T_v was constructed. We find that this is a strictly higher-payoff upper pool, which rules out $u_{pool}(\bar{T}) < v$. $\square$

E Proof of convergence to imitation equilibrium as $N \to \infty$

Here, we first use the separation theorem to prove that outcomes $u_{\sigma_N^*}$ converge to u_{σ^*} in a convergent sequence of games; then, we state a corresponding result that describes the extent to which the sender's messaging strategy also converges, and we give a proof.

To prove Theorem 6, we first introduce some notation. Let payoff frontiers in $\mathcal{G}_N$ under a truth-leaning equilibrium σ_N^* be given by $\hat{M}_N(v)$ — that is,

$$\hat{M}_N(v) = \{t \in \mathcal{T}_N : u_{\sigma_N^*}(t) \geq v,\ \nexists t' \in \mathcal{T}_N \text{ w/ } t' \hat{\subset} t \text{ and } u_{\sigma_N^*}(t') \geq v\}.$$

Let a translation of $M_N(v)$ to the restricted limit type space $\mathcal{T}$ be $\hat{\mu}^N(v)$, where

$$\hat{\mu}_j^N(v) = \{\mu \in [0,1] : u_{\sigma_N^*}(\mu f_j) \geq v, \text{ and } u_{\sigma_N^*}(\mu' f_j) < v\ \forall \mu' < \mu\}.$$

Finally, let

$$u_{pool,N}(T) = \frac{\sum_{t \in T} q_N(t) \mathbb{E}_{\pi(\cdot|t)}[\theta]}{\sum_{t \in T} q_N(t)}$$

be the analog of u_{pool} for finite game $\mathcal{G}_N$. We first give a lemma that shows that average values in the finite game converge to those in the continuous-μ game within upper and lower pools.

Lemma 13. *Fix some $\epsilon > 0$ and $\delta > 0$. We aim to show that there exists large-enough $\hat{N}(\epsilon, \delta)$ such that for $N > \hat{N}(\epsilon, \delta)$, we can ensure that neither $U(\hat{\mu}(v + \epsilon)) \setminus U_N(\hat{M}_N(v))$ nor $U_N(\hat{M}_N(v)) \setminus U(\hat{\mu}(v - \epsilon))$ contain more than a measure δ of types in $\mathcal{T}$.*

Proof of Lemma 13. Fix an integer n. The Glivenko-Cantelli theorem implies that there is a bound on the probability that $\sup_d |\sum_{x=1}^{d} t(x) - \frac{n}{N} F_j(d)| > \eta$ conditional on $|t| = n$ and the true state being θ_j that decreases to 0 for large n, irrespective of N. Because data have a discrete distribution, this implies a similar bound on the empirical probability mass function. Before proceeding further, we formalize the implications for the problem at hand.

It is helpful to formally define a neighborhood of $\{\mu f_j : \mu \in [0,1]\}$ in $\mathcal{T}_N$. For $\eta > 0$, define

$$S_N^j(\eta) = \{t \in \mathcal{T}_N : \exists \theta \text{ s.t. } \sup_d |t(d) - |t| f_\theta(d)| \leq \eta\}.$$

This is the set of datasets in $\mathcal{T}_N$ that differ from some type in $\mathcal{T}$ for which the state is θ_j by no more than a fraction η of observations of each outcome. Furthermore, taking any lower bound

$k^* \in (0,1]$, let $S_N^j(\eta, k^*)$ be the set $S_N^j(\eta) \bigcap \{t : |t| \geq Nk^*\}$ of all datasets in the neighborhood that contain at least a fraction k^* of total possible observations.

We can ensure that, if we know that the state is $\theta = theta_j$ and the number of observations of the data the sender observes is $n > Nk^*$, then the likelihood that their type lies in $S_N^j(\eta, k^*)$ is close to 1 for all η as long as N is sufficiently large. Furthermore, if η is small enough for a given value of k^*, then $S_N^j(\eta, k^*)$ are disjoint. There is a sufficient upper-bound to the value of η that achieves this, $\tilde{\eta}(k^*) = \min_{j,j'} \max_d k \frac{f_j(d) - f_{j'}(d)}{2}$.

For any desired likelihoods $l_1 > 0$ and $l_2 > 0$, and any error rates $\xi_1 > 0$ and $\xi_2 > 0$, there exists a uniform bound $\tilde{N}(k^*, \eta, l_1, l_2, \xi_1, \xi_2)$ such that as long as $k > k^*$ and $N > \tilde{N}(k^*, \eta, l_1, l_2, \xi_1, \xi_2)$, with $\eta \leq \tilde{\eta}(k^*)$,

a) $Pr(t \in S_N^j(\eta, k^*) | \theta = \theta_j, n = Nk) \geq 1 - l_1$

b) $Pr(t \in S_N^j(\eta, k^*) | \theta \neq \theta_j, n = Nk) < l_2$ for any $k > k^*$

c) $|\mathbb{E}_{\pi_N(\cdot|t)}[\theta] - \theta_j| \leq \xi_1$ for all $t \in S_N^j(\eta, k^*)$

d) $|(G_N(Nk) - G_N(Nk')) - (G(k) - G(k'))| < \xi_2|$ for all $0 \leq k' \leq k \leq 1$.

Part (a) follows directly from the Glivenko-Cantelli theorem. Part (b) follows from applying part (a) to $j' \neq j$, although the bound could certainly be tightened more. Part (c) follows from both the previous parts and the fact that the set of possible values of θ is finite. Part (d) follows from the convergence of $\mathcal{G}_1, \ldots,$ to $\mathcal{G}$.

Now we can bound the average value of $U(\hat{\mu}(v + \epsilon)) \setminus U_N(\hat{M}(v))$ in $\mathcal{T}_N$.

