Discussion Paper Series – CRC TR 224

Discussion Paper No. 443
Project B 02

Persuasion With Limited Data: A Case-Based Approach

Shiri Alon [1]
Sarah Auster [2]
Gabi Gayer [3]
Stefania Minardi [4]

July 2023

[1] Department of Economics, Bar-Ilan University, Email: shiri.alon-eron@biu.ac.il
[2] Department of Economics, University of Bonn, Email: sauster@uni-bonn.de
[3] Department of Economics, Bar-Ilan University, Email: gabriella.gayer@biu.ac.il
[4] Department of Economics and Decision Sciences, HEC Paris, Email: minardi@hec.fr

Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Collaborative Research Center Transregio 224 - www.crctr224.de
Rheinische Friedrich-Wilhelms-Universität Bonn - Universität Mannheim

Persuasion with Limited Data: A Case-Based Approach*

Shiri Alon, Sarah Auster, Gabi Gayer, and Stefania Minardi†

July 6, 2023

Abstract

A strategic sender collects data with the goal of persuading a receiver to adopt a new action. The receiver assesses the profitability of adopting the action by following a classical statistics approach: she forms an estimate via the similarity-weighted empirical frequencies of outcomes in past cases, sharing some attributes with the problem at hand. The sender has control over the characteristics of the sampled cases and discloses the outcomes of his study truthfully. We characterize the sender's optimal sampling strategy as the outcome of a greedy algorithm. The sender provides more relevant data—consisting of observations sharing relatively more characteristics with the current problem—when the sampling capacity is low, when a large amount of initial public data is available, and when the estimated benefit of adoption according to this public data is low. Competition between senders curbs incentives for biasing the receiver's estimate and leads to more balanced datasets.

Keywords: Persuasion, case-based inference, similarity-weighted frequencies.
JEL Classification Numbers: D81, D83

*We thank Itai Arieli, Florian Brandl, Gregorio Curello, Drew Fudenberg, Itzhak Gilboa, Sebastian Kodritsch, Marco Li Calzi, Ludvig Sinander, Ina Taneva, Tymon Tatur, and Roberto Weber for valuable comments. Sarah Auster gratefully acknowledges funding from the German Research Foundation (DFG) through Germany's Excellence Strategy—EXC 2126/1—390838866 and CRC TR 224 (Project B02). Stefania Minardi gratefully acknowledges support from the French National Research Agency (ANR), "Investissements d'Avenir" (LabEx Ecodec/ANR-11-LABX-0047).

†Alon: Department of Economics, Bar-Ilan University (email: shiri.alon-eron@biu.ac.il); Auster: Department of Economics, University of Bonn (email: sauster@uni-bonn.de); Gayer: Department of Economics, Bar-Ilan University (email: gabriella.gayer@biu.ac.il); Minardi: Department of Economics and Decision Sciences, HEC Paris (email: minardi@hec.fr)

1 Introduction

Economic agents often face unfamiliar problems and have to make decisions based on a limited amount of data. Examples include policymakers deciding on a new law, health authorities deciding on the approval of a new medicine or vaccine, or a jury coming to a verdict in a criminal case. The canonical approach in economics to such decision problems relies on the Bayesian paradigm, postulating that agents hold prior beliefs and maximize expected utility with respect to these beliefs. The implicit assumption is that agents have sufficient information about the underlying data-generating process, for instance, because they face the same problem repeatedly, they have access to historical data, or they rely on theories regarding the distributions of outcomes.

Forming a prior belief is more intricate when agents face unfamiliar and complex problems with limited data or experience. Following the classical approach to statistical inference, Gilboa and Schmeidler (1995) propose a case-based approach to decision-making in such problems. According to this theory, agents engage in backward-looking reasoning by explicitly relying on past evidence. Facing a new decision problem, they form estimates of potential contingencies by drawing analogies with similar problems encountered in the past, even if those are not identical to the problem at hand. The similarity notion plays a key role: in the absence of data about the current problem, agents seek guidance by examining the performance of actions taken in previous situations that share some attributes with the problem they face. The collection of past cases used for the inference—the database—can include an agent's own experiences as well as external, observable past events.

This paper studies the question of how a decision maker's estimation in such environments can be influenced when external data are provided by a strategic party. This party might, for instance, be a political advisor trying to convince a policymaker, a pharmaceutical company wanting to get a drug approved, or a prosecutor seeking to have a suspect convicted. More specifically, we consider the following design problem: a sender (he) chooses the composition and size of a database with the aim of convincing a receiver (she) to adopt a certain action. We think of the problem under consideration as being characterized by a list of attributes. The receiver estimates the probability of the potential outcomes of an action—assumed to be success or failure—by weighing the frequencies of outcomes in the available data based on their similarity to the current problem. There are some (limited) public data on which both the receiver and the sender base their initial assessment of the action. Given this baseline, the sender seeks to boost the image of the action in the eyes of the receiver by providing additional evidence. Evidence comes in the form of a collection of cases, which we call a database. A typical case in the database is described by its characteristics (referred to as the case type), determining the similarity to the current problem and its realized outcome. The sender has the freedom to choose the size of the database and the case-type composition within the limits of a sampling capacity and physical constraints on the ability to generate observations of the same case type. The sender is committed to disclosing the outcomes of his study truthfully to the receiver.

Formally, our sender chooses an experiment inducing a distribution over posterior beliefs, as in the canonical Bayesian persuasion problem (Kamenica and Gentzkow, 2011). We depart from this literature in three key dimensions. First, we model the receiver as a frequentist rather than a Bayesian. In particular, her statistical inference process is based on similarity-weighted empirical frequencies, as modeled by Billot, Gilboa, Samet, and Schmeidler (2005). Under this criterion, incorporating cases that are not identical to the current case introduces a bias in the belief formation process. As shown later, this aspect is exploited by the sender in our setting to influence the receiver's evaluation of the action. Second, for most of the analysis, we assume that the sender's payoff is linearly increasing in the receiver's final estimate. The sender thus seeks to maximize the receiver's expected revised belief. While this is the standard way to model image or reputation in economics (Holmström, 1999), the assumption trivializes the Bayesian persuasion problem and thus allows us to draw a sharp distinction between that framework and ours. Third, the sender in our setting chooses from a set of databases consisting of a finite collection of case types, which imposes natural restrictions on the set of distributions over revised beliefs that the sender can induce. Instead, the defining feature of the Bayesian persuasion literature is maximal flexibility in the choice of information structures. While this approach is mathematically elegant, it operates on a high level of abstraction. Relative to this benchmark, the set of feasible information structures in our framework is more tangible, allowing for a concrete understanding of the signal characteristics' influence on the receiver's estimate.

In our setting, the sender's key tradeoff is between selecting case types that are promising in terms of probability of success versus selecting those that are sufficiently similar to the current problem. Mathematically, the sender's choice problem amounts to maximizing a non-linear function over sets of integer-valued alternatives, a complex combinatorial task. Leveraging the structure of our specific objective function, we present a greedy search algorithm that solves the sender's problem in a polynomial number of steps. Starting from the empty database, the algorithm replaces in each stage either an unfilled slot or an existing case type with a more similar case type so as to maximize a replacement index, which is defined over the set of all pairs of case types.

Our main result shows that the sender's optimal case-type collection is characterized as the algorithm's outcome. We thus provide a simple, tractable method to find the optimal sampling strategy for the described problem. We use this result to study the properties of the optimal case-type composition. In particular, we show that the sampled case types have a higher similarity to the current problem when there are more relevant (or simply more) public data available and when, according to the public data, the initial chances of success in the current case are smaller. Indeed, a substantial amount of initial public evidence makes it more difficult to change the receiver's view about the right course of action and hence prompts the sender to sample relatively more similar and less promising case types. A higher success probability in the current case has the opposite effect: since the sender is only willing to sample case types that increase the receiver's expected estimate, he will favor case types that have a higher success probability over those that share more characteristics with the current problem.

Next, we consider the special case where the sender faces no constraints on how many instances of the same case type he can sample. We show that, as long as there are some case types with a higher success probability than the current case, the sender exhausts his sampling capacity and generates a database that is entirely homogeneous: all cases in the database have identical attributes. The choice of the optimal case type is described by the maximization of a simple success-similarity index. We then show that the sampled case type will be less similar to the case at hand when the sender's sampling capacity increases. A larger sampling capacity weakens the impact of the initial evidence and thus leads the sender to sacrifice similarity in favor of case types that are more promising in terms of their probability of success. In the limit, as the capacity tends to infinity, similarity loses its relevance, and the sender's sampling decision is driven solely by the case types' likelihood of generating favorable outcomes.

In many contexts, agents obtain information from multiple experts with heterogeneous preferences. In the last part of the paper, we ask whether competition among data providers leads to more relevant data. We address this question by extending the baseline setup to allow for competition between two senders. We model the strategic interaction between the two senders as a zero-sum game by assuming that the senders have opposing state-independent preferences. The first sender seeks to maximize the receiver's estimate of the action generating success, while the second sender seeks to minimize it. We show that a pure-strategy equilibrium exists and provide conditions under which the presence of the competitor causes senders to sample case types that are more similar to the current case than the ones they would sample in isolation. Under these conditions, competition helps the receiver on two dimensions: 1) she receives more data in total; 2) both senders are incentivized to provide data on cases that are more relevant to the current problem. Competition thus improves data on the extensive and intensive margins.

We illustrate the setting and the main results with two examples.

Example 1. The receiver is a policymaker trying to assess the likelihood that a given reform is going to increase the vote share. The more confident the policymaker is in the success of the reform, the more eager she will work to support and, ultimately, implement it. The sender is a lobbyist or think tank aiming to sway the policymaker towards implementing the reform by conducting a study and presenting the results to the policymaker. The think tank has some freedom in shaping the study but cannot tamper with the evidence once it has been conducted. For example, the study might collect information about the performance of related reforms. In this context, a case type describes the attributes of the past reform as well as the social and economic circumstances in which it was implemented. The composition of case types can be affected, for instance, by restricting attention to particular regions or time frames. Alternatively, the study could take the form of a voter survey, with case types describing the attributes of the respondents in the survey. The think tank can choose how the survey is conducted, e.g., by phone or online, thereby affecting the composition of voters reached by the survey. Our results suggest that the think tank will try to target the study to specific demographics or regions that are promising in terms of generating favorable outcomes but are similar enough to the current circumstances so that the policymaker deems the evidence sufficiently relevant.

Example 2. The receiver is a jury member deliberating how to vote on the outcome of a criminal trial. There are two senders, the prosecutor and the defender, both of whom can call on different witnesses to take the stand. The more the jury member is convinced of the guilt or innocence of the accused, the harder she will argue to convict or acquit the accused in jury deliberation. Whether a given witness ultimately helps the prosecutor's case or the defender's case is not known with certainty beforehand. Moreover, some witnesses will have a bigger impact on the jury member's assessment than others. A case type may thus be viewed as the attributes of the witness's relation to the accused and his/her potential knowledge about the case. Based on these attributes, the prosecutor and defender must decide which group of witnesses to call for testimony. For instance, they could call a number of forensic experts, a number of character witnesses, or a combination of them. Our results suggest that, while the prosecutor and the defender will try to influence the jury member's estimate by targeting the witness selection, competition drives both sides to call on witnesses relatively more relevant to the case.

The rest of the paper is organized as follows. We conclude this section with a literature overview. Section 2 introduces our baseline setup with one sender and one receiver. Section 3 introduces the greedy search algorithm and shows that it solves the sender's design problem. Section 4 extends the baseline model to the case of two competing senders. Section 5 concludes. All proofs are contained in the Appendix.

1.1 Related literature

A rapidly expanding literature on persuasion has emerged from the seminal paper of Kamenica and Gentzkow (2011). Promising avenues for further development in this area arise from relaxing two crucial assumptions: the Bayesian formation of beliefs by the receiver and the unconstrained nature of information structures available to the sender. As recently emphasized by Kamenica, Kim, and Zapechelnyuk (2021), these assumptions are demanding and limit the applicability of the theory. In fact, several recent studies introduce non-Bayesian elements into the traditional framework. For instance, Galperti (2019) allows the receiver to question, and sometimes modify, her prior belief if she is successfully persuaded by new evidence that contradicts her initial worldview; de Clippel and Zhang (2022) analyze situations where the receiver makes systematic mistakes in updating beliefs and study the impact of such biases on optimal persuasion within a broad class of non-Bayesian updating rules. In Eliaz, Spiegler, and Thysen (2021), the receiver lacks an understanding of the statistical mapping between states and signals, creating an opportunity for the sender to influence the receiver's beliefs by strategically choosing not only the signal but also a selective interpretation of it. Dworczak and Pavan (2022) as well as Kosterina (2022) adopt a robust approach to information design to study a sender who is uncertain about the receiver's learning environment.[1] Compared with the above papers, we take a different departure from the Bayesian approach by introducing classical statistics inference based on similarity-weighted frequencies.

Closer to our work, Patil and Salant (2023) study the role of statistical inference on optimal persuasion in a setting where the sender provides evidence in the form of Bernoulli experiments. The key difference with our framework is that in Patil and Salant (2023), observations are of only one type. The paper's focus is on the question of *how many* observations the sender will sample, while our main interest lies in the question of *which kind* of data the sender will provide. A shared feature of their setting and ours is that restrictions on the set of feasible information structures naturally arise from the discreteness of the data. Such restrictions also appear in dynamic persuasion settings, where the sender decides how much information to collect for a Bayesian receiver, as in Brocas and Carrillo (2007), Henry and Ottaviani (2019), and McClellan (2021). In contrast, several recent papers have introduced explicit constraints on the set of information structures to study their implications on static Bayesian persuasion, as in Perez-Richet (2014), Le Treust and Tomala (2019), Di Tillio, Ottaviani, and Sørensen (2021), and Ball and Espín-Sánchez (2022).

