Bayesian Persuasion for Algorithmic Recourse

Keegan Harris
Carnegie Mellon University
Pittsburgh, United States
keeganh@cmu.edu

Valerie Chen
Carnegie Mellon University
Pittsburgh, United States
valeriechen@cmu.edu

Joon Sik Kim
Carnegie Mellon University
Pittsburgh, United States
joonkim@cmu.edu

Ameet Talwalkar
Carnegie Mellon University
Pittsburgh, United States
talwalkar@cmu.edu

Hoda Heidari
Carnegie Mellon University
Pittsburgh, United States
hheidari@cmu.edu

Zhiwei Steven Wu
Carnegie Mellon University
Pittsburgh, United States
zstevenwu@cmu.edu

ABSTRACT

When subjected to automated decision-making, decision subjects may strategically modify their observable features in ways they believe will maximize their chances of receiving a favorable decision. In many practical situations, the underlying assessment rule is deliberately kept secret to avoid gaming and maintain competitive advantage. The resulting opacity forces the decision subjects to rely on *incomplete information* when making strategic feature modifications. We capture such settings as a game of *Bayesian persuasion*, in which the decision maker offers a form of recourse to the decision subject by providing them with an action recommendation (or *signal*) to incentivize them to modify their features in desirable ways. We show that when using persuasion, both the decision maker and decision subject are *never worse off* in expectation, while the decision maker can be *significantly better off*. While the decision maker's problem of finding the optimal Bayesian incentive-compatible (BIC) *signaling policy* takes the form of optimization over infinitely-many variables, we show that this optimization can be cast as a linear program over finitely-many regions of the space of possible assessment rules. While this reformulation simplifies the problem dramatically, solving the linear program requires reasoning about exponentially-many variables, even under relatively simple settings. Motivated by this observation, we provide a polynomial-time approximation scheme that recovers a near-optimal signaling policy. Finally, our numerical simulations on semi-synthetic data empirically illustrate the benefits of using persuasion in the algorithmic recourse setting.

KEYWORDS

Algorithmic Recourse, Bayesian Persuasion, Mechanism Design, Strategic Learning

1 INTRODUCTION

High-stakes decision-making systems increasingly utilize data-driven algorithms to assess individuals in such domains as education [28], employment [7, 33], and lending [22]. Individuals subjected to these assessments (henceforth, decision subjects) may strategically modify their observable features in ways they believe maximize their chances of receiving favorable decisions [9, 20]. The decision subject often has a set of actions/interventions available to them. Each of these actions leads to some measurable effect on their observable features, and subsequently, their decision. From the decision maker's perspective, some of these actions may be more desirable than others. Consider credit scoring as an example.[1] Credit scores predict how likely an individual applicant is to pay back a loan on time. Financial institutions regularly utilize credit scores to decide whether to offer applicants their financial products and determine the terms and conditions of their offers (e.g., by setting the interest rate or credit limit). Given their (partial) knowledge of credit scoring instruments, applicants regularly attempt to improve their scores. For instance, a business applying for a loan may improve its score by paying off existing debt or cleverly manipulating its financial records to appear more profitable. While both of these interventions may improve credit score, the former is more desirable than the latter from the perspective of the financial institution offering the loan. The question we are interested in answering in this work is: *how can the decision maker incentivize decision subjects to take such beneficial actions while discouraging manipulations?*

The strategic interactions between decision-making algorithms and decision subjects has motivated a growing literature known as *strategic learning* (see e.g., [12, 18, 19, 27, 37]). While much of the prior work in strategic learning operates under the assumption of full transparency (i.e., the assessment rule is public knowledge), we consider settings where the full disclosure of the assessment rule is not a viable alternative. In many real-world situations, revealing the exact logic of the decision rule is either infeasible or irresponsible. For instance, credit scoring formulae are closely guarded trade secrets, in part to prevent the risk of default rates surging if applicants learn how to manipulate them. In such settings, the decision maker may still have a vested interest in providing *some* information about the decision rule to decision subjects to provide a certain level of *transparency* and *recourse*. In particular, the decision maker may be legally obliged, or economically motivated, to guide decision subjects to take actions that improve their underlying qualifications. To do so, the decision maker can *recommend actions* for decision subjects to take. Of course, such recommendations need to be chosen carefully and credibly; otherwise, self-interested decision

Appears at the 1st Workshop on Learning with Strategic Agents (LSA 2022). Held as part of the Workshops at the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), N. Bishop, M. Han, L. Tran-Thanh, H. Xu, H. Zhang (chairs), May 9–10, 2022, Online. 2022.

[1] Other examples of strategic settings which arise as a result of decision-making include college admissions, in which a college/university (decision maker) decides whether or not to admit a prospective student (decision subject), hiring, in which a company decides whether or not to hire a job applicant, and lending, in which a banking institution decides to accept or reject someone applying for a loan. Oftentimes, the decision maker is aided by automated decision-making tools in these situations (e.g., [22, 28, 35]).

subjects may not follow them or, even worse, they may utilize the recommendations to find pathways for manipulation.

We study a model of strategic learning in which the underlying assessment rule is not revealed to decision subjects. Our model captures several key aspects of the setting described above: First, even though the assessment rule is not revealed to the decision subjects, they often have *prior knowledge* about what the rule may be. Secondly, when the decision maker provides recommendations to decision subjects on which action to take, the recommendations should be *compatible with the subjects' incentives* to ensure they will follow the recommendation. Finally, our model assumes the decision maker discloses how they generate recommendations for recourse—an increasingly relevant requirement under recent regulations (e.g., [10]).

Utilizing our model, we aim to design a mechanism for a decision maker to provide recourse to a decision subject with incomplete information about the underlying assessment rule. We assume the assessment rule makes *predictions* about some future outcome of the decision subject (e.g., whether they pay back the loan in time if granted). Before the assessment rule is trained (i.e., before the model parameters are fit), the decision maker and decision subject have some *prior belief* about the realization of the assessment rule. This prior represents the "common knowledge" about the importance of various observable features for making accurate predictions. After training, the assessment rule is revealed to the decision maker, who then recommends an action for the decision subject to take, based on their pre-determined *signaling policy*. Upon receiving this action recommendation, the decision subject updates their belief about the underlying assessment rule. They then take the action which they believe (according to the update belief) will maximize their expected utility (i.e., the benefit from the decision they receive, minus the cost of taking their selected action). Finally, the decision maker uses the assessment rule to make a prediction about the decision subject.

The interaction described above is an instance of *Bayesian persuasion*, a game-theoretic model of information revelation originally due to Kamenica and Gentzkow. For background on the general Bayesian persuasion model, see Section 1.1. The specific instance of Bayesian persuasion we consider in this work is summarized below.

Interaction protocol for our setting

(1) Before training, the decision maker and decision subject have some prior/belief about the true assessment rule.
(2) After training, the assessment rule is revealed to the decision maker.
(3) The decision maker then uses their signaling policy and knowledge of the assessment rule to recommend an action for the decision subject to take.
(4) The decision subject updates their belief given the recommendation. They then take a (possibly different) action, and receive a prediction through the assessment rule.

Our contributions. Our central conceptual contribution is to cast the problem of offering recourse under partial transparency as a game of Bayesian persuasion. Our key technical contributions consist of comparing optimal action-recommendation policies in this new setup with two natural alternatives: (1) fully revealing the assessment rule to the decision subjects, or (2) revealing no information at all about the assessment rule. We provide new insights about the potentially significant advantages of action recommendation over these baselines, and offer efficient formulations to derive the optimal recommendations. More specifically, our analysis offers the following takeaways:

(1) Using tools from Bayesian persuasion, we show that it is possible for the decision maker to provide incentive-compatible action recommendations that encourage rational decision subjects to modify their features through beneficial interventions. Perhaps most importantly, we show that the optimal signaling policy is more effective than the above two baselines in encouraging positive interventions on the part of the decision subjects (Section 3).
(2) While the decision maker and decision subjects are never worse off in expectation from using optimal incentive-compatible recommendations, we show that situations exist in which the decision maker is *significantly better off* in expectation utilizing the optimal signaling policy (as opposed to the two baselines) (Section 3).
(3) We derive the optimal signaling policy for the decision maker. While the decision maker's optimal signaling policy initially appears challenging (as it involves optimizing over *continuously-many* variables), we show that the problem can naturally be cast as a linear program (Section 4).
(4) We show that even for relatively simple examples, solving this linear program requires reasoning about exponentially-many variables. Motivated by this observation, we provide a polynomial-time algorithm to approximate the optimal signaling policy up to additive terms (Section 5).
(5) Finally, we empirically evaluate our persuasion mechanism on semi-synthetic data based on the Home Equity Line of Credit (HELOC) dataset, and find that the optimal signaling policy performs significantly better than the two natural alternatives in practice (Appendix D).