Let us start with types in $S_N^j(\eta, k) \bigcap U(\hat{\mu}(v + \epsilon)) \setminus U_N(\hat{M}(v))$. In $\mathcal{T}$, we know that $U(\hat{\mu}(v + \epsilon)) \setminus U_N(\hat{M}(v))$ contains μf_j for $\mu \in [\hat{\mu}_j(v + \epsilon), \hat{\mu}_j^N(v)]$. Recall that $|\mathcal{D}| = D$, and define $R = \max_{j,d,d'} \frac{f_j(d)}{f_j(d')}$. We know that for all $k \in [\hat{\mu}_j(v + \epsilon) + DR\eta, \hat{\mu}_j^N(v) - DR\eta]$, if $t \in S_N^j(\eta, k)$ and $|t| = Nk$ then $t \in S_N^j(\eta, k) \bigcap U(\hat{\mu}(v + \epsilon)) \setminus U_N(\hat{M}_N(v))$. This is true because:

- If $t \in S_N^j(\eta, k)$, then the nearest type $\mu f_j \in \mathcal{T}$ differs by adding or deleting at most a mass η of observations of each outcome.
- As a result, the type $(k + DR\eta) f_j$ can imitate t, and the type $(k - DR\eta) f_j$ can be imitated by t.
- Therefore, if, in addition, $k \in [\hat{\mu}_j(v + \epsilon) + RD\eta, \hat{\mu}_j^N(v) - RD\eta]$, then $t \in U(\hat{\mu}(v + \epsilon)) \setminus U_N(\hat{M}_N(v))$.

Likewise, unless $k \in [\hat{\mu}_j(v + \epsilon) - DR\eta, \hat{\mu}_j^N(v) + DR\eta]$, if $t \in S_N^j(\eta, k)$ and $|t| = Nk$ then $t \notin S_N^j(\eta, k) \bigcap U(\hat{\mu}(v + \epsilon)) \setminus U_N(\hat{M}_N(v))$.

The probability that the raw dataset in $\mathcal{G}_N$ lies in $U(\hat{\mu}(v+\epsilon)) \setminus U_N(\hat{M}_N(v))$ is therefore lower-bounded by

$$\beta_0(\theta_j)(1-l_1)[G(\hat{\mu}_j^N(v)+RD\eta(k^*)) - G(\hat{\mu}_j(v+\epsilon)-RD\eta(k^*)) - 2\xi_2] - G_N(k^*N)$$

and upper-bounded by

$$\beta_0(\theta_j)[G(\hat{\mu}_j^N(v)-RD\eta(k^*)) - G(\hat{\mu}_j(v+\epsilon)+RD\eta(k^*)) + 2\xi_2] + l_2$$

whenever $N > \tilde{N}(k^*, \eta, l_1, l_2, \xi_1, \xi_2)$.

In addition, we can define the upper bound $\bar{b} = \max_\mu g(\mu)$ because g is continuous on a compact interval. Then if

$$\sum_{j=1}^{J}[G(\hat{\mu}_j^N(v)) - G(\hat{\mu}_j(v+\epsilon))] \geq \delta,$$

a crude lower bound on the average value of types in $U(\hat{\mu}(v+\epsilon)) \setminus U_N(\hat{M}_N(v))$ under the finite-game type distribution q_N is

$$\begin{aligned}
&\frac{\sum_{j=1}^{J}[\beta_0(\theta_j)(1-l_1)[G(\hat{\mu}_j^N(v)+RD\eta) - G(\hat{\mu}_j(v+\epsilon)-RD\eta) - 2\xi_2] - G(k) - \xi_2](\theta_j - \xi)}{\beta_0(\theta_j)[G(\hat{\mu}_j^N(v)-RD\eta) - G(\hat{\mu}_j(v+\epsilon)+RD\eta) + 2\xi_2] + l_2} \\
&\leq \frac{\delta(v-\xi) - l_1 - 2\bar{b}RD\eta - G(k) - 3\xi_2}{\delta + l_2 + l_1 + 2\bar{b}RD\eta + 2\xi_2} \qquad (16)\\
&\equiv LB(k^*, \eta, l_1, l_2, \xi_1, \xi_2 | \delta)
\end{aligned}$$

as long as $N > \tilde{N}(k^*, \eta, l_1, l_2, \xi_1, \xi_2)$. Note that this bound is independent of v, that is, it applies uniformly to all payoff frontiers.

We have that

$$\lim_{N\to\infty} \min_{\substack{k^*,\eta,l_1,l_2,\xi_1,\xi_2: \\ \tilde{N}(k^*,\eta,l_1,l_2,\xi_1,\xi_2)\leq N}} LB(\eta, k^*, l_1, l_2, \xi_1, \xi_2|\delta) = \frac{\sum_{j=1}^{J}[G(\hat{\mu}_j^N(v)) - G(\hat{\mu}_j(v+\epsilon))]\theta_j}{\sum_{j=1}^{J}[G(\hat{\mu}_j^N(v)) - G(\hat{\mu}_j(v+\epsilon))]} \geq v+\epsilon,$$

by the separation theorem applied to types in $U(\hat{\mu}(v+\epsilon)) \setminus U_N(\hat{M}_N(v))$ under the continuous-data type distribution q. But we also know by applying the separation theorem that the average value of types in $U(\hat{\mu}(v+\epsilon)) \setminus U_N(\hat{M}_N(v))$ under the finite-game type distribution q_N is no more than v. Then there exists large-enough $\hat{N}_+(\epsilon, \delta)$ such that for $N > \hat{N}_+(\epsilon, \delta)$, the above bound ensures that for any v, the set $U(\hat{\mu}(v+\epsilon)) \setminus U_N(\hat{M}_N(v))$ does not contain more than a measure δ of types in $\mathcal{T}$.