Our paper is further related to Glazer and Rubinstein (2004, 2006). In a disclosure setting with verifiable information, they consider a speaker (i.e., the sender) who provides arguments to convince a listener (i.e., the receiver) to take a specific action under the assumption that the amount of evidence that either the speaker can provide or the listener can verify is limited. In related work, Glazer and Rubinstein (2001) study optimal persuasion rules in the context of debates between two speakers who have opposing preferences and compete to persuade a listener to take their respective preferred action. This work recognizes the relevance of certain aspects extraneous to the Bayesian logic for persuasion to be successful. For instance, the success of a counterargument may depend not only on the strength of the evidence presented but also on its similarity to the rival argument. A key distinction is that in these works the sender has full information

[1] Another direction that extends the Bayesian approach studies persuasion problems with ambiguity-averse agents, as in Laclau and Renou (2017), Beauchêne et al. (2019), Cheng (2021), and Liu and Yannelis (2021). These works generalize the notion of Bayes plausibility to explore the role of ambiguous communication.

about the state of the world and, hence, knows which action the receiver should choose.

Finally, our paper relates to a small literature that analyzes the behavior of case-based agents in applications other than persuasion/communication. For instance, Blonski (1999) analyzes social learning of case-based agents in networks, Guerdjikova (2006) and Golosnoy and Okhrin (2008) investigate the portfolio choices of case-based agents in financial markets, and Argenziano and Gilboa (2019) study coordination games with case-based players.

2 The Model

Prediction problem. We consider a receiver who needs to make a prediction about the outcome of some action a. The action may result in either a good outcome (denoted by G) or a bad outcome (denoted by B); hence, the state space is given by $\Omega = \{G, B\}$. The receiver uses a similarity-weighted frequency approach and assesses the consequences of action a based on the observations of past data. An observation is characterized by a list of observable attributes that are believed to affect the probability of the outcome. Such a list of attributes is called a *case type*, with $\mathcal{C}$ denoting the finite set of all case types. An observation is a pair (c, y), where $c \in \mathcal{C}$ is the observed case type, and $y \in \{G, B\}$ is the *outcome*. A *database* $\mathbf{D}$ is a finite collection of observations, modeled as a counter vector

$$\mathbf{D} : \mathcal{C} \times \{G, B\} \to \mathbb{Z}_+,$$

where $\mathbf{D}(c, y)$ counts how many observations (c, y) appear in the database. For two databases $\mathbf{D}$ and $\mathbf{D}'$, their sum $\mathbf{D} + \mathbf{D}'$ is the result of pointwise adding $\mathbf{D}(c, y) + \mathbf{D}'(c, y)$ for every (c, y), such that the new database contains the observations from both databases. The current problem, indexed by 0, is characterized by a case type c_0. The outcome y_0 associated with c_0 is unknown, as a was not yet taken in the current problem.

Following Billot et al. (2005), we assume that, given database $\mathbf{D}$, the receiver's estimate of action a generating a favorable outcome in the current problem is given by the similarity-weighted empirical frequency of outcome G in database $\mathbf{D}$. The underlying assumption is that when two cases are similar, their outcome distribution is similar as well. The similarity between any two case types is quantified by the similarity function $s : \mathcal{C} \times \mathcal{C} \to \mathbb{R}_{++}$.[2] We normalize the similarity function

[2]Throughout our analysis, we assume that similarity values are predetermined and fixed,

so that $s(c_0, c_0) = 1$ and assume that $s(c_0, c) \in (0, 1)$ for all $c \in \mathcal{C}$ with $c \neq c_0$; that is, c_0 is most similar to itself. Given database $\mathbf{D}$, the receiver's estimate that action a be successful in the current problem is given by

$$P(y_0 = G|\mathbf{D}) = \frac{\sum_{c \in \mathcal{C}} s(c_0, c)\mathbf{D}(c, G)}{\sum_{(c,y) \in \mathcal{C} \times \{G,B\}} s(c_0, c)\mathbf{D}(c, y)} \quad . \tag{1}$$

To illustrate the concept of case types and the determination of similarities, let us return to the example of a policymaker deciding on a new reform. The policymaker tries to assess the probability of success of the reform by examining previous cases of similar reforms that have been implemented in other countries or constituencies. A case type in this context contains various factors such as the essential elements of the reform, the characteristics of the electorate, and the economic conditions under which it was implemented. The outcome states whether the policy resulted in a favorable or unfavorable result. The policymaker maintains the belief that past reforms conducted under similar circumstances hold greater significance for predicting the outcome in the current case compared to those implemented under very different conditions. As a result, the policymaker assigns a higher weight to cases with higher similarity values. The similarity $s(c_0, c)$, quantifies the extent to which the features of the reform, the electorate, and the economic circumstances in the current problem resemble those in case type c.

Strategic Data Provision. A sender aims to provide verifiable data that will boost the action's image in the eyes of the receiver. As in the reputation literature (Holmström, 1999), we assume that the sender's payoff is linearly increasing in the receiver's revised belief on action a generating success in the current problem. From a modeling perspective, the assumption is made for analytical tractability, but also as a way to sharply distinguish the effects stemming from case-based inference from the usual (Bayesian) persuasion forces, as we explain below.[3]

The sender can influence the receiver's estimate by providing data. We assume that the sender chooses the case types in the sample without knowing the associated outcomes. Moreover, once the test is concluded, the sender is obliged

based on the notion that similarities are more basic than probabilities and can be used to form probabilities. This assumption applies when the receiver has prior knowledge or experience with the attributes influencing similarity between cases, while it may be violated if the receiver is still in the process of learning similarity values from data, referred to as second-order induction by Gilboa and Schmeidler (2001, pp. 174–183).

[3]See also the discussion in Section 3.2.

to submit all test results truthfully. When choosing the case types that will be sampled, the sender faces two types of constraints: 1) an overall capacity constraint N on the number of sampled cases, and 2) for each case type $c \in \mathcal{C}$, a type-specific constraint N_c, capturing the maximal number of type-c observations that can enter the database. For instance, each case type might be unique, in which case $N_{c_0} = 0$ and $N_c = 1$ for all $c \neq c_0$. Alternatively, there might be some regulatory restriction, which imposes a uniform bound $N_c = \bar{n} < N$ on the number of identical case types in a trial.

Both the sender and the receiver have access to some initial public database $\mathbf{H}$. The probability of action a generating outcome G based on this prior data is $p_0 := P(y = G|\mathbf{H})$, calculated according to formula (1). Once additional data are provided, the receiver revises her estimate by adding the new data $\mathbf{D}$ to the initial database $\mathbf{H}$. The revised probability of success is

$$P(y_0 = G|\mathbf{D} + \mathbf{H}) = \frac{\sum_{c\in\mathcal{C}} s(c_0, c)[\mathbf{D}(c, G) + \mathbf{H}(c, G)]}{\sum_{(c,y)\in\mathcal{C}\times\{G,B\}} s(c_0, c)[\mathbf{D}(c, y) + \mathbf{H}(c, y)]} \quad .$$

Optimization problem. The sender chooses a collection of case types that will be sampled, described as a counter vector $D : \mathcal{C} \to \mathbb{Z}_+$. For each $c \in \mathcal{C}$, $D(c)$ represents the number of cases of type c that appear in the collection. Given the selected collection D, the revised probability $\hat{p}_0(D)$ is a random variable whose realized value depends on the outcomes in the sampled cases. The sender's objective is to maximize the receiver's expected belief. We assume that, just like the receiver, the sender bases his initial assessment of the likelihood that outcome G will be realized in the different types of cases on the public database $\mathbf{H}$. Hence, the sender's estimate of case type $c \in \mathcal{C}$ generating outcome G is the similarity-weighted frequency[4]

$$p_c := \frac{\sum_{c'\in\mathcal{C}} s(c, c')\mathbf{H}(c', G)}{\sum_{(c',y)\in\mathcal{C}\times\{G,B\}} s(c, c')\mathbf{H}(c', y)} \quad .$$

For ease of notation, we represent the collection of case types in the public data by the counter function $H : \mathcal{C} \to \mathbb{Z}_+$ such that $H(c) := \mathbf{H}(c, G) + \mathbf{H}(c, B)$. Given a collection of case types D, the expected revised probability of success with respect

[4]Our assumption that sender and receiver use the same similarity function for statistical inference is akin to the common prior assumption in the standard setting. As shown by Alonso and Câmara (2016), heterogeneous prior beliefs give rise to interesting implications, from which, however, we want to abstract in the current paper.

to $\mathbf{H}$ is then given by

$$\hat{p}_0(D) = \frac{\sum_{c\in\mathcal{C}} s(c_0,c)\,[D(c)p_c + \mathbf{H}(c,G)]}{\sum_{c\in\mathcal{C}} s(c_0,c)[D(c)+H(c)]} \quad .$$

The sender solves the constrained optimization problem

$$\max_D \hat{p}_0(D) \tag{2}$$

subject to

$$D(c) \leq N_c, \forall c \in \mathcal{C} \tag{3}$$
$$\sum_{c\in\mathcal{C}} D(c) \leq N. \tag{4}$$

If agents were Bayesian with a common prior, then the law of iterated expectations would imply that the expected posterior would satisfy $\hat{p}_0(D) = p_0$. Consequently, any sampling strategy would yield the same expected posterior, making all strategies equally optimal. By contrast, in the current framework, the sender can generate an upward bias in the receiver's expected belief and thereby profit from sampling. This bias is facilitated by the fact that the receiver is willing to assign a positive weight to cases that are not identical to the current case.

A similar issue appears in kernel estimation, a non-parametric statistical technique that is used to estimate the conditional expectation of a dependent variable given the values of the independent variables (see Akaike, 1954, Rosenblatt, 1956, Parzen, 1962, Silverman, 1986, among others). The estimator in kernel estimation is a weighted average of the dependent variables in the sample. In order to reduce the variance of the estimator in small samples, the kernel places weight on distant observations, which inevitably yields a biased estimator. This property, known in statistics as the bias-variance trade-off, applies equally to case-based inference.[5]

[5] To ensure asymptotic convergence of the kernel estimator, the relative weights placed on the different observations in the sample generally depend on the size of the sample under this method. Analogously, one could extend our setting to let the receiver's similarity function adapt when the receiver has access to more data (see Gayer, 2010, for an example). The sender would then need to evaluate how the receiver's similarity values evolve and this might affect his disclosure incentives. For instance, if the relative weights placed on distant case types decrease with the number of observations, the sender's incentives to sample such types might be reduced. That said, our main focus is on situations where the sender's sampling capacity is "small", so fixing the similarity function is indeed plausible or, at the very least, serves as a good approximation.

3 Optimal Data Collection

When determining the optimal collection of case types in the sample, the sender faces a tradeoff between selecting case types that have a high likelihood of success and case types that are sufficiently similar to the current problem in order to significantly influence the receiver's beliefs. The first aspect is captured by the success probability p_c derived from the initial data $\mathbf{H}$, while the latter aspect is measured by the similarity $s(c_0, c)$.

The sender, having to choose integer quantities of case types to maximize the receiver's expected estimate, faces an integer programming optimization problem with a non-linear objective function. We will show that the solution to this combinatorial choice problem can be found via an algorithm that makes a greedy choice according to a criterion that we introduce next. Generally, a greedy algorithm selects the best option based on the current situation without taking into account future choices. In our case, the choice of the algorithm at each stage will be guided by a replacement index defined over the set of all pairs of case types. This index will serve as a measure of the benefit associated with replacing one case type with another in a given database.

Definition 1. *For any two case types $c, c' \in \mathcal{C}$, define the replacement index*

$$\phi(c', c) := \frac{s(c_0, c')p'_c - s(c_0, c)p_c}{s(c_0, c') - s(c_0, c)}.$$

The following lemma shows that the replacement index allows us to determine whether, for a given collection of case types, the sender gains or loses from replacing some case type c belonging to the collection with another case type c'.

Lemma 1. *Let D be a case-type collection and $c \in \mathcal{C}$ a case type such that $D(c) > 0$. Define D' as the case-type collection obtained from D by replacing one instance of c with $c' \in \mathcal{C}$.*[6]

(i) If $s(c_0, c') > s(c_0, c)$, then:

$$\hat{p}_0(D') > \hat{p}_0(D) \quad \Longleftrightarrow \quad \phi(c', c) > \hat{p}_0(D).$$

(ii) If $s(c_0, c') < s(c_0, c)$, then:

$$\hat{p}_0(D') > \hat{p}_0(D) \quad \Longleftrightarrow \quad \phi(c', c) < \hat{p}_0(D).$$

[6]Formally: $D'(c) = D(c) - 1$, $D'(c') = D(c') + 1$ and $D'(c'') = D(c'')$ for all $c'' \neq c, c'$.

Lemma 1 provides a simple condition to detect when the sender gains from replacing a certain case type from a given collection. Specifically, suppose that the sender has a collection of case types D and wishes to evaluate the impact on the expected posterior of replacing one instance of case type c represented in D with a more similar case type c'. Part (i) of Lemma 1 guarantees that such a replacement is beneficial if and only if $\phi(c', c) > \hat{p}_0(D)$. Symmetrically, part (ii) characterizes the condition under which replacing an instance of case type c with a less similar case type c' makes the expected posterior increase.

To account for adding a case type c to a database with unfilled capacity, it is useful to introduce a fictitious case type $c_\emptyset$ with $s(c_0, c_\emptyset) = 0$ and $p_{c_\emptyset}$ an arbitrary number. This case type can be thought of as 'no observation' or an observation that is not relevant to the problem at hand.[7] Lemma 1 applies to replacements of $c_\emptyset$ as well: replacing 'no observation' with an observation of case type c in some case-type collection D generates a gain for the sender if and only if $\phi(c, c_\emptyset) = p_c > \hat{p}_0(D)$.[8] For ease of notation, we will treat $\mathcal{C}$ as the set of case types including the fictitious case type $c_\emptyset$. Since not sampling is always feasible, we set $N_{c_\emptyset} \geq N$.

Note that Lemma 1 implies that sampling is beneficial for the sender as long as there exists some case type $c \in \mathcal{C}$ such that $p_c > p_0$. To see this, consider the collection consisting only of instances of the fictitious case type $c_\emptyset$—call it D_0. This can be interpreted as the initial situation where the sender has not sampled yet. Since $\hat{p}_0(D_0) = p_0$, Lemma 1 implies that the sender gains from sampling one instance of case types c if and only if $p_c = \phi(c, c_\emptyset) > \hat{p}_0(D_0) = p_0$. Thus, sampling a case type $c \in \mathcal{C}$ increases the receiver's posterior in expectation if and only if it has a higher success probability than the current case type c_0 according to the initial data $\mathbf{H}$ and, hence, if and only if $p_c > p_0$.