1.1 Related Work

Bayesian Persuasion. In its most basic form, Bayesian persuasion [25] is modeled as a game between a *sender* (with private information) and a *receiver*. At the beginning of the game, the sender and receiver share a *prior* over some unknown *state of nature*, which will eventually be revealed to the sender. Before the state of nature is revealed, the sender commits to a *signaling policy*, a (probabilistic) mapping from states of nature to action recommendations.[2] After the sender commits to a signaling policy, the state of nature is revealed to the sender, who then sends a signal (according to their policy) to the receiver. The receiver uses this signal to form a posterior over the possible states of nature, and then takes an action which affects the payoffs of both players. Several extensions to the original Bayesian persuasion model have been proposed,

[2] Such commitment is especially possible when the sender is a software agent (as is the case in our setting), since the agent is committed to playing the policy prescribed by its code once it is deployed.

including persuasion with multiple receivers [3], persuasion with multiple senders [29], and persuasion with heterogeneous priors [2]. There has been growing interest in persuasion in the computer science and machine learning communities in recent years. Dughmi and Xu [13, 14] characterize the computational complexity of computing the optimal signaling policy for several popular models of persuasion. Castiglioni et al. [6] study the problem of learning the receiver's utilities through repeated interactions. Work in the multi-arm bandit literature [8, 21, 30, 31, 36] leverages techniques from Bayesian persuasion to incentivize agents to perform bandit exploration.

Strategic responses to unknown predictive models. To the best of our knowledge, our work is the first to use tools from persuasion to model the strategic interaction between a decision maker and strategic decision subjects when the underlying predictive model is not public knowledge. Several prior articles have addressed similar problems through different models and techniques. For example, Akyol et al. [1] quantify the "price of transparency", a quantity which compares the decision maker's utility when the predictive model is fully known with their utility when the model is not revealed to the decision subjects. Ghalme et al. [17] compare the prediction error of a classifier when it is public knowledge with the error when decision subjects must learn a version of it, and label this difference the "price of opacity". Bechavod et al. [4] study the effects of information discrepancy across different sub-populations of decision subjects on their ability to improve their observable features in strategic learning settings. Like us, they do not assume the predictive model is fully known to the decision subjects. Instead, the authors model decision subjects as trying to infer the underlying predictive model by learning from their social circle of family and friends, which naturally causes different groups to form within the population. Additionally, while the models proposed by [4, 17] circumvent the assumption of full information about the deployed model, they restrict the decision subjects' knowledge to be obtained only through past data.

Algorithmic recourse. Our work is closely related to recent work on algorithmic recourse [26]. Algorithmic recourse is concerned with providing explanations and recommendations to individuals who are unfavorably treated by automated decision-making systems. A line of algorithmic recourse methods including [23, 39, 40] focus on finding recourses that are *actionable*, or realistic, for decision subjects to take to improve their decision. In contrast, our action recommendations are "actionable" in the sense that they are interventions which promote long-term desirable behaviors while ensuring that the decision subject is not worse off in expectation. Finally, more recent work [38] shows that existing recourse methods based on counterfactual approaches are not robust to manipulations. Our approach to recourse is not counterfactual-based and instead uses a Bayesian persuasion mechanism to ensure decision subject compliance.

2 SETTING AND BACKGROUND

Consider a setting in which a decision maker assigns a predicted label $\hat{y} \in \{-1, +1\}$ (e.g., whether or not someone will repay a loan if granted one) to a decision subject with observable features $\mathbf{x}_0 = (x_{0,1}, \cdots, x_{0,d-1}, 1) \in \mathbb{R}^d$ (e.g., amount of current debt, bank account balance, etc.).[3] We assume the decision maker uses a linear decision rule to make predictions, i.e., $\hat{y} = \text{sign}\{\mathbf{x}_0^\top \boldsymbol{\theta}\}$, where the assessment rule $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^d$ is chosen by the decision maker. The goal of the decision subject is to receive a positive classification (e.g., get approved for a loan). Given this goal, the decision subject may choose to take some *action* a from some set of possible actions $\mathcal{A}$ to modify their observable features (for example, they may decide to pay off a certain amount of existing debt, or redistribute their debt to game the credit score). We assume that the decision subject has m actions $\{a_1, a_2, \ldots a_m\} \in \mathcal{A}$ at their disposal in order to improve their outcomes. For convenience, we add $a_\emptyset$ to $\mathcal{A}$ to denote taking "no action". By taking action a, the decision subject incurs some *cost* $c(a) \in \mathbb{R}$. This could be an actual monetary cost, but it can also represent non-monetary notions of cost such as opportunity cost or the time/effort cost the decision subject may have to exert to take the action. We assume taking an action a changes a decision subject's observable feature values from $\mathbf{x}_0$ to $\mathbf{x}_0 + \Delta\mathbf{x}(a)$, where $\Delta\mathbf{x}(a) \in \mathbb{R}^d$, and $\Delta\mathbf{x}_j(a)$ specifies the change in the jth observable feature as the result of taking action a. For the special case of $a_\emptyset$, we have $\Delta\mathbf{x}(a_\emptyset) = \mathbf{0}$, $c(a_\emptyset) = 0$. As a result of taking action a, a decision subject, *ds*, receives utility $u_{\text{ds}}(a, \boldsymbol{\theta}) = \text{sign}\{(\mathbf{x}_0 + \Delta\mathbf{x}(a))^\top \boldsymbol{\theta}\} - c(a)$. In other words, the decision subject receives some positive (negative) utility for a positive (negative) classification, subject to some *cost* for taking said action.

If the decision subject had exact knowledge of the assessment rule $\boldsymbol{\theta}$ used by the decision maker, they could solve an optimization problem to determine the best action to take in order to maximize their utility. However, in many settings it is not realistic for a decision subject to have perfect knowledge of $\boldsymbol{\theta}$. Instead, we model the decision subject's information through a *prior* Π over $\boldsymbol{\theta}$, which can be thought of as "common knowledge" about the relative importance of each observable feature to the classifier. We will use $\pi(\cdot)$ to denote the probability density function of Π (so that $\pi(\boldsymbol{\theta})$ denotes the probability of the deployed assessment rule being $\boldsymbol{\theta}$). We assume the decision subject is rational and risk-neutral. So at any point during the interaction, if they hold a belief Π' about the underlying assessment rule, they would pick an action a^* that maximize their *expected* utility with respect to that belief. More precisely, they solve $a^* \in \arg\max_{a \in \mathcal{A}} \mathbb{E}_{\boldsymbol{\theta} \sim \Pi'}[u_{\text{ds}}(a, \boldsymbol{\theta})]$.

From the decision maker's perspective, some actions may be more desirable than others. For example, a bank may prefer that an applicant pay off more existing debt than less when applying for a loan. To formalize this notion of action preference, we say that the decision maker receives some utility $u_{dm}(a) \in \mathbb{R}$ when the decision subject takes action a. In the loan example, $u_{dm}(\text{pay off more debt}) > u_{dm}(\text{pay off less debt})$.

2.1 Bayesian Persuasion in the Algorithmic Recourse Setting

The decision maker has an *information advantage* over the decision subject, due to the fact that they know the true assessment rule $\boldsymbol{\theta}$, whereas the decision subject does not. The decision maker may be able to leverage this information advantage to incentivize the decision subject to take a more favorable action (compared to the one they would have taken according to their prior) by recommending

[3] We append a 1 to the decision subject's feature vector for notational convenience.

an action to the decision subject according to a commonly known *signaling policy*.

Definition 2.1 (Signaling Policy). *A signaling policy $\mathcal{S} : \Theta \to \mathcal{A}$ is a (possibly stochastic) mapping from assessment rules to actions.*[4]

We use $\sigma \sim \mathcal{S}(\boldsymbol{\theta})$ to denote the action recommendation sampled from signaling policy $\mathcal{S}$, where σ is a realization from $\mathcal{A}$.

The decision maker's signaling policy is assumed to be fixed and common knowledge. This is because in order for the decision subject to perform a Bayesian update based on the observed recommendation, they must know the signaling policy. Additionally, the decision maker must have the *power of commitment*, i.e., the decision subject must believe that the decision maker will select actions according to their signaling policy. In our setting, this means that the decision maker must commit to their signaling policy before training their assessment rule. This can be seen as a form of transparency, as the decision maker is publicly committing to how they will use their assessment rule to provide action recommendations/recourse before they even train it. For simplicity, we assume that the decision maker shares the same prior beliefs Π as the decision subject over the observable features before the model is trained. These assumptions are standard in the Bayesian persuasion literature (see, e.g., [25, 30, 31]).

In order for the decision subject to be incentivized to follow the actions recommended by the decision maker, the signaling policy $\mathcal{S}$ needs to be *Bayesian incentive-compatible*.