We can show, using a completely symmetric argument, that there also exists large-enough $\hat{N}_-(\epsilon, \delta)$ such that for $N > \hat{N}_-(\epsilon, \delta)$, the set $U_N(\hat{M}_N(v)) \setminus U(\hat{\mu}(v-\epsilon))$ does not contain more than a measure δ of types in $\mathcal{T}$, regardless of v. Rather than lower-bounding the average value of all types in $U_N(\hat{M}_N(v)) \setminus U(\hat{\mu}(v-\epsilon))$ under q_N, we upper-bound it and show that given fixed δ, for large N it is less than ϵ greater than the value of the same set of types evaluated under type distribution q. This shows that it has average value less than v, and fails the separation theorem for the finite game. □

Next, we use the the lemma to prove the theorem.

Proof of Theorem 6. There is an upper bound $\bar{\Delta} = \max_{\mu,j} \frac{d\hat{u}_j(\mu)}{d\mu}$ on the rate of change of payoffs in μ under σ^*.

In addition, for every $y > 0$, there is some $C(y) = \min_{\mu,j} \beta_0(\theta_j)(G(\mu) - G(\mu - y)) > 0$ with $\lim_{x\to 0} C^{-1}(x) = 0$, which shows that upper-bounding the measure of an interval $[\mu - y, \mu]$ of measures of types under state θ_j with respect to prior distribution q also upper-bounds y itself.

From Lemma 13, we know that for every payoff frontier v associated with the payoff of type $\mu f_j \in T$ in the imitation equilibrium of $\mathcal{G}_N$, the frontier associated with type $(\mu - C^{-1}(\delta) f_j)$ represents a payoff no more than $v + \epsilon$ under $\mathcal{G}$ for all $N > \hat{N}(\epsilon, \delta)$, since more than a fraction δ of types lie in $U(\hat{\mu}(v+\epsilon)) \setminus U_N(\hat{M}_N(v))$. Then we know that for $N > \hat{N}(\epsilon, \delta)$,

$$u_{\sigma_N^*}(\mu f_j) \geq u_{\sigma^*}((\mu - C^{-1}\delta) f_j) - \epsilon \geq u_{\sigma^*}(\mu f_j) - C^{-1}(\delta)\bar{\Delta} - \epsilon,$$

and likewise,

$$u_{\sigma_N^*}(\mu f_j) \leq u_{\sigma^*}((\mu + C^{-1}\delta) f_j) + \epsilon \leq u_{\sigma^*}(\mu f_j) + C^{-1}(\delta)\bar{\Delta} + \epsilon.$$

We can take

$$\min_{\delta,\epsilon: N > \hat{N}(\epsilon,\delta)} C^{-1}(\delta)\bar{\Delta} + \epsilon$$

to be the bound on the difference between $u_{\sigma^*}(t) - u^*_{\sigma_N^*}(t)$, and it shrinks to 0 as $N \to \infty$. □

E.1 Strategic convergence

Proposition 14. *Suppose that $\{\mathcal{G}_N\}_{N=1}^{\infty}$ converge to $\mathcal{G}_\infty$. Then for all $p^*, \rho, \eta > 0$, there is $\underline{N}(p^*, \eta)$ such that for all $N > \underline{N}(p^*, \eta)$, conditional on $|u_{\sigma_N^*}(t) - \theta_k| > \eta$ for all k, there is at least probability $1 - p^*$ that t sends a message with (sup norm) distance at most δ from some $t_\infty \in \mathcal{T}_\infty$ that is on-path in σ_∞^* and such that $|u_{\sigma_N^*}(t) - u_{\sigma_\infty^*}(t_\infty)| \leq \rho$.*

This is a partial characterization of large-N equilibrium strategies, saying that among types that obtain payoff bounded away by an arbitrarily small amount from the rewards to certainty about any particular state, the likelihood of playing a message close to their optimal message under limit-game beliefs is very high when there is plentiful access to data. In other words, these types play imitation-like strategies. This follows from the convergence theorem, since in truth-leaning equilibrium a type that receives payoff less than its full-information payoff always discloses its full dataset, so types with $u_{\sigma^*_N}(t) << \mathbb{E}_{\pi(\cdot|t)}[\theta]$ tell the truth; convergence of outcomes implies that they receive payoffs similar to those obtained by nearby types in T_∞ under $\beta_{\sigma^*_\infty}$, and convergence of the type distribution implies that most such senders are indeed near some type in T_∞. In aggregate a similar set of imitators must pool with such senders as the set of imitators pooling with better-state senders in σ^*_∞, which means that types with $u_{\sigma^*_N}(t) >> \mathbb{E}_{\pi(\cdot|t)}[\theta]$ play messages close to $\mathcal{T}_\infty$ with high probability.

The caveat is that when $|\mathbb{E}_{\beta_{\sigma^*_N}(\cdot|m)}[\theta] - \theta_k|$ is small for some k, then there may be no significant mass of senders playing m to earn a payoff much greater or much less than their full-information payoff, which makes it hard to apply the technique of matching imitators to the imitated, though we do not have a counterexample for this case. From Corollary 4, we know that there is a positive-measure set of types that receive payoffs close to their full-information payoffs in σ^*_∞, and the proposition does not pin down the large-N limit of equilibrium strategies of types close to them, but generically, besides these, the set of types excluded from the proposition is measure-0.[23]

Proof of Prop. 14. Define $m_{\sigma^*_N}(t)$ to be the realization of the message played when the sender's type is t – formally, $m_{\sigma^*_N}(t)$ is a random variable with outcomes in $\mathcal{M}_N$ whose distribution is given by the equilibrium strategy $\sigma^*_N(\cdot|t)$.