The algorithm we now specify presents a sequence of 'locally beneficial' replacements from less similar to more similar case types. Starting from the empty database, the algorithm will choose at every stage a replacement of one item in the collection so as to maximize the replacement index. To formalize this procedure, let us define for each case-type collection D, the *set of feasible replacements* of less

[7]Note that for all non-fictitious case types, we require similarities to be strictly positive, in line with Billot et al. (2005). It is not difficult to add the possibility of irrelevant case types that obtain a similarity value of zero; see Alon and Gayer (2020) for a characterization.

[8]Likewise, removing a case type c from a case-type collection D without replacement of another case type in $\mathcal{C}$ is strictly beneficial if and only if $p_c < \hat{p}_0(D)$.

similar case types by more similar case types:

$$R(D) = \{(c', c) \in \mathcal{C} \times \mathcal{C} : s(c_0, c') > s(c_0, c), D(c) > 0 \text{ and } D(c') < N_{c'}\}$$

In the definition and the analysis that follows, we will abstract from the non-generic cases where $\phi(c, c') = \phi(c'', c''')$ for some $(c, c') \neq (c'', c''')$.

Definition 2 (Greedy Search Algorithm (GSA))**.** *Let D_0 be such that $D_0(c_\emptyset) = N$ and $D_0(c) = 0$ for all $c \in \mathcal{C}\backslash\{c_\emptyset\}$. For each $n \in \mathbb{N}$:*

1. *Find $(\tilde{c}', \tilde{c}) \in R(D_n)$ such that $(\tilde{c}', \tilde{c}) \in \arg\max_{(c',c)\in R(D_n)} \phi(c', c)$.*
2. *If $\phi(\tilde{c}', \tilde{c}) \leq \hat{p}_0(D_n)$, then stop.*
3. *If $\phi(\tilde{c}', \tilde{c}) > \hat{p}_0(D_n)$, let D_{n+1} be the case-type collection obtained from D_n by replacing case type $\tilde{c}$ with case type $\tilde{c}'$.*

The algorithm begins at stage 0 with a collection of N instances of the fictitious case type. Then, each stage $n \in \mathbb{N}$ consists of three steps. Given collection D_n, step 1 computes the index-maximizing replacement from less similar to more similar case types subject to the feasibility constraint. Steps 2 and 3 compare the value of the replacement index obtained in step 1 with the expected posterior of D_n. By part (i) of Lemma 1 we know that the replacement is beneficial if and only if the replacement index exceeds the expected posterior. Thus, the algorithm stops at step 2 if this is not the case. Otherwise, we reach step 3, which defines the collection D_{n+1} by implementing the index-maximizing replacement. Since, in each stage, we replace a less similar case type with a more similar case type, every stage of the algorithm contributes to the total similarity of the sampled collection. In other words, the degree of the total similarity to the current case type c_0 increases with the number of stages through which the algorithm runs. We denote by n^* the stage at which the algorithm stops.

We illustrate the algorithm with an example.

Example. There are three possible case types, $\mathcal{C} = \{c, c', c''\}$ with $p_0 < p_c < p_{c'} < p_{c''}$, $s(c_0, c) > s(c_0, c') > s(c_0, c'')$, and

$$\phi(c, c') < \phi(c, c'') < \phi(c', c'').$$

The sender can only sample two observations ($N = 2$), of which at most one can be of case type c''. Hence, $N_{c''} = 1$, and $N_c, N_{c'} > 1$. The algorithm starts with

$D_0(c_\emptyset) = 2$ and $R(D_0) = \{(c, c_\emptyset), (c', c_\emptyset), (c'', c_\emptyset)\}$. The maximizing replacement index over this set is $\phi(c'', c_\emptyset) = p_{c''}$, so we get $D_1(c_\emptyset) = 1, D_1(c'') = 1$, and $R(D_1) = \{(c, c''), (c', c''), (c, c_\emptyset), (c', c_\emptyset)\}$. Case type c'' is now at capacity, so it cannot be added further. Since

$$\phi(c, c'') < \phi(c, c_\emptyset) = p_c, \quad \phi(c', c'') < \phi(c', c_\emptyset) = p_{c'}, \quad \text{and} \quad p_c < p_{c'},$$

the maximizing replacement index in the second stage is $\phi(c', c_\emptyset)$. Provided $\hat{p}_0(D_1) < p_{c'}$ is satisfied, we thus set $D_2(c') = 1, D_2(c'') = 1$. The set of feasible replacements is now given by $R(D_2) = \{(c, c'), (c, c''), (c', c'')\}$ and the maximal replacement index over this set is $\phi(c', c'')$. If $\hat{p}_0(D_2) > \phi(c', c'')$, we stop at D_2. Otherwise we set $D_3(c') = 2$, and so on.

It should be noted that, while the replacement index measures the qualitative effect of a given replacement on the receiver's expected estimate, it does not quantify the magnitude of the change, as this will depend on the public data $\mathbf{H}$ and the other sampled cases in D. Hence, given some case-type collection D, the replacement associated with the highest index is not necessarily the replacement that locally benefits the sender most. Instead, the algorithm proceeds through a sequence of incremental improvements, whose order only depends on the replacement indices $(\phi(c', c))_{(c',c)\in\mathcal{C}\times\mathcal{C}}$. The following result shows that this procedure finds the optimal case-type collection for the sender.

Theorem 1. *The case-type collection D_{n^*}, obtained as the outcome of the GSA, solves the sender's optimization problem* (2).

To show the optimality of D_{n^*}, we start by establishing two preliminary results. First, we demonstrate how the similarity of the different case types with c_0 affects the order of their associated replacement indices (Lemma 4 in the Appendix). In particular, considering three case types c_1, c_2, c_3, with c_1 being most similar to c_0 and c_3 being least similar to c_0, we show that either $\phi(c_1, c_2) < \phi(c_1, c_3) < \phi(c_2, c_3)$ or $\phi(c_2, c_3) < \phi(c_1, c_3) < \phi(c_1, c_2)$ will hold. This ordering has implications for the properties of the sets of feasible replacements $R(D_n)$ at different stages n of the algorithm. For instance, if there is a stage n at which the algorithm replaces case type c_3 with case type c_1, it must be the case that either c_2 is already at capacity ($D_n(c_2) = N_{c_2}$) or c_2 is not represented in the current collection ($D_n(c_2) = 0$). If not, replacing case type c_3 with c_2 and replacing case type c_2 with c_1 would be feasible at stage n, so $\phi(c_1, c_3)$ cannot be the maximizing index across all feasible replacements in $R(D_n)$.

Second, we prove that the maximizing replacement index is monotonically decreasing in the algorithm stages n (Lemma 5). This result is non-trivial since the feasibility set $R(D_n)$ changes with each stage n. We show that having two stages of the algorithm, n and $n+1$, such that the maximizing replacement index at stage n is strictly greater than the one at stage $n+1$ is incompatible with the properties of the sets of feasible replacements implied by Lemma 4.

With these preliminary results at hand, we prove that there is no single case type replacement for D_{n^*} that satisfies the constraints and increases the expected estimate of the receiver. For case types with a greater similarity than those in the final collection D_{n^*}, this conclusion follows directly from the specification of the algorithm. The main challenge lies in showing that replacing any case type belonging to D_{n^*} with another *less similar* case type is either unfeasible or unprofitable. Using the established order on the set indices and the monotonicity property of the maximizing replacement index, we prove that this requirement is indeed satisfied. Finally, we show that if a case-type collection cannot be improved upon by replacing a single case type, then it cannot be improved upon by the replacement of a sub-collection of case types (Lemma 6).

As mentioned earlier, due to the integer constraints, the sender's design problem is a complex combinatorial task. In principle, to solve the problem, one would have to calculate the values of all possible case-type collections, the number of which grows exponentially with the cardinality of $\mathcal{C}$ and N. In contrast, the GSA finds the optimal case-type collection in a number of steps that is polynomial in $|\mathcal{C}|$ and N. Starting from the empty case-type collection, each available slot is replaced at most $|\mathcal{C}|-1$ times. Hence, the maximal number of algorithm stages is $N \cdot (|\mathcal{C}|-1)$. The complexity within each stage is bounded by the total number of indices that need to be compared. This number is given by $\binom{|\mathcal{C}|}{2} = \frac{|\mathcal{C}|(|\mathcal{C}|-1)}{2}$. Hence, the total number of steps of the algorithm is $O(N \cdot |\mathcal{C}|^3)$.

Properties of D_{n^*}. The characterization of the optimal case-type collection in terms of the outcome of the GSA allows us to study the influence of the initial data on the sender's sampling choice. The next proposition shows how the optimal case-type collection D_{n^*} depends on the initial success probability p_0 as well as the total similarity of the initial public data to the current case type, $s_{0H} := \sum_{c \in \mathcal{C}} s(c_0, c)H(c)$.

Proposition 2. *Let $D^*(\cdot \mid p_0, s_{0H})$ denote the solution of the GSA for a given success probability p_0 and similarity s_{0H}. The following statements hold:*

(i) If $s_{0H} > s'_{0H}$, then

$$\sum_{c \in \mathcal{C}} s(c_0, c) D^*(c \,|\, p_0, s_{0H}) \geq \sum_{c \in \mathcal{C}} s(c_0, c) D^*(c \,|\, p_0, s'_{0H}) \quad .$$

(ii) If $p_0 > p'_0$, then

$$\sum_{c \in \mathcal{C}} s(c_0, c) D^*(c \,|\, p_0, s_{0H}) \leq \sum_{c \in \mathcal{C}} s(c_0, c) D^*(c \,|\, p'_0, s_{0H}) \quad .$$

More relevant initial data, as captured by a larger s_{0H}, leads the sender to sample relatively less promising and more similar case types. Indeed, if s_{0H} is relatively large, it is difficult to change the receiver's perception of action a. Bringing to the fore good outcomes in cases that are very different from the current case will have a limited impact on the receiver's beliefs, so similarity becomes a more important factor. Hence, a substantial amount of relevant public data incentivizes the sender to sample more similar case types and thus reduces the bias in the receiver's inference. The GSA captures precisely this intuition. The similarity s_{0H} does not affect the sequence of maximizing replacement indices along the stages of the algorithm but does affect the speed with which the expected posterior $\hat{p}_0(D_n)$ increases with each stage n. The larger s_{0H}, the slower the increase of the expected posterior and, hence, the longer it takes for the algorithm to stop. Thus, as the total similarity of the initial data increases, the algorithm implements the same or strictly more replacements, thereby producing a collection of case types with a higher total similarity to c_0.

An increase in the success probability of the current case type p_0 has the opposite effect on the sender's incentives. If p_0 is relatively high, the sender chooses case types that are likely to have a favorable outcome, even if they share few attributes with the problem at hand. Recall that, by Lemma 1, the sender only gains from case types with $p_c > p_0$. The larger p_0, the more demanding this constraint becomes. Furthermore, the larger p_0, the larger the expected posterior $\hat{p}_0(D_n)$ for each intermediate case-type collection D_n. This implies that Condition 2. of Definition 2 becomes easier to satisfy. Consequently, the algorithm stops earlier, thereby producing a collection of case types that is less similar to c_0.

3.1 Maximal Flexibility

This section considers the special case where the sender faces no constraints on the case-type composition except for the overall sampling constraint. This flexibility allows us to characterize the optimal case-type collection in closed form and to study the implications of tightening/relaxing the sampling capacity. The following proposition characterizes the solution of the sender's optimization problem when the sender has the flexibility to sample any given case type up to N times.

Proposition 3. *Assume $N_c \geq N$ for all $c \in \mathcal{C}$ and define*

$$c^* := \arg\max_{c \in \mathcal{C}} \left(\frac{s(c_0, c)(p_c - p_0)}{N s(c_0, c) + \sum_{c' \in \mathcal{C}} s(c_0, c') H(c')} \right) . \quad (5)$$

The optimal case-type collection D_{n^*} is such that $D_{n^*}(c^*) = N$ and $D_{n^*}(c) = 0$ for all $c \neq c^*$.

A key feature of the environment with maximal flexibility is that the sender chooses a homogeneous database featuring just one case type. The sender's optimization problem thus has a corner solution. There are two situations to be distinguished. If the success probability of the current case type is higher than that of all others ($p_c \leq p_0, \forall c \in \mathcal{C} \backslash \{c_\emptyset\}$), then the sender has no incentives to sample. In this case, we have $c^* = c_\emptyset$.[9]

If instead there is a case type $c \in \mathcal{C}$ such that $p_c > p_0$, then the sender exhausts his sampling capacity, as testing any additional instance of case type c will increase the receiver's estimate in expectation. Proposition 3 demonstrates that the sender cannot gain from combining different case types in one database and that the case type he samples is determined as the maximizer of the *success-similarity index* in (5).[10] This index is increasing in both p_c and $s(c_0, c)$. Hence, a case type is more likely to be sampled if it has a high success probability according to the initial data and if it shares many attributes with the current case. The maximization problem is trivially solved if there exists a case type that dominates all others in terms of both success probability and similarity. Outside of this case, the sender will face a tradeoff along these two dimensions.

[9] Recall $s(c_0, c_\emptyset) = 0$.

[10] Note that the maximizer of the index in (5) is not necessarily unique. For instance, there could be two maximal case types, one having a relatively higher probability of success and the other being relatively more similar. In the absence of additional assumptions, the sender would be indifferent between these two case types. This case is clearly knife-edged and will thus be ignored.

The choice problem is illustrated in Figure 1. The black and gray dots illustrate different case types, identified by their similarity to the current case $s(c_0, c)$ and their success probability p_c. Gray dots represent case types that are strictly dominated by other case types: for each case type represented by a gray dot, there exists another case type, which has a higher similarity and a higher success probability according to the initial data $\mathbf{H}$. The case type maximizing the index (5) is marked by a red circle. The blue curve then shows all combinations of similarity and success probability that give rise to that same value of this index.