Definition 2.2 (Bayesian incentive-compatibility). *Consider a decision subject* ds *with initial observable features* $\mathbf{x}_0$ *and prior* Π. *A signaling policy $\mathcal{S}$ is Bayesian incentive-compatible (BIC) for* ds *if*

$$\mathbb{E}_{\boldsymbol{\theta}\sim\Pi}[u_{ds}(a, \boldsymbol{\theta}) | \sigma = a] \geq \mathbb{E}_{\boldsymbol{\theta}\sim\Pi}[u_{ds}(a', \boldsymbol{\theta}) | \sigma = a], \tag{1}$$

for all actions $a, a' \in \mathcal{A}$ such that $\mathcal{S}(\boldsymbol{\theta})$ had positive support on $\sigma = a$.

In other words, a signaling policy $\mathcal{S}$ is BIC if, *given that the decision maker recommends action a*, the decision subject's expected utility is at least as high as the expected utility of taking any other action a' under the posterior.

We remark that while for the ease of exposition our model focuses the interactions between the decision maker and a *single* decision subject, our results can be extended to a heterogeneous population of decision subjects. Under such a heterogeneous setting, the decision maker would publicly commit to a method of computing the signaling policy, given a decision subject's initial observable features as input. Once a decision subject arrives, their feature values are observed and the signaling policy is computed.

3 THE MOTIVATION BEHIND PERSUASION

As is the case in the Bayesian persuasion literature [14, 24, 25], the decision maker can in general achieve a higher expected utility with an optimized signaling policy than the utilities had they provided no recommendation or fully disclosed the model. To characterize *how much* leveraging the decision maker's information advantage (by recommending actions according to a BIC signaling policy) may improve their expected utility, we study the following example.

[4]Note that since our model is focused on the decision maker's interactions with a single decision subject, we drop the dependence of σ on the decision subject's characteristics.

Consider a simple setting under which a single decision subject has one observable feature x_0 (e.g., credit score) and two possible actions: $a_\emptyset$ = "do nothing" (i.e., $\Delta x(a_\emptyset) = 0$, $c(a_\emptyset) = 0$, $u_{dm}(a_\emptyset) = 0$) and a_1 = "pay off existing debt" (i.e., $\Delta x(a_1) > 0$, $c(a_1) > 0$, $u_{dm}(a_1) = 1$), which in turn raises their credit score. For the sake of our illustration, we assume credit-worthiness to be a mutually desirable trait, and credit scores to be a good measure of credit-worthiness. We assume the decision maker would like to design a signaling policy to maximize the chance of the decision subject taking action a_1, regardless of whether or not the applicant will receive the loan. In this simple setting, the decision maker's decision rule can be characterized by a single threshold parameter θ, i.e., the decision subject receives a positive classification if $x + \theta \geq 0$ and a negative classification otherwise. Note that while the decision subject does not know the exact value of θ, they instead have some prior over it, denoted by Π.

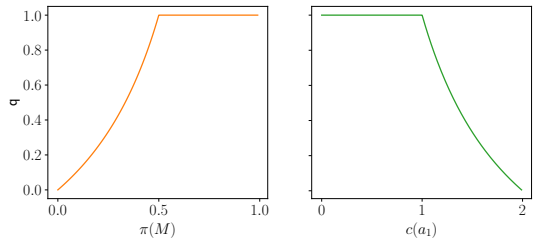
Given the true value of θ, the decision maker recommends an action $\sigma \in \{a_\emptyset, a_1\}$ for the decision subject to take. The decision subject then takes a possibly different action $a \in \{a_\emptyset, a_1\}$, which changes their observable feature from x_0 to $x = x_0 + \Delta x(a)$. Recall that the decision subject's utility takes the form $u_{ds}(a, \theta) = \text{sign}\{(x_0+\Delta x(a))+\theta\} - c(a)$. Note that if $c(a_1) > 2$, then $u_{ds}(a_\emptyset, \theta) > u_{ds}(a_1, \theta)$ holds for any value of θ, meaning that it is impossible to incentivize any rational decision subject to play action a_1. Therefore, in order to give the decision maker a "fighting chance" at incentivizing action a_1, we assume the cost of action a_1 is such that $c(a_1) < 2$.

We observe that in this simple setting, we can bin values of θ into three different "regions", based on the outcome the decision subject would receive if θ were actually in that region. First, if $x_0 + \Delta x(a_1) + \theta < 0$, the decision subject will not receive a positive classification, even if they take action a_1. In this region, the decision subject's initial feature value x_0 is "too low" for taking the desired action to make a difference in their classification. We refer to this region as region L. Second, if $x_0 + \theta \geq 0$, the decision subject will receive a positive classification *no matter what* action they take. In this region, x_0 is "too high" for the action they take to make any difference on their classification. We refer to this region as region H. Third, if $x_0 + \theta < 0$ and $x_0 + \Delta x(a_1) + \theta \geq 0$, the decision subject will receive a positive classification if they take action a_1 and a negative classification if they take action $a_\emptyset$. We refer to this region as region M. Consider the following signaling policy.

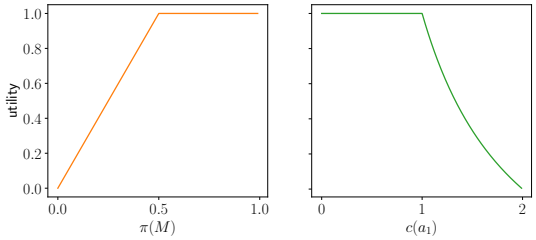
Signaling policy $\mathcal{S}(\theta)$

- **Case 1:** $\theta \in L$. Recommend action a_1 with probability q and action $a_\emptyset$ with probability $1 - q$
- **Case 2:** $\theta \in M$. Recommend action a_1 with probability 1
- **Case 3:** $\theta \in H$. Recommend action a_1 with probability q and action $a_\emptyset$ with probability $1 - q$

In Case 2, $\mathcal{S}$ recommends the action (a_1) that the decision subject would have taken had they known the true θ, with probability 1. However, in Case 1 and Case 3, the decision maker recommends, with probability q, an action (a_1) that the decision subject would not have taken knowing θ, leveraging the fact that the decision subject does not know exactly which case they are currently in. If



(a) q as a function of $\pi(M)$ ($c(a_1) = 1$, left) and $c(a_1)$ ($\pi(M) = \frac{1}{2}$, right).



(b) u as a function of $\pi(M)$ ($c(a_1) = 1$, left) and $c(a_1)$ ($\pi(M) = \frac{1}{2}$, right).

Figure 1: Illustration of how q (the probability of recommending action a_1 when $\theta \notin M$, left) and u (the expected decision maker utility, right) change as a function of $\pi(M)$ and $c(a_1)$. As $\pi(M)$ increases, q and expected utility increase until $\pi(M)$, at which point they remain constant. As $c(a_1)$ increases, q and u remain constant until taking action a_1 becomes prohibitively expensive, at which point both start to decay.

the decision subject follows the decision maker's recommendation from $\mathcal{S}$, then the decision maker expected utility will increase from 0 to q if the realized $\theta \in L$ or $\theta \in H$, and will remain the same otherwise. Intuitively, if q is "small enough" (where the precise definition of "small" depends on the prior over θ and the cost of taking action a_1), then it will be in the decision subject's best interest to follow the decision maker's recommendation, *even though they know that the decision maker may sometimes recommend taking action a_1 when it is not in their best interest to take that action*! That is, the decision maker may recommend that a decision subject pay off existing debt with probability q when it is unnecessary for them to do so in order to secure a loan. We now give a criteria on q which ensures the signaling policy $\mathcal{S}$ is BIC.

Proposition 3.1. *Signaling policy $\mathcal{S}$ is Bayesian incentive-compatible if $q = \min\{\frac{\pi(M)(2-c(a_1))}{c(a_1)(1-\pi(M))}, 1\}$, where $\pi(M) = \pi(x_0 + \theta < 0$ and $x_0 + \Delta x(a_1) + \theta \geq 0)$.*

Proof Sketch. We show that

$$\mathbb{E}_{\theta\sim\Pi}[u_{ds}(a_\emptyset, \theta)|\sigma = a_\emptyset] \geq \mathbb{E}_{\theta\sim\Pi}[u_{ds}(a_1, \theta)|\sigma = a_\emptyset]$$

and

$$\mathbb{E}_{\theta\sim\Pi}[u_{ds}(a_1, \theta)|\sigma = a_1] \geq \mathbb{E}_{\theta\sim\Pi}[u_{ds}(a_\emptyset, \theta)|\sigma = a_1].$$

Since these conditions are satisfied, $\mathcal{S}$ is BIC. See Appendix A for the full proof.

Under this setting, the decision maker will achieve expected utility $u = \pi(M) + q(1 - \pi(M))$. See Figure 1 for an illustration of how q and u vary with $\pi(M)$ and $c(a_1)$.

As we will see in Section 4, the expected utility of the decision maker when recommending actions via the optimal (BIC) signaling policy is trivially *no worse* than their expected utility if they had revealed full information about the assessment rule to the decision subject, or if they had revealed no information and let the decision subject act according to the prior. We now show that the decision maker's expected utility when recommending actions according to the optimal signaling policy can be *arbitrarily higher* than their expected utility from revealing full information or no information.