Define $A_N(x, \Delta; \epsilon)$ to be the set of types $t \in \mathcal{T}_N \bigcap T(\epsilon)$ such that $u_{\sigma^*_N}(t) > \mathbb{E}_{\pi(\cdot|t)}[\theta]$, and $u_{\sigma^*_N}(t) \in (x, x + \Delta]$.

Define $B_N(x, \Delta; \epsilon)$ to be the set of types $t' \in \mathcal{T}_N \bigcap T(\epsilon)$ with $u_{\sigma^*_N}(t') < \mathbb{E}_{\pi(\cdot|t)}[\theta]$, and $u_{\sigma^*_N}(t') \in (x, x + \Delta]$.

Define $X(\eta, \xi, \omega) = \{x \in \left[\max_j \hat{u}_j(\xi) + \omega, u_{\sigma^*}(f_{\theta_J})\right) : \min_k |u_{\sigma^*_N}(t) - \theta_k| > \eta\}$.

For small enough η, ξ and ω, $X(\eta, \xi, \omega)$ is nonempty. On the other hand, $A_N(x, \Delta; \epsilon)$ and $B_N(x, \Delta; \epsilon)$ may be empty, in particular for small N. However, for $x \in X(\eta, \xi, \omega)$, there is large-enough $\underline{N}^*$ so that they are nonempty for all $N > \underline{N}^*$. Continuity of $u_{\sigma^*}(\mu f_j)$ ensures there is positive-measure set of types in $\mathcal{T}$ with $u_{\sigma^*_{infty}}(t) \in [x, x + \Delta]$; the bound away from θ_k for all k ensures that some such types have $u_{\sigma^*_{infty}}(t) - \mathbb{E}_{\pi(\cdot|t)}[\theta] \geq \eta$ and some have $u_{\sigma^*_i nfty}(t) \leq \mathbb{E}_{\pi(\cdot|t)}[\theta] < -\eta$, and so there is a positive-measure set of types nearby with the same properties under σ^*_N in $\mathcal{T}_N$ for large-enough N.

We first prove a claim.

[23] Genericity here can be with respect to perturbations in β_0 or $\theta_1, \ldots, \theta_J$.

Claim 17. *If $\{\sigma^*_N\}_{N=1}^{\infty}$ are truth-leaning equilibria of games $\mathcal{G}_N$ that converge to limit game $\mathcal{G}$ with imitation equilibrium σ^*, then for any $\eta > 0$, $\xi > 0$, $\omega > 0$ and $p > 0$, there exists $\bar{\epsilon} > 0$ and $\bar{\Delta} > 0$ such that, for all $x \in X(\eta, \xi, \omega)$, the probability conditional on $t \in A_N(x, \Delta; \epsilon)$ that $m_{\sigma^*_N}(t) \in B_N(x, \Delta; \epsilon)$ is at least $1 - p$ in the limit as $N \to 0$ for all $\epsilon < \bar{\epsilon}$ and $\Delta < \bar{\Delta}$.*

Proof of Claim 17. Expanding out the realization of $m_{\sigma^*_N}(t)$, this is equivalent to saying that for given $\eta > 0$, $\xi > 0$, $\omega > 0$, $p > 0$, there exists $\bar{\Delta} > 0$ and $\bar{\epsilon} > 0$ so that for all $\Delta < \bar{\Delta}$ and $\epsilon < \bar{\epsilon}$,

$$\lim_{N\to\infty} \frac{\sum_{t\in A_N(x,\Delta;\epsilon)} \left[q_N(t) \sum_{t'\in B_N(x,\Delta;\epsilon)} \sigma^*_N(t'|t)\right]}{\sum_{t\in A_N(x,\Delta;\epsilon)} q_N(t)} \geq 1 - p.$$

for all $x \in X(\eta, \xi, \omega)$.

We have that for any $\xi > 0$, $\omega > 0$,

$$\lim_{\epsilon\to 0} \lim_{N\to\infty} \min_{\theta_j} \max_{t\in\mathcal{T}_N \bigcap T(\epsilon): u_{\sigma^*_N}(t) \geq \max_j \hat{u}_j(\xi)+\omega} |\theta_j - \mathbb{E}_{\pi(\cdot|t)}[\theta]| = 0.$$

Thus for any $\nu > 0$, $\omega > 0$ and $\xi > 0$, there exist small-enough $\bar{\epsilon}(\nu, \xi, \omega) > 0$ and large-enough $\underline{N}(\nu, \xi, \omega, \epsilon)$ defined for $\epsilon < \bar{\epsilon}(\nu, \xi)$ such that $\min_{\theta_j} \max_{t\in\mathcal{T}_N \bigcap T(\epsilon): u_{\sigma^*_N}(t) \geq \max_j \hat{u}_j(\xi)+\omega} |\theta_j - \mathbb{E}_{\pi(\cdot|t)}| < \nu$ for all $\epsilon < \bar{\epsilon}(\nu, \xi, \omega)$, $N > \underline{N}(\nu, \xi, \omega, \epsilon)$.