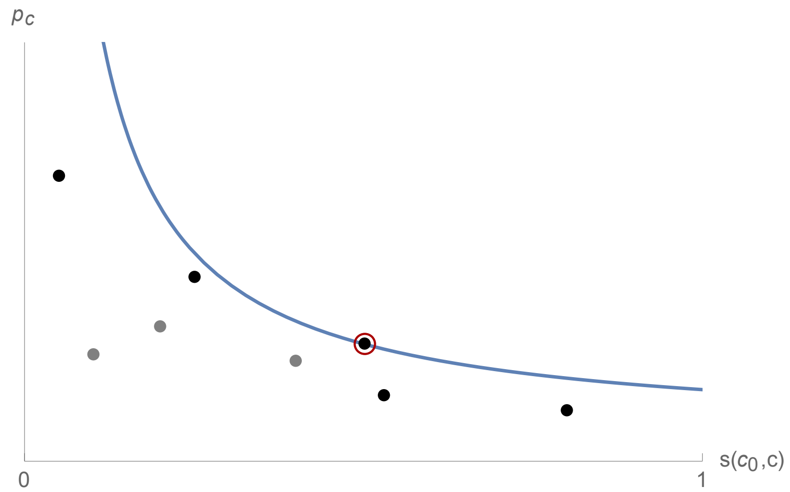

Figure 1: Different case types c in the $(s(c_0, c), p_c)$-space.

The value of the index depends on the sender's sampling capacity N, the size and relevance of the initial data, summarized by s_{0H}, and the success probability of the current case type p_0 according to the initial data. From Proposition 2, we know that the similarity of the optimal case type, $s(c_0, c^*)$, is weakly increasing in the similarity s_{0H} of the public data to the current case type and decreasing in the initial success probability p_0. The following proposition shows how the optimal case type c^* depends on the sampling capacity N.

Proposition 4. *Let $c^*(N)$ denote the case type maximizing* (5) *for a given capacity N. Assume there is some $c \in \mathcal{C}$ such that $p_c > p_0$.*

(i) If $N' > N$, then $s(c_0, c^*(N')) \leq s(c_0, c^*(N))$.

(ii) There is a threshold $\bar{N}$ such that for all $N > \bar{N}$, the sender samples case type

$$c^{max} = \arg\max_{c \in \mathcal{C} \setminus \{c_\emptyset\}} \left(\frac{\sum_{c' \in \mathcal{C}} s(c, c')\mathbf{H}(c', G)}{\sum_{c' \in \mathcal{C}} s(c, c')H(c')} \right) .$$

Proposition 4 shows that a larger sampling capacity N increases the sender's incentives to sample case types that are promising in terms of their success probability but may be rather different from the current case. This is because, with a larger sampling capacity, the impact of the initial data $\mathbf{H}$ on the receiver's estimate can be weakened more easily. Indeed, when the sender has a large sampling capacity, there is no need to rely on cases that share many attributes with p_0 to modify the receiver's beliefs. Intuitively, for N sufficiently large, the sender can overwhelm the initial data using any case type, so similarity becomes relatively less important. In the limit, as $N \to +\infty$, the significance of the similarity to c_0 vanishes altogether, and the sender simply samples the case type c that, according to the initial data, is most likely to generate outcome G.

Figure 2 illustrates the comparative statics results. As N increases, the sender's indifference curve becomes flatter: the sender is more willing to sacrifice similarity for a higher likelihood of generating favorable outcomes. The case type that was optimal before the increase in sampling capacity (see Figure 1) is now dominated by a case type with a strictly lower similarity and a strictly higher success probability, marked by the red circle. In the limit as $N \to \infty$, the indifference curve becomes completely flat so that the optimal case type c is the one with the highest value of p_c (the leftmost point in Figure 2). Note that a qualitatively similar picture obtains when we increase p_0 or decrease s_{0H}. In the latter case, the indifference curve becomes flat in the limit as $s_{0H} \to 0$. Indeed, as the relevance of the initial data vanishes, the receiver's estimate converges to the frequency of G in the new data, even if the sampled case type is very dissimilar to c_0.

In summary, the sender has the ability to influence the receiver's estimate of the outcomes in problem c_0 by selectively sampling case types that are more promising than c_0 based on the initial data $\mathbf{H}$. The sender's ability to bias the receiver is limited by the fact that the receiver discounts observations that share few attributes with the current case. This does not apply when the sender can generate a substantial amount of data or if the relevance of the existing public data to the current problem is negligible. In contrast, when the capacity constraint is tight or there is substantial prior knowledge from existing data, the sender is compelled to trade off the probability of success with the similarity to the current case.

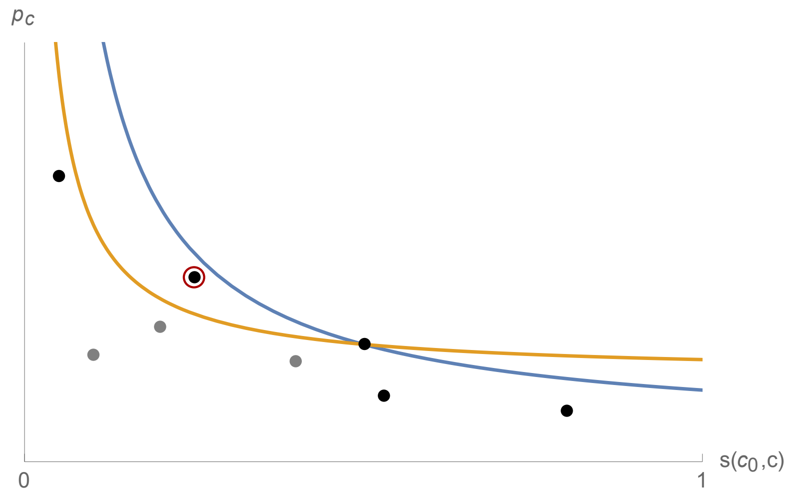

Figure 2: An increase in N results in a flatter indifference curve (graphed in orange). Compared to the situation in Figure 1, the index is now maximized by a case type with lower similarity and higher success probability.

3.2 Discussion

Sampling costs. We assumed throughout that the sender has a fixed sampling capacity N. The cost of sampling $n \leq N$ observations is thus zero, while the cost of sampling more than N observations is arbitrarily large. The setting can be extended to the case where the sender's cost of sampling n cases is described by a smoothly increasing and weakly convex function $K(n)$. The sender's optimization problem then becomes:

$$\max_D \hat{p}_0(D) - K\left(\sum_{c \in \mathcal{C} \setminus \{c_\emptyset\}} D(c)\right)$$

subject to (3).

As long as the cost function does not depend on the sampled case types, it is clear that the optimal case-type composition for any sample size n will continue to be described by Theorem 1. Our previous characterization thus applies, but now the sender has to solve the additional problem of determining the optimal sample size. Since the GSA does not necessarily pick the replacement with the highest marginal gain, it is not straightforward to suitably modify the algorithm, for instance, by comparing in each stage the marginal replacement gain with the associated cost. However, given that the gains from sampling are bounded above by one, the size of the optimal case-type collection is bounded above by $\overline{N} := \min\{n \in \mathbb{N} : K(n) \geq 1\}$. To determine the sender's solution, it thus suffices to compare the net payoff associated with the case-type collection obtained as an outcome of the GSA in $\overline{N}$ cases. Hence, for any increasing convex cost function $K(\cdot)$, our method can be used to solve the sender's optimization problem in a number of steps that is polynomial in the input.

Sampling gains. The sender's payoff in our setting is linearly increasing in the receiver's posterior, so his objective is to maximize the expected value of the receiver's final estimate. The assumption that the sender's payoff increases continuously in the receiver's estimate is natural in settings where the receiver has the ability to influence the choice of action but does not possess full authority over the decision. Linearity is instead a simplifying technical assumption, which isolates the sender's incentives to bias the receiver's estimate from other known persuasion forces. A natural alternative assumption is that the receiver adopts the action if and only if her posterior is above a threshold $\rho \in (0, 1)$.[11] This assumption is plausible if the receiver has full authority to decide on the implementation of the action and there are no additional factors that could influence her decision. The sender's payoff as a function of the receiver's posterior, in this case, is piecewise constant with a single upward jump at ρ.

To illustrate the implications of relaxing linearity, let us focus on this case and assume the receiver has a fixed acceptance threshold ρ. The goal of the sender is then to maximize the probability with which the receiver's estimate exceeds this threshold. Formally, the sender's optimization problem is described by

$$\max_{D} \Pr\left(\frac{\sum_{c\in\mathcal{C}} s(c_0, c)[\mathbf{D}(c, G) + \mathbf{H}(c, G)]}{\sum_{c\in\mathcal{C}} s(c_0, c)[D(c) + H(c)]} \geq \rho\right)$$

subject to (3) and (4), where $\mathbf{D}$ is a random variable induced by the case-type collection D. Clearly, if $p_0 \geq \rho$, the sender's problem is trivial because adoption is guaranteed even if he does not provide additional data. We thus focus on the case $p_0 < \rho$. Since each data point corresponds to a Bernoulli experiment, the collection of observations featuring the same case type follows a binomial distribution. Using this fact, the adoption probability takes the form

$$\Pr\left(\sum_{c\in\mathcal{C}} X_{D(c),p_c} s(c_0, c) \geq \rho \sum_{c\in\mathcal{C}} s(c_0, c) D(c) + (\rho - p_0) \sum_{c\in\mathcal{C}} s(c_0, c) H(c)\right),$$

[11] The literature on Bayesian persuasion considers significantly more general objective functions.

where $X_{n,p} \sim B(n,p)$.

Since our sender does not have full flexibility in choosing the distribution over induced posterior beliefs but is constrained by the underlying structure of finite databases and available case types, standard tools, in particular concavication methods, are not applicable here. The main tradeoff, however, is easy to see when the sender is restricted to homogeneous databases that feature only a single case type $c \in \mathcal{C}$. In this case, the sender's problem amounts to

$$\max_{c \in \mathcal{C}} \quad \Pr\left(X_{n,p_c} \geq \rho n + \frac{\rho - p_0}{s(c_0,c)} \sum_{c' \in \mathcal{C}} s(c_0,c')H(c') \right)$$

subject to $n \leq \min\{N_c, N\}$. The expression clearly shows that the adoption probability is increasing in p_c and increasing in $s(c_0,c)$. When choosing a case type, the sender thus faces a tradeoff that is qualitatively similar to the one of the baseline model: he must weigh the likelihood of generating successes against the similarity to the current case. The expression further reveals that similarity becomes relatively more important when the gap between the acceptance threshold ρ and the prior obtained from the initial data becomes bigger. When $\rho - p_0$ is relatively large, the new data has to be sufficiently relevant for the current case in order to sway the receiver into adopting the desired action.

There are two interesting differences with respect to the baseline model. First, when the receiver has a fixed acceptance threshold, it may be optimal for the sender to sample without exhausting his capacity. In particular, if there is no case type c with $p_c \geq \rho$ or if such case types have little similarity with the current case, then the sender must sample a case type $\tilde{c}$ with a prior success probability below the acceptance threshold ($p_{\tilde{c}} < \rho$). If the sender collects a lot of data on case type $\tilde{c}$, he expects the receiver's posterior belief to be in a small neighborhood of $p_{\tilde{c}}$, and thus below ρ, with high probability. Assuming that $\tilde{c}$ is sufficiently similar to c_0 so that a few successes suffice to persuade the receiver, the sender can generate a higher acceptance probability by relying on a small sample in the hope of getting lucky.

Second, there are now situations where, due to complementarities between case types, the sender wants to sample more than one type even when case-type-specific sampling constraints are non-binding. We illustrate this feature in a simple example. Suppose there are only two case types, c_1 and c_2. Case type c_1 is more similar to the current problem but less likely to generate successes, i.e., $s(c_0,c_1) > s(c_0,c_2)$ and $p_{c_1} < p_{c_2}$. The sender can only generate two data points

($N = 2$), both of which have to be successful in order to persuade the receiver. Within the sampling constraint, the sender is free to choose any combination of case types ($N_{c_1}, N_{c_2} \geq 2$). Now suppose that $s(c_0, c_2)$ is sufficiently small so that, conditional on observing two successes, the receiver is willing to adopt action a if and only if at least one of the two data points is of the more similar case type c_1.[12] The sender's optimal sampling strategy is then $C = (c_1, c_2)$. The adoption probability generated by this strategy is $p_{c_1}p_{c_2}$, which is strictly higher than $p_{c_1}^2$, the adoption probability when sampling only case type c_1. All other sampling strategies lead to certain rejection. The sender's sampling problem is thus solved by a non-homogeneous database. Intuitively, by combining case types c_1 and c_2, the sender maximizes the probability of favorable observations subject to the constraint that, on average, these observations are still sufficiently relevant to the case at hand.

4 Competing Senders

In many situations, decision-makers have access to multiple experts whose preferences may differ. For instance, multiple lobbyists working to advance different political agendas may try to influence a policymaker. The question is then whether the policymaker benefits from the ability to obtain information from both sides of the political aisle. To capture such situations, we modify the setup of the baseline model and allow for competition between senders.

As before, there is one receiver who bases her inference on a similarity-weighted frequency criterion. There are two senders, indexed by $i = A, B$, each of whom can sample up to a total of N^i observations. For the sake of analytical tractability, we focus on the case where senders have maximal flexibility, as in Section 3.1, so that we can ignore case-type specific sampling constraints. Senders have opposing preferences over the receiver's posterior belief. Sender A, as before, would like to boost the image of action a in the eyes of the receiver, whereas Sender B wants the opposite: he has a preference for the status quo and would thus like to discredit action a. More specifically, we maintain the assumption that Sender A's payoff is

[12]Formally, the following two inequalities must hold:

$$\begin{aligned} \rho &> \frac{s(c_0,c_1) + \sum_{c\in\mathcal{C}} s(c_0,c)\mathbf{H}(c,G)}{s(c_0,c_1) + s_{0H}}, \frac{2s(c_0,c_2) + \sum_{c\in\mathcal{C}} s(c_0,c)\mathbf{H}(c,G)}{2s(c_0,c_2) + s_{0H}} \\ \rho &< \frac{s(c_0,c_1) + s(c_0,c_2) + \sum_{c\in\mathcal{C}} s(c_0,c)\mathbf{H}(c,G)}{s(c_0,c_1) + s(c_0,c_2) + s_{0H}} \quad . \end{aligned}$$

linear and increasing in the realized posterior and analogously add that Sender B's payoff is linear and decreasing in the realized posterior. Sender A's optimization problem then remains to maximize the receiver's expected posterior belief, while Sender B's optimization problem is to minimize it. Both senders share the same prior on the success probabilities of the different case types, which is again based on the similarity-weighted frequency computed over the initial database $\mathbf{H}$.