Theorem 3.2. *For any $\epsilon > 0$, there exists a problem instance such that the expected decision maker utility from recommending actions according to the optimal signaling policy is $1 - \epsilon$ and the expected decision maker utility for revealing full information or revealing no information is at most ϵ.*

See Appendix B for the proof. The decision maker's expected utility as a function of their possible strategies is summarized in Table 1. Note that when $\mathbb{1}\{\pi(M) \geq \frac{c(a_1)}{2}\} = 1$, $q = 1$. Therefore, the decision maker's expected utility is always as least as good as the two natural alternatives of revealing no information about the assessment rule, or revealing full information about the rule.

	No information	Signaling with $\mathcal{S}$	Full information
Decision maker utility	$\mathbb{1}\{\pi(M) \geq \frac{c(a_1)}{2}\}$	$\pi(M) + q(1 - \pi(M))$	$\pi(M)$

Table 1: Decision maker's expected utility when (1) revealing no information about the model, (2) recommending actions according to $\mathcal{S}$, and (3) revealing full information about the model.

4 OPTIMAL SIGNALING POLICY

In Section 3, we show a one-dimensional setting, where a signaling policy can obtain unbounded better utilities compared to revealing full information and revealing no information. We now derive the decision maker's optimal signaling policy for the general setting with arbitrary numbers of observable features and actions described in Section 2. Under the general setting, the decision maker's optimal signaling policy can be described by the following optimization:

$$\begin{aligned} \max_{p(\sigma=a|\boldsymbol{\theta}),\forall a\in\mathcal{A}} \quad & \mathbb{E}_{\sigma\sim\mathcal{S},\boldsymbol{\theta}\sim\Pi}[u_{dm}(\sigma)] \\ \text{s.t.} \quad & \mathbb{E}_{\boldsymbol{\theta}\sim\Pi}[u_{ds}(a, \boldsymbol{\theta}) - u_{ds}(a', \boldsymbol{\theta})|\sigma = a] \geq 0, \ \forall a, a' \in \mathcal{A}, \end{aligned} \tag{2}$$

where we omit the valid probability constraints over $p(\sigma = a|\boldsymbol{\theta}), a \in \mathcal{A}$ for brevity. In words, the decision maker wants to design a signaling policy $\mathcal{S}$ in order to maximize their expected utility, subject to the constraint that the signaling policy is BIC. At first glance, the optimization may initially seem hopeless as there are infinitely many values of $p(\sigma = a|\boldsymbol{\theta})$ (one for every possible $\boldsymbol{\theta} \in \boldsymbol{\Theta}$) that the decision maker's optimal policy must optimize over. However, we will show that the decision maker's optimal policy can actually be recovered by optimizing over finitely many *equivalence regions*.

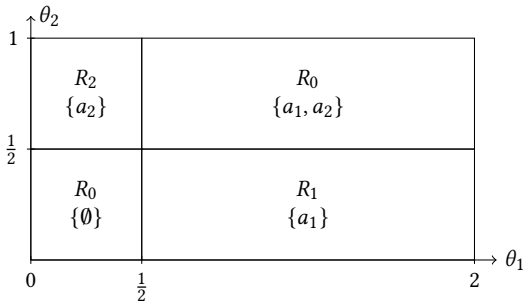


Figure 2: An illustration of the equivalence regions for a two action (a_1, a_2) and two observable feature (x_1, x_2) setting, where $\Theta = [0,2] \times [0,1] \times \{\frac{1}{2}\}$. Consider an individual with $\mathbf{x}_0 = [0,0,1]^\top$, $\Delta\mathbf{x}(a_1) = [1,0,0]^\top$, and $\Delta\mathbf{x}(a_2) = [0,1,0]^\top$. The equivalence regions of Θ are quadrants described the set of actions the decision subject could take in order to receive a positive classification. Region R_0 contains the bottom-left and top-right quadrants of Θ, region R_1 contains the bottom-right quadrant of Θ, and region R_2 contains the top-left quadrant of Θ.

Definition 4.1 (Equivalence Region). *Two assignments θ, θ' are* equivalent *(w.r.t. u_{ds}) if $u_{ds}(a, \theta) - u_{ds}(a', \theta) = u_{ds}(a, \theta') - u_{ds}(a', \theta')$, $\forall a, a' \in \mathcal{A}$. An* equivalence region *R is a subset of Θ such that for any $\theta \in R$, all θ' equivalent to θ are also in R. We denote the set of all equivalence regions by $\mathcal{R}$.*

In Figure 2, we show an example of how different equivalence regions might partition the space of possible assessment rules Θ. In this example, there are two actions and two observable features, and the space of Θ is partitioned into three different equivalence regions. Note that as long as the set of actions $\mathcal{A}$ is finite, $|\mathcal{R}| < \infty$. Using the definition of equivalence region, we are able to simplify the optimization into the following form, where $p(R)$ is the probability under Π that $\boldsymbol{\theta}$ is in equivalence region R and $u_{ds}(a, R)$ is the decision subject's utility of taking action a in equivalence region R.

Theorem 4.2 (Optimal signaling policy). *The decision maker's optimal signaling policy can be characterized by the following linear program OPT-LP:*

$$\max_{p(\sigma=a|R), \forall a \in \mathcal{A}, R \in \mathcal{R}} \quad \sum_{a \in \mathcal{A}} \sum_{R \in \mathcal{R}} p(R) p(\sigma = a|R) u_{dm}(a)$$

$$\text{s.t.} \quad \sum_{R \in \mathcal{R}} p(\sigma = a|R) p(R) (u_{ds}(a, R) - u_{ds}(a', R)) \geq 0, \quad \forall a, a' \in \mathcal{A}$$

$$\sum_{a \in \mathcal{A}} p(\sigma = a|R) = 1, \ \forall R, \quad p(\sigma = a|R) \geq 0, \quad \forall R \in \mathcal{R}, a \in \mathcal{A}, \tag{OPT-LP}$$

where $p(\sigma = a|R)$ denotes the probability of sending recommendation $\sigma = a$ if $\boldsymbol{\theta} \in R$. For the full proof, see Appendix C. Note that the linear program OPT-LP is always feasible, as the decision maker can always recommend the action the decision subject would play according to the prior, which is BIC.

ALGORITHM 1: Approximation Algorithm for (OPT-LP)

Input: $\boldsymbol{\theta} \in \Theta, \epsilon > 0, \delta > 0$

Output: Signaling policy $\widehat{\mathcal{S}} := \{p(\sigma = a|R_{\boldsymbol{\theta}})\}_{\forall a \in \mathcal{A}}$ (where region $R_{\boldsymbol{\theta}}$ contains $\boldsymbol{\theta}$)

Set $K = \left\lceil \frac{2}{\epsilon^2} \log\left(\frac{2(m^2+1)}{\delta}\right) \right\rceil$

Pick $\ell \in \{1, \ldots, K\}$ uniformly at random. Set $\boldsymbol{\theta}_\ell = \boldsymbol{\theta}$.

Sample $\widetilde{\boldsymbol{\Theta}} = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{\ell-1}, \boldsymbol{\theta}_{\ell+1}, \ldots, \boldsymbol{\theta}_K\} \sim \pi(\boldsymbol{\theta})$.

Let $\widetilde{\mathcal{R}}$ denote the set of observed regions. Compute $\tilde{p}(R)$, $\forall R \in \widetilde{\mathcal{R}}$, where $\tilde{p}(R)$ is the empirical probability of $\boldsymbol{\theta}' \in R$.

Solve (APPROX-LP) and return signaling policy $\widehat{\mathcal{S}} := \{p(\sigma = a|R_{\boldsymbol{\theta}})\}_{\forall a \in \mathcal{A}}$.

5 COMPUTING THE OPTIMAL SIGNALING POLICY

In Section 4, we show that the problem of determining the decision maker's optimal signaling policy can be transformed from an optimization over infinitely many variables into an optimization over the set of finitely many equivalence regions $\mathcal{R}$ (Theorem 4.2). However, computing the decision maker's optimal signaling policy by solving (OPT-LP) requires reasoning over *exponentially-many* variables, even in relatively simple settings (see Appendix C.1). Motivated by this result, we aim to design a computationally efficient approximation scheme to compute an approximately optimal signaling policy for the decision maker. In particular, we adapt the sampling-based approximation algorithm of Dughmi and Xu to our setting in order to compute an ϵ-optimal and ϵ-approximate signaling policy in polynomial time, as shown in Algorithm 1. At a high level, Algorithm 1 samples polynomially-many times from the prior distribution over the space of assessment rules, and solves an empirical analogue of (OPT-LP) ((APPROX-LP), see Appendix E). We show that the resulting signaling policy is ϵ-BIC, and is ϵ-optimal with high probability, for any $\epsilon > 0$.