If we take $\Delta < \eta/3$ and $\nu < \eta/3$, then whenever $|\theta_k - \mathbb{E}_{\pi(\cdot|t)}| < \nu$ for some k and $u_{\sigma^*_N}(t) \in [x, x + \Delta]$ for x in $X(\eta, \xi, \omega)$, we have that $|u_{\sigma^*_N}(t) - \mathbb{E}_{\pi(\cdot|t)}| > \eta/3$. In particular, $u_{\sigma^*_N}(t) \neq \mathbb{E}_{\pi(\cdot|t)}$, so, for any $x \in X(\eta, \xi, \omega)$, and for any $\Delta, \nu < \eta/3$ and any ξ, ω and $\epsilon < \bar{\epsilon}(\nu, \xi, \omega)$, $N > \underline{N}(\nu, \xi, \omega, \epsilon)$,

$$A_N(x, \Delta; \epsilon) \bigcup B_N(x, \Delta; \epsilon) = \{t \in T(\epsilon) \bigcap \mathcal{T}_N : u_{\sigma^*_N}(t) \in (x, x + \Delta]\}.$$

In addition, the uniform convergence of outcomes on $\mathcal{T}$ and of the type distribution conditional on each state ensures that for all ϵ, ξ, Δ, and for all $x \geq \max_j \hat{u}_j(\xi) + \omega$,

$$\lim_{N\to\infty} \sum_{t\in A_N(x,\Delta;\epsilon)} q(t)\mathbb{E}_{\pi(\cdot|t)}[\theta] = \sum_{\theta_j<\theta_k} \theta_j \beta_0(\theta_j)[G(\hat{\mu}_j(x + \Delta)) - G(\hat{\mu}_j(x))] \tag{17}$$

and

$$\lim_{N\to\infty} \sum_{t\in A_N(x,\Delta;\epsilon)} q(t) = \sum_{\theta_j<\theta_k} \beta_0(\theta_j)[G(\hat{\mu}_j(x + \Delta)) - G(\hat{\mu}_j(x))] \tag{18}$$

and likewise

$$\lim_{N\to\infty} \sum_{t\in B_N(x,\Delta;\epsilon)} q(t)\mathbb{E}_{\pi(\cdot|t)}[\theta] = \sum_{\theta_j>\theta_k} \theta_j \beta_0(\theta_j)[G(\hat{\mu}_j(x + \Delta)) - G(\hat{\mu}_j(x))] \tag{19}$$

and

$$\lim_{N\to\infty} \sum_{t\in B_N(x,\Delta;\epsilon)} q(t) = \sum_{\theta_j>\theta_k} \beta_0(\theta_j)[G(\hat{\mu}_j(x+\Delta)) - G(\hat{\mu}_j(x))]. \tag{20}$$

Supposing that $x \in X(\eta,\xi,\omega)$ for some θ_k, and $\epsilon < \bar{\epsilon}(\nu,\xi,\omega)$, and $\Delta, \nu < \eta/3$, we know from the above that

$$\begin{aligned} x &\le \frac{\sum_{\theta_j} \theta_j \beta_0(\theta_j)[G(\hat{\mu}_j(x+\Delta)) - G(\hat{\mu}_j(x))]}{\sum_{\theta_j} \beta_0(\theta_j)[G(\hat{\mu}_j(x+\Delta)) - G(\hat{\mu}_j(x))]} \\ &= \lim_{N\to\infty} \frac{\sum_{t\in B_N(x,\Delta;\epsilon)} q(t)\mathbb{E}_{\pi(\cdot|t)}[\theta] + \sum_{t\in A_N(x,\Delta;\epsilon)} q(t)\mathbb{E}_{\pi(\cdot|t)}[\theta]}{\sum_{t\in B_N(x,\Delta;\epsilon)} q(t) + \sum_{t\in A_N(x,\Delta;\epsilon)} q(t)} \end{aligned} \tag{21}$$

On the other hand,

$$\begin{aligned} &x+\Delta \\ &\ge \lim_{N\to\infty} \frac{\sum_{t\in B_N(x,\Delta;\epsilon)} q(t)\mathbb{E}_{\pi(\cdot|t)}[\theta] + \sum_{t\in A_N(x,\Delta;\epsilon)} \left[q(t)\mathbb{E}_{\pi(\cdot|t)}[\theta] \sum_{t'\in B_N(x,\Delta;\epsilon)} \sigma_N^*(t'|t)\right]}{\sum_{t\in B_N(x,\Delta;\epsilon)} q(t) + \sum_{t\in A_N(x,\Delta;\epsilon)} \left[q(t) \sum_{t'\in B_N(x,\Delta;\epsilon)} \sigma_N^*(t'|t)\right]} \\ &= \lim_{N\to\infty} \frac{\sum_{\theta_j} \theta_j\beta_0(\theta_j)[G(\hat{\mu}_j(x+\Delta)) - G(\hat{\mu}_j(x))] - \sum_{t\in A_N(x,\Delta;\epsilon)} q(t)\mathbb{E}_{\pi(\cdot|t)}[\theta] \sum_{t'\in B_N(x,\Delta;\epsilon)}(1-\sigma_N^*(t|t'))}{\sum_{\theta_j} \beta_0(\theta_j)[G(\hat{\mu}_j(x+\Delta)) - G(\hat{\mu}_j(x))] - \sum_{t\in A_N(x,\Delta;\epsilon)} q(t) \sum_{t'\in B_N(x,\Delta;\epsilon)}(1-\sigma_N^*(t|t'))}. \end{aligned} \tag{22}$$

But, we know that $\mathbb{E}_{\pi(\cdot|t)}[\theta] < u_{\sigma_N^*}(t) - \eta/3 \le x+\Delta-\eta/3$ for any $t \in A_N(x,\Delta;\epsilon)$. Then, combining, we have