Best responses. The strategic interaction between the two senders takes the form of a zero-sum game. Suppose that Sender A's strategy is to sample a collection D^A of case types, while Sender B's strategy is to sample the collection D^B. These samples jointly induce a random database. The receiver's expected posterior belief is

$$\hat{p}_0(D^A + D^B) := \frac{\sum_{c \in \mathcal{C}} s(c_0, c)[(D^A(c) + D^B(c))p_c + \mathbf{H}(c, G)]}{\sum_{c \in \mathcal{C}} s(c_0, c)(D^A(c) + D^B(c) + H(c))} .$$

From the perspective of each sender, the problem of maximizing or minimizing this expectation is similar to the problem analyzed in Section 3.1. Compared to the single-sender benchmark, each sender now optimizes against a larger initial database, which is the combination of $\mathbf{H}$ and the database generated by the other sender. From the viewpoint of a given sender, competition thus changes the total similarity of the 'external database' to case type c_0—i.e., the data publicly available or supplied by the other sender— as well as the expected baseline probability of success.

Lemma 2. *For each case-type collection D, define*

$$\tilde{p}_0(D) := \frac{\sum_{c \in \mathcal{C}} s(c_0, c)[D(c)p_c + \mathbf{H}(c, G)]}{\sum_{c \in \mathcal{C}} s(c_0, c)[D(c) + H(c)]} .$$

A sender's best response to the other sender playing a pure strategy D is:

1. *For Sender A, to sample N^A observations of type*

$$c_A^* = \arg\max_{c \in \mathcal{C}} \left(\frac{s(c_0, c)(p_c - \tilde{p}_0(D))}{N^A s(c_0, c) + \sum_{c \in \mathcal{C}} s(c_0, c)[D(c) + H(c)]} \right) .$$

2. *For Sender B, to sample N^B observations of type*

$$c_B^* = \arg\min_{c \in \mathcal{C}} \left(\frac{s(c_0, c)(p_c - \tilde{p}_0(D))}{N^B s(c_0, c) + \sum_{c \in \mathcal{C}} s(c_0, c)[D(c) + H(c)]} \right) .$$

Lemma 2 shows that senders best respond to a pure strategy by either not sampling ($c_i^* = c_\emptyset, i = A, B$) or by sampling a single case type up to capacity, just as in the single sender setting. As stated above, the best response of each sender depends on the choice of the other sender through the similarity weight $\sum_{c \in \mathcal{C}} s(c_0, c)[D(c) + H(c)]$ as well as through $\tilde{p}_0(D)$, the expected probability that will obtain if the considered sender provides no additional data.

The description of the best response in Lemma 2 is valid only if the senders play a pure strategy. It is well known that zero-sum games often do not have pure strategy equilibria. The following lemma shows that this is not a concern here.

Lemma 3. *There exists an equilibrium in pure strategies.*

The existence of a pure strategy equilibrium is proved by contradiction. Given Lemma 2, a pure strategy can be described by the case type a given sender samples. If no pure strategy equilibrium exists, then there must be a set of pure strategies that form a cycle of best responses (recall that there is a finite number of pure strategies available to each sender). We show that within this cycle, there is a sub-cycle containing four case types, c_j^A, c_{j+1}^A for Sender A and c_j^B, c_{j+1}^B for Sender B, that yield the following preferences: Sender A prefers sampling case type c_j^A to case type c_{j+1}^A when Sender B chooses case type c_j^B and the converse when Sender B chooses case type c_{j+1}^B. At the same time, Sender B prefers case type c_j^B to case type c_{j+1}^B when Sender A chooses case type c_{j+1}^A and the converse when Sender A chooses case type c_j^A. These preferences correspond to four inequalities, where at least one of them must be strict. Summing over these inequalities generates the contradiction.

Equilibrium. We now examine how competition between senders impacts equilibrium sampling. Specifically, we investigate the conditions under which each sender provides data to the receiver. To this end, let $\bar{c}^*$ denote the case type maximizing the success-similarity index (5) under the initial data $\mathbf{H}$ and $\underline{c}^*$ denote the case type minimizing this index. These two case types describe, respectively, Sender A's and Sender B's optimal sampling if they act in isolation.

Proposition 5. *There is a pure-strategy equilibrium with only one sender sampling if and only if for all $c \in \mathcal{C}$,*

$$p_c \geq \frac{N^A s(c_0, \bar{c}^*) p_{\bar{c}^*} + \sum_{c \in \mathcal{C}} s(c_0, c) \mathbf{H}(c, G)}{N^A s(c_0, \bar{c}^*) + \sum_{c \in \mathcal{C}} s(c_0, c) H(c)}, \tag{6}$$

or for all $c \in \mathcal{C}$,

$$p_c \leq \frac{N^B s(c_0, \underline{c}^*) p_{\underline{c}^*} + \sum_{c \in \mathcal{C}} s(c_0, c)\mathbf{H}(c, G)}{N^B s(c_0, \underline{c}^*) + \sum_{c \in \mathcal{C}} s(c_0, c) H(c)}. \tag{7}$$

If neither of these conditions holds, then there is a pure strategy equilibrium in which both senders sample up to capacity.

The result directly follows from the characterization of the senders' best responses in Lemma 2 and therefore its proof is omitted. Despite its simplicity, the result carries interesting implications in terms of competition between the two senders and their incentives to take up sampling. Consider the case where Sender B samples the case type $\underline{c}^*$ that would be optimal if Sender A were not present. Note that if $\mathcal{C}$ contains case types with a success probability below p_0, then $p_{\underline{c}^*} < p_0$ and the right-hand side of (7) is strictly smaller than p_0. Suppose now there is some $c \in \mathcal{C}$ such that

$$\frac{N^B s(c_0, \underline{c}^*) p_{\underline{c}^*} + \sum_{c \in \mathcal{C}} s(c_0, c)\mathbf{H}(c, G)}{N^B s(c_0, \underline{c}^*) + \sum_{c \in \mathcal{C}} s(c_0, c) H(c)} < p_c < p_0,$$

and that all case types with a higher success probability have an arbitrarily low similarity with c_0. Then, Sender A will find it optimal to sample case type c, even though it has a lower success probability than c_0 according to the initial data. This is because by sampling case type c, Sender A reduces the impact of Sender B's sampling on the receiver's final estimate. Providing evidence on case type c decreases the weight the receiver assigns to even more detrimental case types.

In order to weaken the evidence provided by his opponent, Sender A, in this case, chooses a case type that is less promising but more similar to c_0 than the one he would choose in isolation. We next show that this property holds more generally.

Proposition 6. *In any pure strategy equilibrium where both senders sample,*

1. *Sender A chooses a case type c such that $s(c_0, c) \geq s(c_0, \overline{c}^*)$;*
2. *Sender B chooses a case type c' such that $s(c_0, c') \geq s(c_0, \underline{c}^*)$.*

In the proof, we show that from the perspective of Sender A, the receiver's expected estimate can be expressed as a weighted average of the receiver's estimate when Sender A acts in isolation and chooses to sample case type c and the success probability of the case type chosen by Sender B. The weights in this average depend on the similarity value of Sender A's chosen case type c. We then establish that choosing any case type with a lower similarity than that of case type $\bar{c}^*$ is dominated by case type $\bar{c}^*$. An analogous argument applies to Sender B.

Our findings are related to some results in the literature on competition between multiple senders under Bayesian persuasion. Among others, Gentzkow and Kamenica (2017), Au and Kawai (2020), and Koessler, Laclau, Renault, and Tomala (2021) show that competition leads to more informative equilibria in comparison to the collusive outcome. Consistently with the existing literature, our results indicate that competing senders will generate a more balanced composition of observations as compared to when they are the sole provider of information. In particular, competition in our case-based environment enhances the relevance of the data made available to the receiver by inducing both senders to select case types more similar to the current one in comparison to their sampling behavior in isolation.

5 Conclusions

This paper introduces a classical statistics approach to persuasion, considering a receiver who forms an estimate of the outcomes of a new action via the similarity-weighted frequencies in the provided data. The sender decides on the case types in the database and faces a novel tradeoff between selecting types that are relatively promising according to the initial data and sampling types that share more characteristics with the problem faced by the receiver. Due to the fact that signals are discrete databases, the sender's design problem becomes a combinatorial optimization problem over sets of integer-valued alternatives. We characterize the sender's optimal sampling strategy as the outcome of a greedy search algorithm and study the impact of the initial public data and the sender's sampling capacity on the solution.

The Bayesian approach to information design has generated many important economic insights. We believe that our model complements the literature by offering a number of new benefits. First, being less abstract than the canonical approach, it can help to clarify the link between the characteristics of signals and the belief-formation process. Second, the explicit formulation of data facilitates the modeling of possible constraints on the set of feasible posterior distributions, for instance, via the set of feasible case types, the tightening/relaxation of sampling constraints, or regulatory restrictions on experimental trials. Finally, our model can be seen as a stepping stone to the analysis of more general statistical inference procedures in games with strategic data provision.

A Appendix

Notation To ease notation, this appendix abbreviates by $s_{cc'}$ the similarity $s(c, c')$, between case types c and c', with s_{0c} denoting the similarity $s(c_0, c)$ between case types c_0 and c. Furthermore, for a database $\mathbf{D}$ and a case type c', we define

$$S(c, \mathbf{D}) = \sum_{(c',y)\in\mathcal{C}\times\{G,B\}} s(c, c')\mathbf{D}(c', y)$$

as the total similarity of c with cases belonging to $\mathbf{D}$. Finally, we denote by $\mathbf{D}_G$ the sub-database of $\mathbf{D}$ containing only those observations for which the realized outcome was G. Hence, $S(c, \mathbf{D}_G) = \sum_{c'\in\mathcal{C}} s(c, c')\mathbf{D}(c', G)$ is the total similarity of c' with the successful observations in $\mathbf{D}$.

A.1 Proof of Lemma 1

Let D be a case-type collection and $c, c' \in \mathcal{C}$ be two case types such that $D(c) > 0$ and $D(c') < N_{c'}$. Let D' be a case-type collection such that $D'(c) = D(c) - 1$, $D'(c') = D(c) + 1$ and $D'(c'') = D(c'')$ for all $c'' \neq c, c'$.

(i) Suppose that $s_{0c'} > s_{0c}$. Then:

$$\begin{aligned}
\hat{p}_0(D') = \frac{\sum\limits_{c\in\mathcal{C}} p_c s_{0c} D(c) + s_{0c'}p_{c'} - s_{0c}p_c + S(c_0, \mathbf{H}_G)}{\sum\limits_{c\in\mathcal{C}} s_{0c} D(c) + s_{0c'} - s_{0c} + S(c_0, \mathbf{H})} \quad &> \quad \frac{\sum\limits_{c\in\mathcal{C}} p_c s_{0c} D(c) + S(c_0, \mathbf{H}_G)}{\sum\limits_{c\in\mathcal{C}} s_{0c} D(c) + S(c_0, \mathbf{H})} = \hat{p}_0(D) \\
\iff \quad \frac{s_{0c'}p_{c'} - s_{0c}p_c}{s_{0c'} - s_{0c}} \quad &> \quad \frac{\sum\limits_{c\in\mathcal{C}} p_c s_{0c} D(c) + S(c_0, \mathbf{H}_G)}{\sum\limits_{c\in\mathcal{C}} s_{0c} D(c) + S(c_0, \mathbf{H})} = \hat{p}_0(D) \\
\iff \quad \phi(c', c) \quad &> \quad \hat{p}_0(D)
\end{aligned}$$

(ii) Suppose that $s_{0c'} < s_{0c}$. Then:

$$\begin{aligned}
\hat{p}_0(D') = \frac{\sum\limits_{c\in\mathcal{C}} p_c s_{0c} D(c) + s_{0c'}p_{c'} - s_{0c}p_c + S(c_0, \mathbf{H}_G)}{\sum\limits_{c\in\mathcal{C}} s_{0c}D(c) + s_{0c'} - s_{0c} + S(c_0,\mathbf{H})} &> \frac{\sum\limits_{c\in\mathcal{C}} p_c s_{0c} D(c) + S(c_0, \mathbf{H}_G)}{\sum\limits_{c\in\mathcal{C}} s_{0c}D(c) + S(c_0,\mathbf{H})} = \hat{p}_0(D) \\
\iff \quad \frac{s_{0c'}p_{c'} - s_{0c}p_c}{s_{0c'} - s_{0c}} &< \frac{\sum\limits_{c\in\mathcal{C}} p_c s_{0c} D(c) + S(c_0, \mathbf{H}_G)}{\sum\limits_{c\in\mathcal{C}} s_{0c}D(c) + S(c_0,\mathbf{H})} = \hat{p}_0(D) \\
\iff \quad \phi(c', c) &< \hat{p}_0(D)
\end{aligned}$$

A.2 Proof of Theorem 1

For any case-type collection C, denote by

$$\phi^*(C) := \max_{(c,c')\in R(C)} \phi(c, c')$$

the maximal index over the set of feasible replacements and define $\phi_n^* := \phi^*(D_{n-1})$.

A.2.1 Preliminary Lemmas

Before moving to the main argument, let us prove three preliminary lemmas.