Theorem 5.1. *Algorithm 1 runs in poly(m, $\frac{1}{\epsilon}$) time (where $m = |\mathcal{A}|$), and implements an ϵ-BIC signaling policy that is ϵ-optimal with probability at least $1 - \delta$.*

See Appendix E for the proof.

Bi-criteria approximation. It is important to note that the signaling policy from Algorithm 1 is both ϵ-optimal and ϵ-incentive compatible. While one may wonder whether (i) an ϵ-optimal and exactly incentive compatible signaling policy exists, or (ii) an exactly optimal and ϵ-incentive compatible signaling policy exists, Dughmi and Xu show that this is generally not possible for sampling-based approximation algorithms for Bayesian persuasion (see Theorem 27 in Dughmi and Xu [14]). Note that unlike the other results in Dughmi and Xu [14], these results directly apply to the setting we consider.

Computational complexity. Recall that the algorithm for computing the optimal policy runs in time polynomial in the number of equivalence regions $|\mathcal{R}|$, which can scale exponentially in the number of actions m. However, without any structural assumptions, the input prior over the space of assessment rules Θ can scale exponentially in the number of features d. When m and d

are comparable, our algorithm runs in time polynomial in the input size. We leave open the question of whether there are classes of succinctly represented prior distributions that permit efficient algorithms for computing the optimal policy in time polynomial in d and m. It is also plausible to design efficient algorithms that only require some form of query access to the prior distribution. However, information-theoretic lower bounds of [14] rule out the query access through sampling, as they show that no sampling-based algorithm can compute the optimal signaling policy with finite samples across all problem instances.

6 CONCLUSION

In this work, we investigated the problem of offering algorithmic recourse without requiring full transparency (i.e., revealing the assessment rule). We cast this problem as a game of Bayesian persuasion, and offered several new insights regarding how a decision maker can leverage their information advantage over decision subjects to incentivize mutually beneficial actions. Our stylized model relies on several simplifying assumptions, which suggest important directions for future work:

Alternative models of information design. Cheap talk [11] and verifiable disclosure [15] are two alternative models of information disclosure which may be applicable whenever the sender *does not* have the power to commit to a signaling policy before the state of nature is revealed. As a consequence, the resulting equilibria are often difficult to characterize, and the players may face an equilibrium selection problem. Nevertheless, it may be worthwhile to analyze these alternative games in the algorithmic recourse setting to capture situations in which the decision maker cannot commit to a signaling policy.

Beyond linear decision rules. Finally, we focus on settings with *linear* decision rules and assume all decision subject parameters (e.g., cost function, initial observable features, etc.) are known to the decision maker. We leave it for future work to extend our findings to non-linear decision rules, or settings in which some of the decision subjects' parameters are unknown to the decision maker.

REFERENCES

[1] Emrah Akyol, Cedric Langbort, and Tamer Basar. Price of transparency in strategic machine learning. *arXiv preprint arXiv:1610.08210*, 2016.

[2] Ricardo Alonso and Odilon Câmara. Bayesian persuasion with heterogeneous priors. *Journal of Economic Theory*, 165:672–706, 2016.

[3] Itai Arieli and Yakov Babichenko. Private bayesian persuasion. *Journal of Economic Theory*, 182:185–217, 2019.

[4] Yahav Bechavod, Chara Podimata, Zhiwei Steven Wu, and Juba Ziani. Information discrepancy in strategic learning. *arXiv preprint arXiv:2103.01028*, 2021.

[5] Ralph Allan Bradley and Milton E. Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952. ISSN 00063444. URL http://www.jstor.org/stable/2334029.

[6] Matteo Castiglioni, Andrea Celli, Alberto Marchesi, and Nicola Gatti. Online bayesian persuasion. *Advances in Neural Information Processing Systems*, 33, 2020.

[7] Aaron Chalfin, Oren Danieli, Andrew Hillis, Zubin Jelveh, Michael Luca, Jens Ludwig, and Sendhil Mullainathan. Productivity and selection of human capital with machine learning. *American Economic Review*, 106(5):124–27, 2016.

[8] Bangrui Chen, Peter Frazier, and David Kempe. Incentivizing exploration by heterogeneous users. In *Conference On Learning Theory*, pages 798–818. PMLR, 2018.

[9] Danielle Keats Citron and Frank Pasquale. The scored society: Due process for automated predictions. *Wash. L. Rev.*, 89:1, 2014.

[10] Council of European Union. Council regulation (EU) no 679/2016, 2016. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679.

[11] Vincent P Crawford and Joel Sobel. Strategic information transmission. *Econometrica: Journal of the Econometric Society*, pages 1431–1451, 1982.

[12] Jinshuo Dong, Aaron Roth, Zachary Schutzman, Bo Waggoner, and Zhiwei Steven Wu. Strategic classification from revealed preferences. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 55–70, 2018.

[13] Shaddin Dughmi and Haifeng Xu. Algorithmic persuasion with no externalities. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 351–368, 2017.

[14] Shaddin Dughmi and Haifeng Xu. Algorithmic bayesian persuasion. *SIAM Journal on Computing*, (0):STOC16–68, 2019.

[15] Ronald A Dye. Disclosure of nonproprietary information. *Journal of accounting research*, pages 123–145, 1985.

[16] FICO. Explainable machine learning challenge. https://community.fico.com/s/explainable-machine-learning-challenge, 2018.

[17] Ganesh Ghalme, Vineet Nair, Itay Eilat, Inbal Talgam-Cohen, and Nir Rosenfeld. Strategic classification in the dark. *arXiv preprint arXiv:2102.11592*, 2021.

[18] Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. Strategic classification. In *Proceedings of the 2016 ACM conference on innovations in theoretical computer science*, pages 111–122, 2016.

[19] Keegan Harris, Hoda Heidari, and Zhiwei Steven Wu. Stateful strategic regression. *arXiv preprint arXiv:2106.03827*, 2021.

[20] Tatiana Homonoff, Rourke O'Brien, and Abigail B Sussman. Does knowing your fico score change financial behavior? evidence from a field experiment with student loan borrowers. *Review of Economics and Statistics*, 103(2):236–250, 2021.

[21] Nicole Immorlica, Jieming Mao, Aleksandrs Slivkins, and Zhiwei Steven Wu. Bayesian exploration with heterogeneous agents. In *The World Wide Web Conference*, pages 751–761, 2019.

[22] Julapa Jagtiani and Catharine Lemieux. The roles of alternative data and machine learning in fintech lending: evidence from the lendingclub consumer platform. *Financial Management*, 48(4):1009–1029, 2019.

[23] Shalmali Joshi, Oluwasanmi Koyejo, Warut Vijitbenjaronk, Been Kim, and Joydeep Ghosh. Towards realistic individual recourse and actionable explanations in black-box decision making systems. *arXiv preprint arXiv:1907.09615*, 2019.

[24] Emir Kamenica. Bayesian persuasion and information design. *Annual Review of Economics*, 11:249–272, 2019.

[25] Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615, 2011.

[26] Amir-Hossein Karimi, Gilles Barthe, Bernhard Schölkopf, and Isabel Valera. A survey of algorithmic recourse: definitions, formulations, solutions, and prospects, 2021.

[27] Jon Kleinberg and Manish Raghavan. How do classifiers induce agents to invest effort strategically? *ACM Transactions on Economics and Computation (TEAC)*, 8 (4):1–23, 2020.

[28] Danijel Kučak, Vedran Juričić, and Goran Đambić. Machine learning in education-a survey of current research trends. *Annals of DAAAM & Proceedings*, 29, 2018.

[29] Fei Li and Peter Norman. On bayesian persuasion with multiple senders. *Economics Letters*, 170:66–70, 2018.

[30] Yishay Mansour, Aleksandrs Slivkins, and Vasilis Syrgkanis. Bayesian incentive-compatible bandit exploration. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pages 565–582, 2015.

[31] Yishay Mansour, Aleksandrs Slivkins, Vasilis Syrgkanis, and Zhiwei Steven Wu. Bayesian exploration: Incentivizing exploration in bayesian games. In Vincent

Conitzer, Dirk Bergemann, and Yiling Chen, editors, *Proceedings of the 2016 ACM Conference on Economics and Computation, EC '16, Maastricht, The Netherlands, July 24-28, 2016*, page 661. ACM, 2016. doi: 10.1145/2940716.2940755. URL https://doi.org/10.1145/2940716.2940755.

[32] Colin McDiarmid et al. On the method of bounded differences. *Surveys in combinatorics*, 141(1):148–188, 1989.

[33] Manish Raghavan, Solon Barocas, Jon Kleinberg, and Karen Levy. Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pages 469–481, 2020.

[34] Kaivalya Rawal and Himabindu Lakkaraju. Beyond individualized recourse: Interpretable and interactive summaries of actionable recourses. *Advances in Neural Information Processing Systems*, 33, 2020.