$$\begin{aligned} 0 \le &(x+\Delta)\left(\sum_{\theta_j} \beta_0(\theta_j)[G(\hat{\mu}_j(x+\Delta)) - G(\hat{\mu}_j(x))] - \lim_{N\to\infty} \sum_{t\in A_N(x,\Delta;\epsilon)} q(t) \sum_{t'\in B_N(x,\Delta;\epsilon)} (1-\sigma_N^*(t|t'))\right) \\ &- x\left(\sum_{\theta_j} \beta_0(\theta_j)[G(\hat{\mu}_j(x+\Delta)) - G(\hat{\mu}_j(x))]\right) - \lim_{N\to\infty} \sum_{t\in A_N(x,\Delta;\epsilon)} q(t)\mathbb{E}_{\pi(\cdot|t)}[\theta] \sum_{t'\in B_N(x,\Delta;\epsilon)} (1-\sigma_N^*(t|t')) \\ \le &\Delta\left(\sum_{\theta_j} \beta_0(\theta_j)[G(\hat{\mu}_j(x+\Delta)) - G(\hat{\mu}_j(x))]\right) \\ &\qquad - [(x+\Delta) - (x+\Delta-\eta/3)] \lim_{N\to\infty} \sum_{t\in A_N(x,\Delta;\epsilon)} q(t) \sum_{t'\in B_N(x,\Delta;\epsilon)} (1-\sigma_N^*(t|t')) \\ = &\lim_{N\to\infty} \Delta\left(\sum_{t\in A_N(x,\Delta;\epsilon)\bigcup B_N(x,\Delta;\epsilon)} q(t)\right) - \frac{\eta}{3} \lim_{N\to\infty} \sum_{t\in A_N(x,\Delta;\epsilon)} q(t) \sum_{t'\in B_N(x,\Delta;\epsilon)} (1-\sigma_N^*(t|t')). \end{aligned} \tag{23}$$

Finally, since

$$\lim_{N\to\infty} \frac{\sum_{t\in A_N(x,\Delta;\epsilon)\bigcup B_N(x,\Delta;\epsilon)} \mathbb{E}_{\pi(\cdot|t)}[\theta]q(t)}{\sum_{t\in A_N(x,\Delta;\epsilon)\bigcup B_N(x,\Delta;\epsilon)} q(t)} \le x+\Delta,$$

we have

$$\sum_{t\in B_N(x,\Delta;\epsilon)} (\theta_{k+1}-x-\Delta)q(t) \le \sum_{t\in A_N(x,\Delta;\epsilon)} (x+\Delta-\theta_1)q(t)$$

and so

$$\frac{\sum_{t\in A_N(x,\Delta;\epsilon)} q(t)}{\sum_{t\in B_N(x,\Delta;\epsilon)} q(t)} \ge \frac{\eta/3}{\theta_J}.$$

From the above and eq. 23, we have

$$\lim_{N\to\infty} \frac{\sum_{t\in A_N(x,\Delta;\epsilon)} \left[q_N(t)\sum_{t'\in B_N(x,\Delta;\epsilon)}(1-\sigma_N^*(t'|t))\right]}{\sum_{t\in A_N(x,\Delta;\epsilon)} q_N(t)} \le \frac{\Delta(\theta_J+\eta)}{(\eta/3)^2}.$$

Then for given p, ξ, ω, η, and $\nu < \eta/3$ and $\epsilon < \bar{\epsilon}(\nu,\xi,\omega)$, as long as $\Delta \le \frac{p\eta^2}{9(\theta_J+\eta)}$, we have

$$\lim_{N\to\infty} \frac{\sum_{t\in A_N(x,\Delta;\epsilon,\xi)} \left[q_N(t)\sum_{t'\in B_N(x,\Delta;\epsilon,\xi)}\sigma_N^*(t'|t)\right]}{\sum_{t\in A_N(x,\Delta;\epsilon,\xi)} q_N(t)} \ge 1-p$$

and this bound is independent of x. So, letting $\bar{\epsilon} = \bar{\epsilon}(\eta/3,\xi,\omega)$ and $\bar{\Delta} = \frac{p\eta^2}{9(\theta_J+\eta)}$, we have proven the claim. □

Next, let $\underline{u} = \hat{u}_j(0)$, where the choice of j for the definition does not matter. Suppose the following condition holds for some positive η:

Condition 1. $\min_k |\theta_k - \underline{u}| > \eta$ *and there is* $\tilde{\xi} > 0$ *such that* $G(\tilde{\xi}) > 0$ *for all* j*, and* $\hat{u}_j(\tilde{\xi}) = \underline{u}$ *for some* j*.*

This says that a sender with a positive amount $\tilde{\xi}$ of distribution f_j gets the same payoff as the sender with no data, and that that payoff is bounded away from any θ_j by η. When showing that senders must play similarly under σ_N^* in the limit as the average dataset becomes large, we consider separately the small fraction of senders that, by chance, receive very little data, i.e. those with $|t| \le \xi$, and in this case

$$\lim_{\xi\to 0}\lim_{\omega\to 0} \min\{\mu : \exists j \text{ s.t. } \hat{u}_j(\mu) > \xi+\omega\} > 0,$$

and there is a positive-measure set of types that may be pooled with those low-data senders.

Let us prove a similar claim to the previous one.

Claim 18. *If there exists η such that condition 1 holds, then for given $p > 0$, there exists $\bar{\epsilon} > 0$ such that for $\epsilon < \bar{\epsilon}$, if we define*

$$S(\xi, \omega, \epsilon) = \{t \in \mathcal{T}_N \bigcap T(\epsilon) : |t| \leq \xi \text{ and } u_{\sigma_N^*}(t) \in (x, x + \Delta]\},$$

then letting $a_N(\xi, \omega, \epsilon) = A_N(\underline{u} - \omega, 2\omega; \epsilon) \setminus S(\xi, \omega, \epsilon)$ and $b_N(\xi, \omega, \epsilon) = B_N(\underline{u} - \omega, 2\omega; \epsilon) \bigcup S(\xi, \omega, \epsilon)$, we have

$$\lim_{\xi \to 0} \lim_{\omega \to 0} \lim_{N \to \infty} \frac{\sum_{t \in a_N(\xi,\omega,\epsilon)} \left[q_N(t) \sum_{t' \in b_N(\xi,\omega,\epsilon)} \sigma_N^*(t'|t) \right]}{\sum_{t \in a_N(\xi,\omega,\epsilon)} q_N(t)} \geq 1 - p.$$

Proof of Claim 18. To start, note that in this case, we have for all $\xi > 0$ that

$$\lim_{\epsilon \to 0} \lim_{N \to 0} \min_{\theta_j} \max_{t:|t| \leq \xi} |\theta_j - \mathbb{E}_{\pi(\cdot|t)}[\theta]| = 0.$$

Then for all $t \in a_N(\xi, \omega, \epsilon)$, there is some $\bar{\epsilon}'(\eta, \xi)$ and $\underline{N}'(\eta, \xi, \epsilon)$ so that for all $\epsilon < \bar{\epsilon}'(\eta, \xi)$ and $N > \underline{N}'(\eta, \xi, \epsilon)$, we have $u_{\sigma_N^*}(t) - \mathbb{E}_{\pi(\cdot|t)}[\theta] \geq \eta/3$.