Lemma 4. *Consider three case types c, c', c'' such that $s_{0c} > s_{0c'} > s_{0c''}$. Then*

$$\min\{\phi(c, c'), \phi(c', c'')\} < \phi(c, c'') < \max\{\phi(c, c'), \phi(c', c'')\}$$

Proof. Towards a contradiction, assume first $\phi(c, c'') > \phi(c, c'), \phi(c', c'')$. The inequality $\phi(c, c'') > \phi(c, c')$ can be written as

$$\begin{aligned}
\frac{s_{0c}p_c - s_{0c''}p_{c''}}{s_{0c} - s_{0c''}} &> \frac{s_{0c}p_c - s_{0c'}p_{c'}}{s_{0c} - s_{0c'}} \\
\iff \quad (s_{0c} - s_{0c'})(s_{0c}p_c - s_{0c''}p_{c''}) &> (s_{0c} - s_{0c''})(s_{0c}p_c - s_{0c'}p_{c'}) \\
\iff \quad s_{0c}(s_{0c'}p_{c'} - s_{0c''}p_{c''}) + s_{0c''}(s_{0c}p_c - s_{0c'}p_{c'}) &> s_{0c'}(s_{0c}p_c - s_{0c''}p_{c''}) \ .
\end{aligned}$$

The inequality $\phi(c, c'') > \phi(c', c'')$ can be written as

$$\begin{aligned}
\frac{s_{0c}p_c - s_{0c''}p_{c''}}{s_{0c} - s_{0c''}} &> \frac{s_{0c'}p_{c'} - s_{0c''}p_{c''}}{s_{0c'} - s_{0c''}} \\
\iff \quad (s_{0c'} - s_{0c''})(s_{0c}p_c - s_{0c''}p_{c''}) &> (s_{0c} - s_{0c''})(s_{0c'}p_{c'} - s_{0c''}p_{c''}) \\
\iff \quad s_{0c'}(s_{0c}p_c - s_{0c''}p_{c''}) &> s_{0c}(s_{0c'}p_{c'} - s_{0c''}p_{c''}) + s_{0c''}(s_{0c}p_c - s_{0c'}p_{c'}) \ .
\end{aligned}$$

Both conditions together yield a contradiction. By an analogous argument (with reversed inequalities), we can also exclude the case $\phi(c, c'') < \phi(c, c'), \phi(c', c'')$. $\square$

The next lemma shows that the maximizing replacement index is (weakly) decreasing with each stage of the algorithm.

Lemma 5. *ϕ_n^* is monotonically decreasing in n.*

Proof. Suppose the statement is not true: there exists some n such that $\phi_n^* < \phi_{n+1}^*$. Let $\phi(c_i, c_j) = \phi_n^*$ and $\phi(c_k, c_l) = \phi_{n+1}^*$ with $s_{0c_i} > s_{0c_j}$ and $s_{0c_k} > s_{0c_l}$ (c_j, c_l could be the ficticious case type). By assumption, we have $\phi(c_i, c_j) < \phi(c_k, c_l)$, so it must be the case that at stage n replacing type c_l with type c_k was not feasible. This can be for two reasons:

1. D_{n-1} contains no c_l-types: $D_{n-1}(c_l) = 0$;

2. at stage n, the capacity of case types c_k was exhausted: $D_{n-1}(c_k) = N_{c_k}$.

We start with the first case. Since D_{n-1} contains no c_l-types and since replacing type c_l with type c_k is feasible in stage $n+1$, we must have $c_i = c_l$ (case type c_l enters in stage n) and hence $p_{c_k} < p_{c_l} = p_{c_i} < p_{c_j}$. Since $\phi(c_i, c_j) = \phi(c_l, c_j) < \phi(c_k, c_l)$, Lemma 4 implies $\phi(c_l, c_j) < \phi(c_k, c_j)$. But this implies that in stage n case type c_j is replaced by case type c_k rather than $c_i = c_l$, unless adding case c_k is not feasible. In this case, however, it will not be feasible in stage $n+1$ either.

Moving to the second case, assume $D_{n-1}(c_k) = N_{c_k}$. In order for the replacement of case type c_l with case type c_k to be feasible at stage $n+1$, it must be that case type c_k is replaced in stage n. We thus have $c_j = c_k$ and hence $p_{c_i} < p_{c_j} = p_{c_k} < p_{c_l}$. The assumption $\phi(c_i, c_j) = \phi(c_i, c_k) < \phi(c_k, c_l)$ implies $\phi(c_i, c_k) < \phi(c_i, c_l)$ (again by Lemma 4). Since the latter inequality holds and the algorithm replaces case type c_k with case type c_i in stage n, it means that replacing case type c_l with case type c_i is not feasible at this stage. Hence, D_{n-1} contains no c_l types, which brings us back to case 1. $\square$

Single replacements. We will now show that D_{n^*} cannot be improved upon by replacing a single case type c with another case type c'. For case types c, c' with $s_{0c'} > s_{0c}$, this property follows directly from the specification of the algorithm and Lemma 1. We can thus focus on similarity-reducing replacements.

Towards a contradiction, suppose then there is a pair of case types (c, c') with $s_{0c} > s_{0c'}$ such that replacing case type c with case type c' is optimal, i.e. $\phi(c, c') < \hat{p}_0(D_{n^*})$, and feasible, i.e. $D_{n^*}(c) > 0$ and $D_{n^*}(c') < N_{c'}$. Noting $\phi(c, c') < \hat{p}_0(D_{n^*}) < \phi_{n^*}^*$, let $n_* \leq n^*$ be the first stage in the algorithm such that there is a pair (c_i, c_j) with $s_{0c_i} > s_{0c_j}$ satisfying the properties

1. $\phi(c_i, c_j) < \phi^*_{n_*}$:

2. $D_{n_*}(c_i) > 0$;

3. $D_{n_*}(c_j) < N_{c_j}$.

Property (b) implies that there exists a subsequence of algorithm stages $S^i = (n^i_1, n^i_2, ..., n^i_K)$ with $n^i_K \leq n_*$ and an ordered set of case types $C^i = \{c^i_1, c^i_2, ..., c^i_K\}$ with $s(c^i_1, c_0) < ... < s(c^i_K, c_0)$ and $c^i_K = c_i$ such that[13]

- $\phi^*_{n^i_k} = \phi(c^i_k, c^i_{k-1})$, where $c^i_0 := c_\emptyset$;
- $D_n(c^i_k) > 0$ for all $n \in \{n^i_k, ..., n^i_{k+1} - 1\}$, $k \in \{1, ..., K-1\}$;
- $D_n(c_i) > 0$ for all $n \in \{n^i_K, ..., n_*\}$.

By monotonicity of the algorithm (Lemma 5) and definition of S^i, $\phi(c^i_k, c^i_{k-1})$ is decreasing in k. By, Lemma 4, this implies $\phi(c^i_k, c^i_{k-1}) < \phi(c^i_k, c^i_{k-2}) < \phi(c^i_{k-1}, c^i_{k-2})$, which in turn implies $\phi(c^i_k, c^i_{k-2}) < \phi(c^i_k, c^i_{k-3}) < \phi(c^i_{k-2}, c^i_{k-3})$, and so on. Hence, for all $k \in \{1, ..., K\}$ and $k' < k$, $\phi(c^i_k, c^i_{k'})$ is decreasing in k'. Since $\phi(c_i, c_j) < \phi(c_i, c^i_{K-1})$, a direct implication of this property is $c_j \notin C^i$.

Case type c_j must be at capacity when c_i enters. Let S^i be a subsequence of algorithm stages as defined above and let $n^i_{\tilde{k}}$ be the first stage of the subsequence S^i such that the corresponding case type $c^i_{\tilde{k}}$ has a similarity $s(c^i_{\tilde{k}}, c_0)$ greater than $s(c_j, c_0)$. That is, $\tilde{k}$ is the smallest number such that $s(c_i, c_0) > s(c^i_k, c_0) > s(c_j, c_0)$ for all $k \in \{\tilde{k}, ..., K-1\}$. Since $\phi(c_i, c_j) < \phi^*_{n_*} < \phi(c_i, c^i_k)$ for all k, Lemma 4 implies

$$\phi(c^i_k, c_j) < \phi(c_i, c_j) \text{ for all } k \in \{\tilde{k}, ..., K-1\}, \tag{8}$$

which further implies $\phi(c^i_k, c_j) < \phi(c^i_k, c^i_{k-1})$ for all $k \in \{\tilde{k}, ..., K-1\}$. Considering $k = \tilde{k}$ with $s(c^i_{\tilde{k}}, c_0) > s(c_j, c_0) > s(c^i_{\tilde{k}-1}, c_0)$, we then have $\phi(c^i_{\tilde{k}}, c^i_{\tilde{k}-1}) < \phi(c_j, c^i_{\tilde{k}-1})$, again by Lemma 4.[14] Hence, at stage $n^i_{\tilde{k}}$, case type c_j must be at capacity:

$$D_{n^i_{\tilde{k}}}(c_j) = N_{c_j}.$$

Let $n_j \leq n_*$ be the first stage following $n^i_{\tilde{k}}$ at which case type c_j is replaced by another case type. Suppose $n^i_k < n_j < n^i_{k+1}$ for some $k \in \{\tilde{k}, ..., K-1\}$. Then, by condition 8, Properties 1-3 are satisfied for the pair (c^i_k, c_j) at stage n_j, which

[13] S^i describes the sequence of replacements that lead to c_i.

[14] If $\tilde{k} = 1$, $c^i_{\tilde{k}-1} = c_\emptyset$.

violates our definition of c_i. We thus have $n_j = n_* < n^i_K$.

Replacement of c_j. Let c_l be such that $\phi^*_{n_*} = \phi(c_l, c_j)$. Since $n_* < n^i_K$, we have $\phi(c_l, c_j) < \phi(c_i, c^i_{K-1})$. We distinguish two cases according to the relative similarity of case type c_l:

1. $s_{0c_i} > s_{0c_l}$: The inequality $\phi(c_i, c_j) < \phi(c_l, c_j)$ implies, by Lemma 4, $\phi(c_i, c_l) < \phi(c_l, c_j)$. Since c_l replaces c_j in stage n_*, we must have $D_{n_*-1}(c_l) < N_{c_l}$. Note further that

$$\phi(c_i, c_l) < \phi(c_l, c_j) = \phi^*_{n_*} < \phi^*_{n_*-1}.$$

   These two properties, together with $D_{n_*-1}(c_i) > 0$, violate the definition of n_* as the first stage in the algorithm satisfying Properties 1-3.

2. $s_{0c_l} > s_{0c_i}$: The inequality $\phi(c_i, c_j) < \phi(c_l, c_j)$ now implies $\phi(c_l, c_j) < \phi(c_l, c_i)$, again by Lemma 4. Given $D_{n_*-1}(c_i) > 0$, this contradicts $\phi^*_{n_*} = \phi(c_l, c_j)$, as c_l would be replaced by c_i rather than c_j according to the algorithm.

Combined Replacements. What remains to be shown is if the sender cannot increase $\hat{p}_0$ locally by replacing a single case type in D_{n^*}, then $\hat{p}_0$ cannot be increased either by replacing a subset of case types in the collection.

Lemma 6. *Let D and D' be two case-type collections such that $\hat{p}_0(D) < \hat{p}_0(D')$. Then there exist two case types $c, c' \in \mathcal{C}$ such that $D(c) > 0$ and a case-type collection $\tilde{D}$ defined by $\tilde{D}(c) = D(c) - 1$, $\tilde{D}(c') = D(c') + 1$ and $\tilde{D}(c'') = D(c)$ for all $c'' \neq c, c'$ such that $\hat{p}_0(\tilde{D}) > \hat{p}_0(D)$.*

Proof. Consider case-type collection D and suppose there is no single replacement that yields an increase of the expected posterior. That is, for all c with $D(c) > 0$ and $c' \neq c$, we have $\phi(c', c) \leq \hat{p}_0(D)$ if $s_{0c'} > s_{0c}$ and $\phi(c', c) \geq \hat{p}_0(D)$ if $s_{0c'} < s_{0c}$. Towards a contradiction, suppose then there is an alternative case-type collection D' such that $\hat{p}_0(D') > \hat{p}_0(D)$. Let us define $\Delta : \mathcal{C} \to \mathbb{Z}$ such that $\Delta(c) = D'(c) - D(c)$ for all $c \in \mathcal{C}$. Following the same argument as in the proof of Lemma 1, $\hat{p}_0(D') > \hat{p}_0(D)$ requires

$$\frac{\sum_{c\in\mathcal{C}} p_c s_{0c} \Delta(c)}{\sum_{c\in\mathcal{C}} s_{0c} \Delta(c)} > \hat{p}_0(D)$$

if $\sum_{c\in\mathcal{C}} s_{0c}\Delta(c) > 0$ and the reverse inequality if $\sum_{c\in\mathcal{C}} s_{0c}\Delta(c) < 0$. In either case, we can write this condition as

$$\sum_{c\in\mathcal{C}} p_c s_{0c}\Delta(c) > \sum_{c\in\mathcal{C}} \hat{p}_0(D) s_{0c}\Delta(c)\ .$$

Since $\sum_{c\in\mathcal{C}}\Delta(c) = 0$, there must then exist two case types c' and c such that $\Delta(c') > 0$, $\Delta(c) < 0$, and

$$p_{c'}s_{0c'} - p_c s_{0c} > \hat{p}_0(D)(s_{0c'} - s_{0c}).$$

If $s_{0c'} > s_{0c}$, this inequality is equivalent to $\phi(c', c) > \hat{p}_0(D)$, whereas if $s_{0c'} < s_{0c}$, it is equivalent to $\phi(c', c) < \hat{p}_0(D)$. Both cases contradict the assumption that case-type collection D cannot be improved upon via a single replacement.

□

A.3 Proof of Proposition 2

Let $D^*(p_0, s_{0H})$ denote the solution of the algorithm for a given probability of success p_0 and similarity s_{0H}. Note that the order of steps through which the algorithm runs and the associated case-type collections, $\langle D_0, D_1, D_2, ...\rangle$, are independent of $\mathbf{H}$. The public data only matters for the stopping point of the algorithm. According to Definition 2, the index of the last stage, denoted by $n^*(p_0, s_{0H})$, is the largest number such that for all $n < n^*(p_0, s_{0H})$

$$\hat{p}_0(D_n; p_0, s_{0H}) = \frac{\sum_{c\in\mathcal{C}} p_c s_{0c} D_n(c) + p_0 s_{0H}}{\sum_{c\in\mathcal{C}} s_{0c} D_n(c) + s_{0H}} < \phi^*(D_n).$$

We observe that $\hat{p}_0(D_n; s_{0H}, p_0)$ is increasing in p_0 and, for all $n \leq n^*(p_0, s_{0H})$, decreasing in s_{0H}. To see the latter, consider

$$\frac{\partial(\hat{p}_0(D_n; p_0 s_{0H}))}{\partial s_{0H}} = -\frac{\sum_{c\in\mathcal{C}}(p_c - p_0) s_{0c} D_n(c)}{\left(\sum_{c\in\mathcal{C}} s_{0c} D_n(c) + s_{0H}\right)^2}. \tag{9}$$

Since each replacement stage of the GSA constitutes an improvement of the expected posterior, we have $p_c > p_0$ for all $c \in D_n$ and all $n \leq n^*(p_0, s_{0H})$ (see Lemma 1). Hence, (9) is strictly negative.