[35] Javier Sánchez-Monedero, Lina Dencik, and Lilian Edwards. What does it mean to'solve'the problem of discrimination in hiring? social, technical and legal perspectives from the uk on automated hiring systems. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pages 458–468, 2020.

[36] Mark Sellke and Aleksandrs Slivkins. The price of incentivizing exploration: A characterization via thompson sampling and sample complexity. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 795–796, 2021.

[37] Yonadav Shavit, Benjamin Edelman, and Brian Axelrod. Causal strategic linear regression. In *International Conference on Machine Learning*, pages 8676–8686. PMLR, 2020.

[38] Dylan Slack, Sophie Hilgard, Himabindu Lakkaraju, and Sameer Singh. Counterfactual explanations can be manipulated. *arXiv preprint arXiv:2106.02666*, 2021.

[39] Berk Ustun, Alexander Spangher, and Yang Liu. Actionable recourse in linear classification. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 10–19, 2019.

[40] Sandra Wachter, Brent Mittelstadt, and Chris Russell. Counterfactual explanations without opening the black box: Automated decisions and the gdpr. *Harv. JL & Tech.*, 31:841, 2017.

A PROOF OF PROPOSITION 3.1

Proof. Based on the decision subject's prior over θ, they can calculate

(1) $\pi(L) = \pi(x_0 + \Delta x(a_1) + \theta < 0)$, i.e., the probability the decision subject is in region L according to the prior
(2) $\pi(M) = \pi(x_0 + \theta < 0 \text{ and } x_0 + \Delta x(a_1) + \theta \geq 0)$, i.e., the probability the decision subject is in region M according to the prior
(3) $\pi(H) = \pi(x_0 + \theta \geq 0)$, i.e., the probability the decision subject is in region H according to the prior

Case 1: $\sigma = a_0$. Given the signal $\sigma = a_\emptyset$, the decision subject's *posterior* probability density function $\pi(\cdot|\sigma = a_\emptyset)$ over L, M, and H will take the form

$$\pi(L|\sigma = a_\emptyset) = \frac{p(\sigma=a_\emptyset|L)\pi(L)}{p(\sigma=a_\emptyset)} = \frac{\pi(L)}{\pi(L)+\pi(H)}$$
$$\pi(M|\sigma = a_\emptyset) = \frac{p(\sigma=a_\emptyset|M)\pi(M)}{p(\sigma=a_\emptyset)} = 0$$
$$\pi(H|\sigma = a_\emptyset) = \frac{p(\sigma=a_\emptyset|H)\pi(H)}{p(\sigma=a_\emptyset)} = \frac{\pi(H)}{\pi(L)+\pi(H)}$$

If the decision subject receives signal $\sigma = a_0$, they know with probability 1 that they are *not* in region M with probability 1. Therefore, they know that taking action a_1 will not change their classification, so they will follow the decision maker's recommendation and take action $a_\emptyset$.

Case 2: $\sigma = a_1$. Given the signal $\sigma = a_1$, the decision subject's posterior density over L, M, and H will take the form

$$\pi(L|\sigma = a_1) = \frac{p(\sigma=a_1|L)\pi(L)}{p(\sigma=a_1)} = \frac{q\pi(L)}{\pi(M)+q(\pi(L)+\pi(H))} = \frac{q\pi(L)}{\pi(M)+q(1-\pi(M))}$$
$$\pi(M|\sigma = a_1) = \frac{p(\sigma=a_1|M)\pi(M)}{p(\sigma=a_1)} = \frac{\pi(M)}{\pi(M)+q(\pi(L)+\pi(H))} = \frac{\pi(M)}{\pi(M)+q(1-\pi(M))}$$
$$\pi(H|\sigma = a_1) = \frac{p(\sigma=a_1|H)\pi(H)}{p(\sigma=a_1)} = \frac{q\pi(H)}{\pi(M)+q(\pi(L)+\pi(H))} = \frac{q\pi(H)}{\pi(M)+q(1-\pi(M))}$$

The decision subject's expected utility of taking actions $a_\emptyset$ and a_1 under the posterior induced by $\sigma = a_1$ are

$$\begin{aligned}&\mathbb{E}_{\theta\sim\Pi}[u_{ds}(a_\emptyset, \theta)|\sigma = a_1]\\ &= \pi(H|\sigma = a_1)\cdot(1-0) + \pi(M|\sigma = a_1)\cdot(-1-0) + \pi(L|\sigma = a_1)\cdot(-1-0)\\ &= \pi(H|\sigma = a_1) - \pi(M|\sigma = a_1) - \pi(L|\sigma = a_1)\end{aligned}$$

and

$$\begin{aligned}&\mathbb{E}_{\theta\sim\Pi}[u_{ds}(a_1, \theta)|\sigma = a_1]\\ &= \pi(H|\sigma = a_1)\cdot(1-c(a_1))\\ &+ \pi(M|\sigma = a_1)\cdot(1-c(a_1)) + \pi(L|\sigma = a_1)\cdot(-1-c(a_1))\end{aligned}$$

In order for $\mathcal{S}$ to be BIC,

$$\mathbb{E}_{\theta\sim\Pi}[u_{ds}(a_1, \theta)|\sigma = a_1] \geq \mathbb{E}_{\theta\sim\Pi}[u_{ds}(a_\emptyset, \theta)|\sigma = a_1].$$

Plugging in our expressions for $\mathbb{E}_{\theta\sim\Pi}[u_{ds}(a_1, \theta)|\sigma = a_1]$ and $\mathbb{E}_{\theta\sim\Pi}[u_{ds}(a_\emptyset, \theta)|\sigma = a_1]$, we see that

$$\begin{aligned}&\pi(H|\sigma = a_1)\cdot(1-c(a_1)) +\\ &\pi(M|\sigma = a_1)\cdot(1-c(a_1)) + \pi(L|\sigma = a_1)\cdot(-1-c(a_1))\\ &\geq \pi(H|\sigma = a_1) - \pi(M|\sigma = a_1) - \pi(L|\sigma = a_1)\end{aligned}$$

After canceling terms and simplifying, we see that

$$-(\pi(L|\sigma = a_1) + \pi(H|\sigma = a_1))c(a_1) + \pi(M|\sigma = a_1)(2 - c(a_1)) \geq 0$$

Next, we plug in for $\pi(L|\sigma = a_1)$, $\pi(M|\sigma = a_1)$, and $\pi(H|\sigma = a_1)$. Note that the denominators of $\pi(L|\sigma = a_1)$, $\pi(M|\sigma = a_1)$, and $\pi(H|\sigma = a_1)$ cancel out.

$$\begin{aligned} &- q(\pi(L) + \pi(H))c(a_1) + \pi(M)(2 - c(a_1)) \\ &= -q(1 - \pi(M))c(a_1) + \pi(M)(2 - c(a_1)) \geq 0 \end{aligned}$$

Solving for q, we see that

$$q \leq \frac{\pi(M)(2 - c(a_1))}{c(a_1)(1 - \pi(M))}.$$

Note that $q \geq 0$ always. Finally, in order for q to be a valid probability, we restrict q such that

$$q = \min\{\frac{\pi(M)(2 - c(a_1))}{c(a_1)(1 - \pi(M))}, 1\}.$$

This completes the proof. □

B PROOF OF THEOREM 3.2

Proof. Consider the example in Section 3.

Expected utility from revealing no information. If the decision subject acts exclusively according to the prior, they will select action a_1 with probability 1 if $\mathbb{E}_{\theta\sim\Pi}[u_{ds}(a_1, \theta)] \geq \mathbb{E}_{\theta\sim\Pi}[u_{ds}(a_\emptyset, \theta)]$ and with probability 0 otherwise. Plugging in our expressions for $\mathbb{E}_{\theta\sim\Pi}[u_{ds}(a_1, \theta)]$ and $\mathbb{E}_{\theta\sim\Pi}[u_{ds}(a_\emptyset, \theta)]$, we see that the decision subject will select action a_1 only if

$$\begin{aligned} &\pi(L)(-1 - c(a_1)) + \pi(M)(1 - c(a_1)) + \pi(H)(1 - c(a_1)) \\ &\geq \pi(L)(-1 - 0) + \pi(M)(-1 - 0) + \pi(H)(1 - 0) \end{aligned}$$

Canceling terms and simplifying, we see that

$$-c(a_1)(\pi(L) + \pi(H)) + \pi(M)(2 - c(a_1)) \geq 0$$

must hold for the decision subject to select action a_1. Finally, substituting $\pi(L) + \pi(H) = 1 - \pi(M)$ gives us the condition $2\pi(M) - c(a_1) \geq 0$. Alternatively, if $\frac{\pi(M)}{c(a_1)} < \frac{1}{2}$, the decision subject will select action $a_\emptyset$ with probability 1. Intuitively, this means that a rational decision subject would take action a_1 if the ratio of $\pi(M)$ (the probability according to the prior that taking action a_1 is in the decision subject's best interest) to $c(a_1)$ (the cost of taking action a_1) is high, and would take action $a_\emptyset$ otherwise.