We know that

$$\begin{aligned} \underline{u} - \omega &\leq \lim_{N \to \infty} \frac{\sum_{t \in b_N(\xi,\omega,\epsilon)} q(t) \mathbb{E}_{\pi(\cdot|t)}[\theta] + \sum_{t \in a_N(\xi,\omega,\epsilon)} q(t) \mathbb{E}_{\pi(\cdot|t)}[\theta]}{\sum_{t \in b_N(\xi,\omega,\epsilon)} q(t) + \sum_{t \in a_N(\xi,\omega,\epsilon)} q(t)} \\ &= \frac{\sum_{\theta_j} \theta_j \beta_0(\theta_j) G(\hat{\mu}_j(\underline{u} + \omega))}{\sum_{\theta_j} \beta_0(\theta_j) G(\hat{\mu}_j(\underline{u} + \omega))} \end{aligned} \tag{24}$$

and

$$\begin{aligned} \underline{u} + \omega &\geq \lim_{N \to \infty} \frac{\sum_{t \in b_N(\xi,\omega,\epsilon)} q(t) \mathbb{E}_{\pi(\cdot|t)}[\theta] + \sum_{t \in a_N(\xi,\omega,\epsilon)} \left[q(t) \mathbb{E}_{\pi(\cdot|t)}[\theta] \sum_{t' \in b_N(\xi,\omega,\epsilon)} \sigma_N^*(t'|t) \right]}{\sum_{t \in b_N(\xi,\omega,\epsilon)} q(t) + \sum_{t \in a_N(\xi,\omega,\epsilon)} \left[q(t) \sum_{t' \in b_N(\xi,\omega,\epsilon)} \sigma_N^*(t'|t) \right]} \\ &= \lim_{N \to \infty} \frac{\sum_{\theta_j} \theta_j \beta_0(\theta_j) G(\hat{\mu}_j(\underline{u} + \omega)) - \sum_{t \in a_N(\xi,\omega,\epsilon)} q(t) \mathbb{E}_{\pi(\cdot|t)}[\theta] \sum_{t' \in b_N(\xi,\omega,\epsilon)} (1 - \sigma_N^*(t|t'))}{\sum_{\theta_j} \beta_0(\theta_j) G(\hat{\mu}_j(\underline{u} + \omega)) - \sum_{t \in a_N(\xi,\omega,\epsilon)} q(t) \sum_{t' \in b_N(\xi,\omega,\epsilon)} (1 - \sigma_N^*(t|t'))}. \end{aligned} \tag{25}$$

Then, just as in eq. 23, we have when $\epsilon < \bar{\epsilon}'(\eta, \xi, \omega)$ that

$$\lim_{N \to \infty} 2\omega \left(\sum_{t \in a_N(\xi,\omega,\epsilon) \bigcup b_N(\xi,\omega,\epsilon)} q(t) \right) - \frac{\eta}{3} \lim_{N \to \infty} \sum_{t \in a_N(x,\Delta;\epsilon)} q(t) \sum_{t' \in B_N(x,\Delta;\epsilon)} (1 - \sigma_N^*(t|t')) \geq 0.$$

Since there is some j such that $\hat{u}_j(\tilde{\xi}) = \underline{u}$, we have the bound

$$\lim_{N\to\infty} \sum_{t\in a_N(\xi,\omega,\epsilon)} q_N(t) \geq \beta_0(\theta_j)[G(\tilde{\xi}) - G(\xi)].$$

So, we have that

$$\lim_{N\to\infty} \frac{\sum_{t\in a_N(\xi,\omega,\epsilon)} \left[q_N(t) \sum_{t'\in b_N(\xi,\omega,\epsilon)} (1-\sigma_N^*(t'|t))\right]}{\sum_{t\in a_N(\xi,\omega,\epsilon)} q_N(t)} \leq \frac{2\omega(1+\beta_0(\theta_j)[G(\tilde{\xi}) - G(\xi)])}{\beta_0(\theta_j)[G(\tilde{\xi}) - G(\xi)]\eta/3}.$$

This implies the claim. □

Finally, we use these claims to prove the proposition.

For any δ and ρ, there are ξ^*, $\epsilon^* > 0$ and N^* such that for all $\xi < \xi^*$, $\epsilon < \epsilon^*$, $N > N^*(\epsilon, \xi)$, and $\omega > 0$, any t' is in either $B_N(x, \Delta; \epsilon)$ for some Δ and $x > \xi + \omega$, or in $b_N(\xi, \omega, \epsilon)$ if it is at most a distance δ away from some $t \in \mathcal{T}$ with $|u_{\sigma^*}(t) - u_{\sigma_N^*}(t')| \leq \rho$.