Part(i). Suppose now database $\mathbf{H}$ changes to $\mathbf{H}'$ such that $s_{0H} < s_{0H'}$ and p_0 remains invariant. We then have for all $n < n^*(p_0, s_{0H})$,

$$\hat{p}_0(D_n; s_{0H'}, p_0) < \hat{p}_0(D_n; s_{0H}, p_0) < \phi^*(D_n).$$

Hence, $n^*(p_0, s_{0H'}) \geq n^*(p_0, s_{0H})$. Since with each stage n of the algorithm the total similarity $s(c_0, D_n)$ becomes larger, this implies $s(c_0, D^*(p_0, s_{0H'})) \geq s(c_0, D^*(p_0, s_{0H}))$.

Part(ii). Similarly, suppose database $\mathbf{H}$ changes to $\mathbf{H}'$ such that $p_0 > p_0'$ and s_{0H} remains invariant. We have for all $n \leq n^*(p_0, s_{0H})$,

$$\hat{p}_0(D_n; s_{0H}, p_0') < \hat{p}_0(D_n; s_{0H}, p_0) < \phi^*(D_n).$$

Hence, $n^*(p_0', s_{0H}) \geq n^*(p_0, s_{0H})$ and $s(c_0, D^*(p_0', s_{0H})) \geq s(c_0, D^*(p_0, s_{0H}))$.

A.4 Proof of Proposition 3

To prove the statement, we distinguish two cases.

Case (i). Suppose $p_c \leq p_0$ for all $c \in \mathcal{C}\backslash\{c_\emptyset\}$. In this case, we have $c^* = c_\emptyset$, i.e., not sampling is optimal for the sender. This directly follows from Lemma 1 . Indeed, consider a case-type collection D of size $n = \sum_{c\in\mathcal{C}} D(c) \leq N$ and let case type $c' = \arg\min_{c:D(c)>0} p_c$ be the type minimizing the success probability among all case types sampled under D, with $p_{c'} \leq p_0$. The sender then benefits from excluding c' if and only if $\hat{p}_0(D') \geq \hat{p}_0(D)$, where D' is the case-type collection obtained from D by omitting one instance of c'. Namely, $D'(c') = D(c') - 1$ and $D'(c) = D(c)$ for all $c \neq c'$. Since $0 = s(c_0, c_\emptyset) < s(c_0, c')$, by applying Lemma 1 (part (ii)) we have that

$$\hat{p}_0(D') \geq \hat{p}_0(D) \quad \text{if and only if} \quad p_{c'} = \phi(c_\emptyset, c') \leq \hat{p}_0(D).$$

The last inequality holds because $p_{c'} \leq p_c$ for all c such that $D(c) > 0$. Thus, it is always (weakly) better not to sample c'. By the same token, the sender is weakly better off by discarding all case types with prior probability equal to or smaller than p_0.

Case (ii). Assume now there exists some $c \in \mathcal{C}\backslash\{c_\emptyset\}$ such that $p_c > p_0$. According to Proposition 3, the sender will exhaust his sampling capacity by choosing a homogenous database featuring a single case type that maximizes the index (5).

Step 1. At stage 1, we have $\phi^*(D_0) = \phi(\bar{c}, c_\emptyset) = p_{\bar{c}}$, where $p_{\bar{c}} := \max_{c\in\mathcal{C}\backslash\emptyset} p_c$. The replacement is clearly beneficial since $p_{\bar{c}} > p_0 = \hat{p}_0(D_0)$. In the next stage, the set of feasible replacements still contains $(\bar{c}, c_\emptyset)$ and, hence, the maximizer is unchanged. Proceeding in this way, the algorithm arrives at a collection containing N instances of case type $\bar{c}$.

Step 2. We claim that, if at any stage n, the maximal replacing index is $\phi^*(D_n) = \phi(c', c) > \hat{p}_0(D_n)$, then the GSA will replace all instances of c with c', where $s(c_0, c') > s(c_0, c)$.

Indeed, let $\phi^*(D_n) = \phi(c', c)$ at stage n. Then, at stage $n+1$, if not all instances of c were replaced, the maximal replacing index is unchanged. This follows from

$$\phi(c', c) > \hat{p}_0(D_{n+1}) = \frac{\sum_{c\in\mathcal{C}} p_c s_{0c} D_n(c) - p_c s_{0c} + p_{c'} s_{0c'} + S(c_0, \mathbf{H}_G)}{\sum_{c\in\mathcal{C}} s_{0c} D_n(c) - s_{0c} + s_{0c'} + S(c_0, \mathbf{H})}$$

if and only if

$$\phi(c', c) > \hat{p}_0(D_n) = \frac{\sum_{c\in\mathcal{C}} p_c s_{0c} D_n(c) + S(c_0, \mathbf{H}_G)}{\sum_{c\in\mathcal{C}} s_{0c} D_n(c) + S(c_0, \mathbf{H})}.$$

We thus conclude that the optimal collection of case types D_{n*} is homogeneous.

Step 3. For every $c \in \mathcal{C}$, let D_c^N be the case-type collection containing N instances of case type c, that is, $D_c^N(c) = N$ and $D_c^N(c') = 0$ for every $c' \neq c$. Then, D_c^N is optimal if $\hat{p}_0(D_c^N) \geq \hat{p}_0(D_{c'}^N)$ for every $c' \in \mathcal{C}$. Namely, if,

$$\begin{aligned}
\hat{p}_0(D_c^N) = \frac{N s_{0c} p_c + S(c_0, \mathbf{H}_G)}{N s_{0c} + S(c_0, \mathbf{H})} &\geq \frac{N s_{0c'} p_{c'} + S(c_0, \mathbf{H}_G)}{N s_{0c'} + S(c_0, \mathbf{H})} = \hat{p}_0(D_{c'}^N) \\
\iff N s_{0c} p_c s_{0c'} + s_{0c} p_c S(c_0, \mathbf{H}) + s_{0c'} s(c_0, \mathbf{H}_G) &\geq N s_{0c'} p_{c'} s_{0c} + s_{0c'} p_{c'} S(c_0, \mathbf{H}) + s_{0c} S(c_0, \mathbf{H}_G) \\
\iff N s_{0c} s_{0c'} (p_c - p_0) + s_{0c} S(c_0, \mathbf{H})(p_c - p_0) &\geq N s_{0c} s_{0c'} (p_{c'} - p_0) + s_{0c'} S(c_0, \mathbf{H})(p_{c'} - p_0) \\
\iff s_{0c}(p_c - p_0)(N s_{0c'} + S(c_0, \mathbf{H})) &\geq s_{0c'}(p_{c'} - p_0)(N s_{0c} + S(c_0, \mathbf{H})) \\
\iff \frac{s_{0c}(p_c - p_0)}{N s_{0c} + S(c_0, \mathbf{H})} &\geq \frac{s_{0c'}(p_{c'} - p_0)}{N s_{0c'} + S(c_0, \mathbf{H})}.
\end{aligned}$$

A.5 Proof of Proposition 4

Part (I). Let $c^* := c^*(N, p_0, \mathbf{H})$ denote the optimal case type for a given capacity N, a prior probability of success p_0, and an initial database $\mathbf{H}$. That is, for all $c \in \mathcal{C}$,

$$\begin{aligned}
\frac{s_{0c^*}(p_{c^*} - p_0)}{N s_{0c^*} + S(c_0, \mathbf{H})} &\geq \frac{s_{0c}(p_c - p_0)}{N s_{0c} + S(c_0, \mathbf{H})} \\
\iff N s_{0c^*} s_{0c}(p_{c^*} - p_c) &\geq S(c_0, \mathbf{H})[(s_{0c}(p_c - p_0) - s_{0c^*}(p_{c^*} - p_0)]. \quad (10)
\end{aligned}$$

Let the capacity increase to N', and the new optimum be $c'^* := c^*(N', p_0, \mathbf{H})$. That is,

$$\frac{s_{0c'^*}(p_{c'^*} - p_0)}{N's_{0c'^*} + S(c_0, \mathbf{H})} \geq \frac{s_{0c^*}(p_{c^*} - p_0)}{N's_{0c^*} + S(c_0, \mathbf{H})} \iff$$
$$N's_{0c^*}s_{0c'^*}(p_{c'^*} - p_{c^*}) \geq S(c_0, \mathbf{H})[(s_{0c^*}(p_{c^*} - p_0) - s_{0c'^*}(p_{c'^*} - p_0)]. \quad (11)$$

Combining eq. (10) and (11), we obtain that $(N'-N)s_{0c^*}s_{0c'^*}(p_{c'^*}-p_{c^*}) \geq 0$, which holds if and only if $p_{c'^*} \geq p_{c^*}$. Moreover, it must be that $s_{0c'^*} \leq s_{0c^*}$ (otherwise case type c^* would be dominated by c' for all sampling capacities).

Part (ii). Consider case-type collection D_c^N such that $D_c^N(c) = N$ and $D_c^N(c') = 0$ for every $c' \neq c$. Then, the expected posterior probability is

$$\hat{p}_0(D_c^N) = \frac{Ns_{0c}p_c + S(c_0, \mathbf{H})p_0}{Ns_{0c} + S(c_0, \mathbf{H})} = \frac{p_c}{1 + \frac{S(c_0,\mathbf{H})}{Ns_{0c}}} + \frac{S(c_0, \mathbf{H})p_0}{Ns_{0c} + S(c_0, \mathbf{H})} \to p_c = \frac{s(c, \mathbf{H}_G)}{s(c, \mathbf{H})}$$

as $N \to \infty$. Hence, for N sufficiently large, the optimal case type c^* is the case type that maximizes p_c over $c \in \mathcal{C}$.

A.6 Proof of Lemma 2

Suppose that Sender $i = A, B$ samples a case-type collection D. The opponent sender will then best respond to the joint similarity $\sum_{c\in\mathcal{C}}[D(c) + H(c)]$ and the revised posterior probability $\tilde{p}_0(D)$, as defined in Lemma 2, based on $\mathbf{H}$ and the expected similarity-weighted frequency of success within D. This probability is of the same form as p_0 in the single-sender problem. Hence, Proposition 3 applies, where $\sum_{c\in\mathcal{C}}[D(c) + H(c)]$ takes the place of $\sum_{c\in\mathcal{C}} H(c)$ and $\tilde{p}_0(D)$ takes the place of p_0.

A.7 Proof of Lemma 3

From Lemma 2, we know that the best response of a sender to a pure strategy of the opponent is either to sample nothing or to sample a database of maximal capacity composed of only one case type. As before, for a case type c and a positive integer m, let D_c^m denote the case-type collection consisting of m observations of type c, namely, $D_c^m(c) = m$ and $D_c^m(c') = 0$ for every case type $c' \neq c$.

First, note that an equilibrium where both senders do not sample can exist only if there are no types with success probability different from p_0. Otherwise, if there is a type c_1 with probability p_{c_1} that is different from p_0, we have the following: if $p_{c_1} > p_0$, then Sender A will be better off sampling c_1 than not sampling, and if $p_{c_1} < p_0$, then Sender B will be better off sampling c_1 than not sampling.

Suppose that there is a case type with probability p_c higher than p_0. According to Lemma 2, if Sender B does not sample, then Sender A's best response is to sample at full capacity, choosing a case-type collection composed of only one case type. If the generated expected posterior is below the success probabilities of all available case types, then Sender B's best response remains not sampling since any sampled case type would only increase the expected value of the receiver's final estimate. However, if there is a case type c_m whose probability of success p_{c_m} is lower than this revised probability, then Sender B's best response is to sample c_m as it lowers the revised probability. This will result in a revised probability of success that is still higher than p_{c_m}. Consequently, Sender A, will respond by sampling a type that keeps the revised probability above p_{c_m} (which may or may not be the same case type as in the previous stage); in turn, Sender B will still want to sample. The same analysis holds when the roles of the senders are reversed (i.e. when there is a case type with probability lower than p_0). Hence, there cannot exist a cycle of best responses where senders switch back and forth between sampling and not sampling. To prove the existence of a pure-strategy equilibrium, it thus suffices to rule out the possibility of best-response cycles where both senders sample.

By contradiction, suppose that there is no equilibrium in pure strategies. Then there should exist $r \in \mathbb{N}$ pairs of strategies that form a cycle of best responses that strictly improve on one another. That is, there should exist pairs of case types $c_j^A, c_j^B,\ j = 1, \dots, r$, where $c_r^A = c_1^A, c_r^B = c_1^B$, that satisfy, for all $j = 1, \dots, r-1$:

- Sampling $D_{c_j^A}^{N^A}$ is a best response of Sender A to a sample of $D_{c_j^B}^{N^B}$ by Sender B.

- Sampling $D_{c_{j+1}^B}^{N^B}$ is a best response of Sender B to a sample of $D_{c_j^A}^{N^A}$ by Sender A.

- Sampling $D_{c_{j+1}^A}^{N^A}$ in response to a sample of $D_{c_{j+1}^B}^{N^B}$ by Sender B is strictly better for Sender A than sampling $D_{c_j^A}^{N^A}$ (otherwise Sender A sampling $D_{c_j^A}^{N^A}$ and Sender B sampling $D_{c_{j+1}^B}^{N^B}$ would be a pure strategy equilibrium).

- Sampling $D_{c_{j+1}^B}^{N^B}$ in response to a sample of $D_{c_j^A}^{N^A}$ by Sender A is strictly better for Sender B than sampling $D_{c_j^B}^{N^B}$.