Expected utility from revealing full information. If the decision maker reveals the assessment rule to the decision subject, they will select action a_1 when $\theta \in M$ and action $a_\emptyset$ otherwise. Therefore since $u_{dm}(a_1) = 1$ and $u_{dm}(a_\emptyset) = 0$, the decision maker's expected utility if they reveal full information is $\pi(M)$.

Expected utility from $\mathcal{S}$. Recall that the decision maker's signaling policy $\mathcal{S}$ from Section 3 sets $q = \min\{\frac{\pi(M)(2-c(a_1))}{c(a_1)(1-\pi(M))}, 1\}$. Under this setting, the decision maker's expected utility is $\min\{1 \cdot \pi(M) + q \cdot (1 - \pi(M)), 1\}$. Substituting in our expression for q and simplifying, we see that the decision maker's expected utility for recommending actions via $\mathcal{S}$ is $\min\{\frac{2\pi(M)}{c(a_1)}, 1\}$.

Suppose that $2\pi(M) = c(a_1)(1 - \epsilon)$ and $c(a_1) = 2\epsilon$, for some small $\epsilon > 0$. The decision maker's expected utility will always be 0 from revealing no information because $\frac{2\pi(M)}{c(a_1)} = 1 - \epsilon < 1$. The decision maker's expected utility from recommending actions via $\mathcal{S}$ will be $\frac{2\pi(M)}{c(a_1)} = 1 - \epsilon$. Since $\pi(M) = \epsilon(1 - \epsilon) < \epsilon$, the decision maker's expected utility from revealing full information will be less than ϵ. Therefore, as ϵ approaches 0, the decision maker's expected utility from revealing full information approaches 0 (the smallest value possible), and the decision maker's expected utility from $\mathcal{S}$ approaches 1 (the highest value possible). This completes the proof. □

C PROOF OF THEOREM 4.2

By rewriting the BIC constraints as integrals over $\boldsymbol{\Theta}$ and applying Bayes' rule, our optimization over $p(\sigma = a|\boldsymbol{\theta}), a \in \mathcal{A}$ takes the following form

$$\begin{aligned} \max_{p(\sigma=a|\boldsymbol{\theta}),\forall a\in\mathcal{A}} \quad & \mathbb{E}_{\sigma\sim\mathcal{S},\boldsymbol{\theta}\sim\Pi}[u_{dm}(\sigma)] \\ \text{s.t.} \quad & \int_{\boldsymbol{\Theta}} p(\sigma = a|\boldsymbol{\theta})\pi(\boldsymbol{\theta})(u_{ds}(a, \boldsymbol{\theta}) - u_{ds}(a', \boldsymbol{\theta}))d\boldsymbol{\theta} \geq 0, \ \forall a, a' \in \mathcal{A}. \end{aligned}$$

Note that if $u_{ds}(a, \boldsymbol{\theta}) - u_{ds}(a', \boldsymbol{\theta})$ is the same for some "equivalence region" $R \subseteq \boldsymbol{\Theta}$ (which we formally define below), we can pull $u_{ds}(a, \boldsymbol{\theta}) - u_{ds}(a', \boldsymbol{\theta})$ out of the integral and instead sum over the different equivalence regions. Intuitively, an equivalence region can be thought of as the set of all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ pairs that are indistinguishable from a decision subject's perspective because they lead to the exact same utility for any possible action the decision subject could take. Based on this idea, we formally define a region of $\boldsymbol{\Theta}$ as follows.

After pulling the decision subject utility function out of the integral, our optimization takes the following form:

$$\begin{aligned} \max_{p(\sigma=a|\boldsymbol{\theta}),\forall a\in\mathcal{A}} \quad & \mathbb{E}_{\sigma\sim\mathcal{S},\boldsymbol{\theta}\sim\Pi}[u_{dm}(\sigma)] \\ \text{s.t.} \quad & \sum_{R\in\mathcal{R}} (u_{ds}(a, R) - u_{ds}(a', R)) \cdot \\ & \int_{\boldsymbol{\theta}\in R} p(\sigma = a|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} \geq 0, \ \forall a, a' \in \mathcal{A}. \end{aligned}$$

Now that the decision subject's utility $u_{ds}(\cdot)$ no longer depends on $\boldsymbol{\theta}$, we can integrate $p(\sigma = a|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$ over each equivalence region R. We denote $p(R)$ as the probability that the true $\boldsymbol{\theta} \in R$ according to the prior.

$$\begin{aligned} \max_{p(\sigma=a|R),\forall a\in\mathcal{A},R\in\mathcal{R}} \quad & \mathbb{E}_{\sigma\sim\mathcal{S},\theta\sim\Pi}[u_{dm}(\sigma)] \\ \text{s.t.} \quad & \sum_{R\in\mathcal{R}} p(\sigma = a|R)\pi(R)(u_{ds}(a, R) - u_{ds}(a, R)) \geq 0, \\ & \forall a, a' \in \mathcal{A}. \end{aligned}$$

Since it is possible to write the constraints in terms of $p(\sigma = a|R)$, $\forall a \in \mathcal{A}, R \in \mathcal{R}$, it suffices to optimize directly over these quantities. The final step is to rewrite the objective. For completeness, we include the constraints which make each $\{p(\sigma = a_1|R), p(\sigma = a_2|R), \ldots, p(\sigma = a_m|R)\}$, $\forall R$ a valid probability distribution.

C.1 Computational Barriers

In this section, we show that even in the setting where each action only affects one observable feature (e.g., as shown in Figure 3), the number of equivalence regions in (OPT-LP) is still exponential in the size of the input. While somewhat simplistic, we believe this action

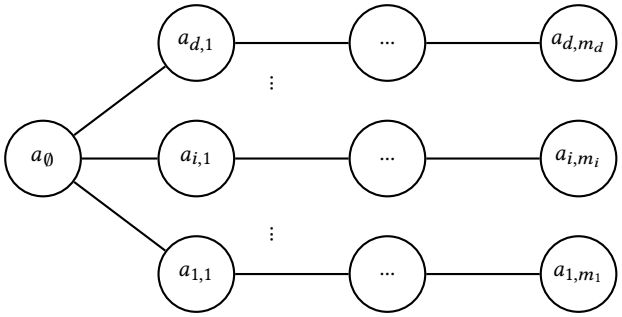


Figure 3: Graphical representation of special ordering over the actions available to each decision subject. Each branch corresponds to an observable feature and each node corresponds to a possible action the decision subject may take. The root corresponds to taking no action (denoted by $a_\emptyset$). Nodes further away from the root on branch i correspond to higher $\Delta\mathbf{x}_i$, i.e., $\Delta\mathbf{x}_i(a_\emptyset) \prec \Delta\mathbf{x}_i(a_{i,1}) \prec \ldots \prec \mathbf{x}_i(a_{i,m_i})$.

scheme reasonably reflects real-world settings in which the decision subjects are under time or resource constraints when deciding which action to take. For example, the decision subject may need to choose between paying off some amount of debt and opening a new credit card when strategically modifying their observable features before applying for a loan.

Under this setting, (OPT-LP) optimizes over $\Theta(m|\mathcal{R}|)$ variables, where m is the number of actions available to each agent and $|\mathcal{R}|$ is the number of equivalence regions. In order to determine the size of $\mathcal{R}$, we note that an equivalence region can be alternatively characterized by observing that assessment rules $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ belong to the same equivalence region if the difference in their predictions for any two actions a and a' is the same. (This follows from straightforward algebraic manipulation of Definition 4.1.) As such, an equivalence region R can essentially be characterized by the set of actions $A_R \subseteq \mathcal{A}$ which receive a positive classification when $\boldsymbol{\theta} \in R$.[5]

Armed with this new characterization of an equivalence region, we are now ready to show the scale of $|\mathcal{R}|$ for the setting described in Figure 3.

Proposition C.1. *For the setting described in Figure 3, there are $|\mathcal{R}| = \Pi_{i=1}^{d} m_i - 1$ equivalence regions, where d is the number of observable features of the decision subject and m_i ($\forall i \in [d]$) is the number of actions the decision subject has at their disposal to improve observable feature i.*

Proof. In order to characterize the number of equivalence regions $|\mathcal{R}|$, we define the notion of a *dominated* action a, where an action a is dominated by some other action a' if $\Delta\mathbf{x}(a) \preceq \Delta\mathbf{x}(a')$, with strict inequality holding for at least one index. Using this notion of dominated actions and our refined characterization of an equivalence region, it is straightforward to see that if action a is dominated by action a', then $a' \in A_R$ for any equivalence region R where $a \in A_R$. Proposition C.1 then follows directly from the fact that each action only affects one observable feature. □

Proposition C.1 shows that the computation of (OPT-LP) quickly becomes intractable as the number of observable features grows large, even in this relatively simple setting. This motivates the need for an approximation algorithm for (OPT-LP).