In particular, find l such that $\theta_l \leq \bar{u} < \theta_{l+1}$, and then for any K we can construct the collection of sets

$$\left\{B_N\left(\xi + \omega + k\frac{\theta_l - (\xi+\omega)}{K}, \frac{\theta_l - (\xi+\omega)}{K}; \epsilon\right)\right\}_{k=0}^{K-1}$$

and

$$\left\{B_N\left(\theta_{j-1} + \eta + k\frac{\theta_j - \eta - (\theta_{j-1}+\eta)}{K}, \frac{\theta_j - \eta - (\theta_{j-1}+\eta)}{K}; \epsilon\right)\right\}_{k=0}^{K}, \quad \text{for all } j > l,$$

essentially partitioning the imitated senders by the payoffs they receive, into intervals that are disjoint, cover all attained payoffs except $[0, \xi+\omega]$ and the intervals $[\theta_j - \eta, \theta_j + \eta)$ and are arbitrarily small as $K \to \infty$.

Let $C(N, K, \xi, \omega, \eta; \epsilon)$ be the collection that is the union of these collections, and also includes, if condition 1 holds for η, the set $b_N(\xi, \omega, \epsilon)$. Call the elements of $C(N, K, \xi, \omega, \eta; \epsilon)$ by $C_1(N, K, \xi, \omega, \eta; \epsilon), \ldots, C_I(N, K, \xi, \omega, \eta; \epsilon)$.

Likewise, we can construct the collection of sets

$$\left\{A_N\left(\xi + \omega + k\frac{\theta_l - (\xi+\omega)}{K}, \frac{\theta_l - (\xi+\omega)}{K}; \epsilon\right)\right\}_{k=0}^{K-1}$$

and

$$\left\{A_N\left(\theta_{j-1} + \eta + k\frac{\theta_j - \eta - (\theta_{j-1}+\eta)}{K}, \frac{\theta_j - \eta - (\theta_{j-1}+\eta)}{K}; \epsilon\right)\right\}_{k=0}^{K}, \quad \text{for all } j > l,$$

which are corresponding sets of imitating types; let $D(N, K, \xi, \omega, \eta; \epsilon)$ be the collection containing these as well as $a_N(\xi, \omega, \epsilon)$ if condition 1 holds for η. Call the elements of $D(N, K, \xi, \omega, \eta; \epsilon)$ by $D_1(N, K, \xi, \omega, \eta; \epsilon), \ldots, D_I(N, K, \xi, \omega, \eta; \epsilon)$.

The proposition follows from proving that the ex-ante probability that the sender imitates some t' that is in an element of $C(N, K, \xi, \omega, \eta; \epsilon)$ converges to 1 in the large N limit and in the limit as $K \to \infty$ and $\epsilon, \omega, \xi, \eta \to 0$.

To see this, first observe that, from the above two claims, if we define

$$LB(N, K, \xi, \omega, \eta; \epsilon) = \min_i Pr(m_{\sigma_N^*}(t) \in C_i(N, K, \xi, \omega, \eta; \epsilon) | t \in D_i(N, K, \xi, \omega, \eta; \epsilon)),$$

then $\lim_{\xi \to 0} \lim_{\omega \to 0} \lim_{K \to \infty} \lim_{\epsilon \to 0} \lim_{N \to \infty} UB(N, K, \xi, \omega, \eta; \epsilon) = 1$; note that this is a uniform bound over all i.

Then, letting T denote a set of types that is an element of $C(N, K, \xi, \omega, \eta; \epsilon)$, we have

$$\begin{aligned} & Pr\left(m_{\sigma_N^*}(t) \in [\bigcup_{C_N(K,\xi,\omega,\eta;\epsilon)} T] \middle| \min_k |u_{\sigma_N^*}(t) - \theta_k| > \eta \right) \\ & \geq \sum_{i=1}^{I} \Bigg[Pr\left(t \in D_i(N, K, \xi, \omega, \eta; \epsilon) | \min_k |u_{\sigma_N^*}(t) - \theta_k| > \eta \right) \\ & \qquad \cdot Pr(m_{\sigma_N^*}(t) \in C_i(N, K, \xi, \omega, \eta; \epsilon) | t \in D_i(N, K, \xi, \omega, \eta; \epsilon)) \\ & \quad + Pr(m_{\sigma_N^*}(t) \in C_i(N, K, \xi, \omega, \eta; \epsilon) \Big| \min_k |u_{\sigma_N^*}(t) - \theta_k| > \eta) \Bigg] \\ & \geq \frac{Pr(t \in \bigcup_i D_i(N, K, \xi, \omega, \eta; \epsilon)) LB(N, K, \xi, \omega, \eta; \epsilon) + Pr(t \in \bigcup_i C_i(N, K, \xi, \omega, \eta; \epsilon))}{Pr(\min_k |u_{\sigma_N^*}(t) - \theta_k| > \eta)}. \end{aligned} \tag{26}$$

Since $\lim_{\xi \to 0} \lim_{\omega \to 0} \lim_{K \to \infty} \lim_{\epsilon \to 0} \lim_{N \to \infty} PR(t \in \bigcup_i [D_i(N, K, \xi, \omega, \eta; \epsilon) \bigcup D_i(N, K, \xi, \omega, \eta; \epsilon)]) = Pr(\min_k |u_{\sigma_N^*}(t) - \theta_k > \eta)$, we have

$$\begin{aligned} & \lim_{\xi \to 0} \lim_{\omega \to 0} \lim_{K \to \infty} \lim_{\epsilon \to 0} \lim_{N \to \infty} \\ & \quad \frac{Pr(t \in \bigcup_i D_i(N, K, \xi, \omega, \eta; \epsilon)) LB(N, K, \xi, \omega, \eta; \epsilon) + Pr(t \in \bigcup_i C_i(N, K, \xi, \omega, \eta; \epsilon))}{Pr(\min_k |u_{\sigma_N^*}(t) - \theta_k| > \eta)} \\ & = 1 \end{aligned} \tag{27}$$

for all η, thus proving the proposition.

□