If Sender A samples collection $D^{N^A}_{c^A}$ and Sender B samples collection $D^{N^B}_{c^B}$, then the revised posterior probability of success is,

$$\hat{p}_0(c^A, c^B) = \frac{N^A p_{c^A} s_{0c^A} + N^B p_{c^B} s_{0c^B} + p_0 S(c_0, \mathbf{H})}{N^A s_{0c^A} + N^B s_{0c^B} + S(c_0, \mathbf{H})} .$$

The above conditions on senders' strategy pairs imply that, for $j = 1, \ldots, r-1$, and for any case type c,

$$\begin{aligned} \hat{p}_0(c^A_j, c^B_j) &\geq \hat{p}_0(c, c^B_j) \\ \hat{p}_0(c^A_j, c) &\geq \hat{p}_0(c^A_j, c^B_{j+1}) \\ \hat{p}_0(c^A_{j+1}, c^B_{j+1}) &> \hat{p}_0(c^A_j, c^B_{j+1}) \\ \hat{p}_0(c^A_j, c^B_j) &> \hat{p}_0(c^A_j, c^B_{j+1}) . \end{aligned}$$

Therefore, according to the first inequality, we also have $\hat{p}_0(c^A_j, c^B_j) \geq \hat{p}_0(c^A_{j+1}, c^B_j)$ for all $j = 1, \ldots, r-1$.

We first show that whenever a cycle as above exists, then there must exist $j \in \{1, \ldots, r-1\}$ for which $\hat{p}_0(c^A_{j+1}, c^B_{j+1}) \geq \hat{p}_0(c^A_{j+1}, c^B_j)$. Suppose, on the contrary, that there does not exist such a j. Then for all $j = 1, \ldots, r-1$, $\hat{p}_0(c^A_{j+1}, c^B_j) > \hat{p}_0(c^A_{j+1}, c^B_{j+1})$, implying, by the above inequality, that for all $j = 1, \ldots, r-1$, $\hat{p}_0(c^A_j, c^B_j) > \hat{p}_0(c^A_{j+1}, c^B_{j+1})$. As $c^A_r = c^A_1$ and $c^B_r = c^B_1$, we obtain a contradiction. Therefore, for some $j \in \{1, \ldots, r-1\}$, the following holds:

$$\hat{p}_0(c^A_j, c^B_j) \geq \hat{p}_0(c^A_{j+1}, c^B_j) \tag{12}$$

$$\hat{p}_0(c^A_j, c^B_j) > \hat{p}_0(c^A_j, c^B_{j+1}) \tag{13}$$

$$\hat{p}_0(c^A_{j+1}, c^B_{j+1}) > \hat{p}_0(c^A_j, c^B_{j+1}) \tag{14}$$

$$\hat{p}_0(c^A_{j+1}, c^B_{j+1}) \geq \hat{p}_0(c^A_{j+1}, c^B_j) . \tag{15}$$

We show that these four conditions cannot be simultaneously satisfied. We define

$$\begin{aligned} E_k &:= N^A p_{c^A_k} s_{0c^A_k} + p_0 S(c_0, \mathbf{H}) , \\ F_k &:= N^A s_{0c^A_k} + S(c_0, \mathbf{H}) , \\ e_k &:= N^B p_{c^B_k} s_{0c^B_k} , \\ f_k &= N^B s_{0c^B_k} . \end{aligned}$$

Using this notation, inequalities (12-15) can be written as,

$$\begin{aligned} \frac{E_j + e_j}{F_j + f_j} &\geq \frac{E_{j+1} + e_j}{F_{j+1} + f_j} \\ \frac{E_j + e_j}{F_j + f_j} &> \frac{E_j + e_{j+1}}{F_j + f_{j+1}} \\ \frac{E_{j+1} + e_{j+1}}{F_{j+1} + f_{j+1}} &> \frac{E_j + e_{j+1}}{F_j + f_{j+1}} \\ \frac{E_{j+1} + e_{j+1}}{F_{j+1} + f_{j+1}} &\geq \frac{E_{j+1} + e_j}{F_{j+1} + f_j} \ . \end{aligned}$$

Multiplying by the denominators and summing over the inequalities, we obtain

$$(E_j + E_{j+1} + e_j + e_{j+1})(F_j + F_{j+1} + f_j + f_{j+1}) > (E_j + E_{j+1} + e_j + e_{j+1})(F_j + F_{j+1} + f_j + f_{j+1}) \ .$$

A contradiction. We thus conclude that a cycle of strictly improving best responses cannot exist. Therefore, there exists an equilibrium in pure strategies.

A.8 Proof of Proposition 6

If all case types $c \in \mathcal{C}$ satisfy $p_c = p_0$, the proof is trivial. Note that as long as there exists a case type $c \in \mathcal{C}$ whose probability $p_c \geq p_0$ ($p_c \leq p_0$) then in response to any action of Sender B (resp., A), not sampling is weakly dominated by sampling a case type with the maximal (resp., minimal) probability in $\mathcal{C}$. Therefore, if a sender chose to sample a positive number of observations in isolation, he will certainly choose to sample when facing an opponent who is also allowed to sample.

Suppose now there exists a case type $c \in \mathcal{C}$ such that $p_c > p_0$. Then, in isolation, Sender A would sample N^A instances of case type $\bar{c}^*$. This will also be Sender A's equilibrium strategy if, in equilibrium, Sender B does not sample. We thus consider equilibria where both Sender A and Sender B choose to sample a positive number of observations. We wish to show that in this case, Sender A samples a case type c with $s_{0c} \geq s_{0\bar{c}^*}$. As before, we denote by D_c^N the case-type collection with N instances of case type c. The receiver's expected posterior when Sender A samples case type $\bar{c}^*$ and Sender B samples N^B instances of case type

c' is:

$$\frac{N^B s_{0c'} p_{c'} + N^A s_{0\bar{c}^*} p_{\bar{c}^*} + S(c_0, \mathbf{H}) p_0}{N^B s_{0c'} + N^A s_{0\bar{c}^*} + S(c_0, \mathbf{H})} = \frac{N^B s_{0c'}}{N^B s_{0c'} + N^A s_{0\bar{c}^*} + S(c_0, \mathbf{H})} p_{c'} + \frac{N^A s_{0\bar{c}^*} + S(c_0, \mathbf{H})}{N^B s_{0c'} + N^A s_{0\bar{c}^*} + S(c_0, \mathbf{H})} \hat{p}_0(D_{\bar{c}^*}^{N^A}). \tag{16}$$

Thus, the resulting expected posterior of the receiver is a weighted average of $p_{c'}$ and $\hat{p}_0(D_{\bar{c}^*}^{N^A})$. By Proposition 5, we know that Sender B would never choose a case type c' such that $p_{c'} > \hat{p}_0(D_{\bar{c}^*}^{N^A})$; hence, it must be that $p_{c'} \leq \hat{p}_0(D_{\bar{c}^*}^{N^A})$.

We argue next that since $p_{c'} \leq \hat{p}_0(D_{\bar{c}^*}^{N^A})$, choosing case type $\bar{c}^*$ for Sender A weakly dominates choosing any other case type c'' such that $p_{c''} > p_{\bar{c}^*}$. Since $\bar{c}^*$ is optimal for Sender A when he acts in isolation, we have $\hat{p}_0(D_{c''}^{N^A}) \leq \hat{p}_0(D_{\bar{c}^*}^{N^A})$; furthermore, it must be that $s_{0c''} < s_{0\bar{c}^*}$, otherwise choosing $\bar{c}^*$ would always be dominated by c''. Therefore, if Sender A switches from $\bar{c}^*$ to c'', then the weight on $p_{c'}$—the first probability in the weighted average in (16)—would increase, while the second probability in this weighted average would decrease from $\hat{p}_0(D_{\bar{c}^*}^{N^A})$ to $\hat{p}_0(D_{c''}^{N^A})$, leading to an unambiguous decline of the receiver's expected posterior. Thus, in equilibrium, Sender A samples a case type c'' whose probability is $p_{c''} \leq p_{\bar{c}^*}$. If $s_{0c''} < s_{0\bar{c}^*}$, sampling a database with case type c'' is dominated by sampling $\bar{c}^*$. Hence in equilibrium, Sender A will choose case type c'' for which $s_{0c''} \geq s_{0\bar{c}^*}$.

An analogous argument follows for Sender B if there exists a case type $c \in \mathcal{C}$ such that $p_c < p_0$. This proves the proposition for the case where $\bar{c}^* \neq c_\emptyset$ (resp. $\underline{c}^* \neq c_\emptyset$). On the other hand, if a sender does not sample when acting in isolation ($\bar{c}^* = c_\emptyset$ or $\underline{c}^* = c_\emptyset$), then the similarity of his sampled database when acting with the other sender can only increase.

References

AKAIKE, H. (1954): "An approximation to the density function," *Annals of the Institute of Statistical Mathematics*, 6, 127–132.

ALON, S. AND G. GAYER (2020): "From ordinal ordering to a prior probability," Manuscript.

ALONSO, R. AND O. CÂMARA (2016): "Bayesian persuasion with heterogeneous priors," *Journal of Economic Theory*, 165, 672–706.

ARGENZIANO, R. AND I. GILBOA (2019): "Second-Order Induction in Prediction Problems," *PNAS*, 116.

AU, P. H. AND K. KAWAI (2020): "Competitive information disclosure by multiple senders," *Games and Economic Behavior*, 119, 56–78.

BALL, I. AND J.-A. ESPÍN-SÁNCHEZ (2022): "Experimental Persuasion," *Cowles Foundation Research paper 2298.*

BEAUCHÊNE, D., J. LI, AND M. LI (2019): "Ambiguous persuasion," *Journal of Economic Theory*, 179, 312–365.

BILLOT, A., I. GILBOA, D. SAMET, AND D. SCHMEIDLER (2005): "Probabilities as Similarity-Weighted Frequencies," *Econometrica*, 73, 1125–1136.

BLONSKI, M. (1999): "Social learning with case-based decisions," *Journal of Economic Behavior & Organization*, 38, 59–77.

BROCAS, I. AND J. D. CARRILLO (2007): "Influence through Ignorance," *The RAND Journal of Economics*, 38, 931–947.

CHENG, X. (2021): "Ambiguous Persuasion Under Dynamic Consistency," *working paper.*

DE CLIPPEL, G. AND X. ZHANG (2022): "Non-Bayesian Persuasion," *Journal of Political Economy*, 130, 2594–2642.

DI TILLIO, A., M. OTTAVIANI, AND P. N. SØRENSEN (2021): "Strategic Sample Selection," *Econometrica*, 89, 911–953.

DWORCZAK, P. AND A. PAVAN (2022): "Preparing for the Worst but Hoping for the Best: Robust (Bayesian) Persuasion," *Econometrica*, 90, 2017–2051.

Eliaz, K., R. Spiegler, and H. C. Thysen (2021): "Strategic interpretations," *Journal of Economic Theory*, 192, 105192.

Galperti, S. (2019): "Persuasion: The Art of Changing Worldviews," *American Economic Review*, 109, 996–1031.

Gayer, G. (2010): "Perception of probabilities in situations of risk: A case based approach," *Games and Economic Behavior*, 68, 130–143.

Gentzkow, M. and E. Kamenica (2017): "Competition in persuasion," *Review of Economic Studies*, 84, 300–322.

Gilboa, I. and D. Schmeidler (1995): "Case-Based Decision Theory," *The Quarterly Journal of Economics*, 110, 605–639.

——— (2001): *A Theory of Case-Based Decisions*, Cambridge University Press.

Glazer, J. and A. Rubinstein (2001): "Debates and Decisions: On a Rationale of Argumentation Rules," *Games and Economic Behavior*, 36, 158–173.

——— (2004): "On Optimal Rules of Persuasion," *Econometrica*, 72, 1715–1736.

——— (2006): "A study in the pragmatics of persuasion: a game theoretical approach," *Theoretical Economics*, 1, 395–410.

Golosnoy, V. and Y. Okhrin (2008): "General uncertainty in portfolio selection: A case-based decision approach," *Journal of Economic Behavior & Organization*, 67, 718–734.

Guerdjikova, A. (2006): "Evolution of wealth and asset prices in markets with case-based investors," *mimeo.*

Henry, E. and M. Ottaviani (2019): "Research and the Approval Process: The Organization of Persuasion," *American Economic Review*, 109, 911–955.

Holmström, B. (1999): "Managerial incentive problems: A dynamic perspective," *The review of Economic studies*, 66, 169–182.

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

Kamenica, E., K. Kim, and A. Zapechelnyuk (2021): "Bayesian persuasion and information design: perspectives and open issues," *Economic Theory*, 72, 701–704.

Koessler, F., M. Laclau, J. Renault, and T. Tomala (2021): "Long Information Design," *Theoretical Economics.*

Kosterina, S. (2022): "Persuasion with unknown beliefs," *Theoretical Economics*, 17, 1075–1107.

Laclau, M. and L. Renou (2017): "Public Persuasion," *working paper.*

Le Treust, M. and T. Tomala (2019): "Persuasion with limited communication capacity," *Journal of Economic Theory*, 184: 104940.

Liu, Z. and N. C. Yannelis (2021): "Persuasion in an asymmetric information economy: a justification of Wald's maxmin preferences," *Economic Theory*, 72, 801–833.

McClellan, A. (2021): "Experimentation and Approval Mechanisms," *Econometrica, forthcoming.*

Parzen, E. (1962): "On Estimation of a Probability Density Function and Mode," *Annals of Mathematical Statistics*, 33, 1065–1076.

Patil, S. and Y. Salant (2023): "Optimal Sample Sizes and Statistical Decision Rules," *Theoretical Economics, forthcoming.*

Perez-Richet, E. (2014): "Interim Bayesian Persuasion: First Steps," *American Economic Review: Papers & Proceedings*, 104, 469–474.

Rosenblatt, M. (1956): "Remarks on Some Nonparametric Estimates of a Density Function," *Annals of Mathematical Statistics*, 27, 832–837.

Silverman, B. W. (1986): *Density Estimation for Statistics and Data Analysis*, London and New York: Chapman and Hall.