D EXPERIMENTS

In this section, we provide experimental results that validate our findings using a semi-synthetic setting where decision subjects are based on individuals in the Home Equity Line of Credit (HELOC) dataset [16]. We compare the decision maker utility for different models of information revelation: our optimal signaling, revealing full information, revealing no information. To do so, we first estimate agent costs using the Bradley-Terry model [5] and compute the decision maker's expected utility for each information revelation scheme we consider. We find that the expected decision maker utility when recommending actions according to the optimal signaling policy either matches or exceeds the expected utility from revealing full information or no information about the assessment rule across all problem instances. Moreover, the expected decision maker utility from signaling is *significantly* higher on average. Next, we explore how the decision maker's expected utility changes when action costs and changes in observable features are varied jointly. Our results are summarized in Figures 4, 5, and 6.

The HELOC dataset contains information about 9,282 customers who received a Home Equity Line of Credit. Each individual in the dataset has 23 observable features related to an applicant's financial history (e.g., percentage of previous payments that were delinquent) and a label which characterizes their loan repayment status (repaid/defaulted). In order to adapt the HELOC dataset to our strategic setting, we select four features from the original 23 and define five hypothetical actions $\mathcal{A} = \{a_\emptyset, a_1, a_2, a_3, a_4\}$ that decision subjects may take in order to improve their observable features. Actions $\{a_1, a_2, a_3, a_4\}$ result in changes to each of the decision subject's four observable features, whereas action $a_\emptyset$ does not. For simplicity, we view actions $\{a_1, a_2, a_3, a_4\}$ as equally desirable to the decision maker, and assume they are all more desirable than $a_\emptyset$. See Table 2 for details about the observable features and actions we consider. Using these four features, we train a logistic regression model that predicts whether an individual is likely to pay back a loan if given one, which will serve as the decision maker's realized assessment rule.

Common prior. We assume the common prior over the realized assessment rule $\boldsymbol{\theta}$ takes the form of a multivariate Gaussian $\mathcal{N}(\boldsymbol{\theta}, \sigma^2 I_{4\times4})$ before training. This captures the setting in which both the decision maker and decision subjects have a good estimate of what the true model will be, but are somewhat uncertain about their estimate. We note that our methods extend to more complicated priors beyond the isotropic Gaussian prior we consider in this setting.

Changes in observable features. In order to examine the effects that different $\Delta\mathbf{x}(a_i)(i \in \{1, 2, 3, 4\})$ have on the decision maker's expected utility, we consider settings in which each $\Delta\mathbf{x}(a_i)$ takes a value in $\{0, 0.25, 0.5, 0.75, 1\}$.

[5] Specifically, if taking action a results in a positive classification for some $\boldsymbol{\theta} \in \Theta$ and a negative classification for $\boldsymbol{\theta} \in \Theta$, the only way for $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ to be in the same equivalence region is if taking *any* action in $\mathcal{A}$ results in a positive classification for $\boldsymbol{\theta}$ and a negative classification for $\boldsymbol{\theta}'$. Besides this special case, if $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ result in different classifications for the same action, they are in different equivalence regions.

Pair	Feature (x_i)	Action (a_i)
(x_1, a_1)	# payments with high-utilization ratio	decrease this value
(x_2, a_2)	# satisfactory payments	increase this value
(x_3, a_3)	% payments that were not delinquent	increase this value
(x_4, a_4)	revolving balance to credit limit ratio	decrease this value

Table 2: Decision subject's observable features from the HELOC dataset and corresponding actions to improve each feature. For simplicity, we assume that each action only affects one observable feature, although our model generally allows for more intricate relationships between actions and changes in observable features.

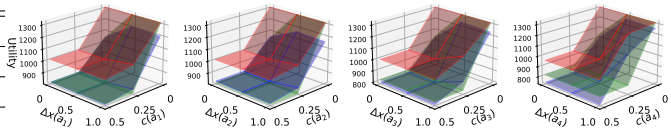


Figure 5: Utility surface across different $c(a)$ and $\Delta\mathbf{x}(a)$ pairs for $\sigma^2 = 0.4$. Optimal signaling policy (red) effectively upper-bounds the two baselines, revealing everything (blue) and revealing nothing (green) in all settings.

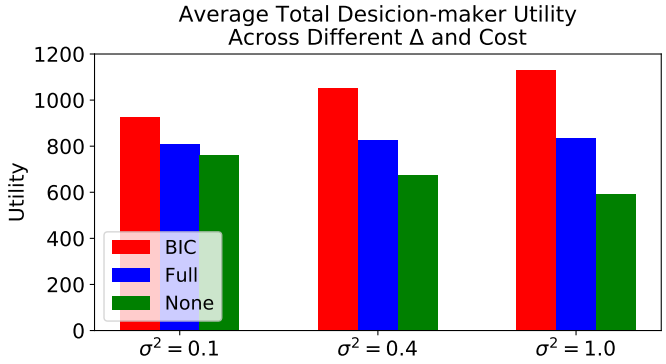


Figure 4: Total decision maker utility averaged across all cost and $\Delta\mathbf{x}(a)$ configurations for three different prior variances ($\sigma^2 = 0.1, 0.4, 1.0$). See Figure 8 to view individual plots of the settings which were averaged in order to generate this plot. The optimal signaling policy (red) consistently yields higher utility compared to the two baselines: revealing full information (blue) and no information (green). This gap increases when the decision subjects are less certain about the model parameters being used (higher σ^2).

Utilities and costs of actions. As the decision maker views actions $\{a_1, a_2, a_3, a_4\}$ as equally desirable, we define $u_{dm}(a_i) = 1$, $i \in \{1, 2, 3, 4\}$ and $u_{dm}(a_\emptyset) = 0$.[6] Since there are 1,320 individuals in our test dataset, the maximum utility the decision maker can obtain is 1,320. As proposed in [34], we use the Bradley-Terry model [5] to generate the decision subject's cost $c(a_i)$ of taking action a_i, for $i = 1, 2, 3, 4$. See Appendix G.2 for details on our exact generation methods.

Results. Given a $\{(c(a_i), \Delta\mathbf{x}(a_i))\}_{i=1}^4$ instance and information revelation scheme, we calculate the decision maker's total expected utility by summing their expected utility for each applicant. Figure 4 shows the average total expected decision maker utility across different $\Delta\mathbf{x}(a)$ and cost configurations for priors with varying amounts of uncertainty. See Figure 8 in Appendix G.3 for plots of all instances which were used to generate Figure 4. Across all instances, the optimal signaling policy (red) achieves higher average total utility compared to the other information revelation schemes (blue and green). The difference is further amplified whenever the decision subjects are less certain about the true assessment rule (i.e., when σ is large). Intuitively, this is because the decision maker leverages the decision subjects' uncertainty about the true assessment rule in order to incentivize them to take desirable actions, and as the uncertainty increases, so does their ability of persuasion.

[6] We set $u_{dm}(a_1) = u_{dm}(a_2) = u_{dm}(a_3) = u_{dm}(a_4)$ for ease of exposition — in general, actions can have different utility values based on their relative importance.

D.1 Patterns under different $\Delta\mathbf{x}(a)$ and $c(a)$

To better understand how the decision maker's expected utility changes as a function of $c(a)$ and $\Delta\mathbf{x}(a)$, we sweep through multiple $\{(c(a_i), \Delta\mathbf{x}(a_i))\}_{i=1}^4$ tuples on a grid of $(c(a_i), \Delta\mathbf{x}(a_i)) \in \{0, 0.25, 0.5\} \times \{0, 0.5, 1.0\}$ for $i \in \{1, 2, 3, 4\}$ and measure the effectiveness of the three information revelation schemes. Figure 5 shows the surface of the decision maker utility as a function of $(c(a_i), \Delta\mathbf{x}(a_i))$ for the optimal signaling policy (red), revealing full information (blue), and revealing no information (green). When $c(a_i)$ is high and $\Delta\mathbf{x}(a_i)$ is low, the total expected decision maker utility is low as there is less incentive for the decision subject to take actions (although even under this setting, the optimal signaling policy outperforms the other two baselines). As $c(a_i)$ decreases and $\Delta\mathbf{x}(a_i)$ increases, the total expected decision maker utility increases.

In Figure 6, we show 2-D slices of Figure 5 along the $c(a)$ axis (left) and $\Delta\mathbf{x}(a)$ axis (right). As is expected, with small cost and sufficiently large $\Delta\mathbf{x}(a)$ (top row, right), the two baselines become as effective as the optimal signaling policy. Interestingly, we note that changes in different $(c(a_i), \Delta\mathbf{x}(a_i))$ result in significantly different rates of change in decision maker utility. For example, the optimal signaling policy (red) and revealing full information (blue) are more resistant to the increase in $c(a_4)$ in range $[0, 0.25]$ than they are for the increase in other $c(a_i)$, $i \neq 4$, showing a concave drop in utility rather than a convex one (bottom row, left). Such behavior can be attributed to the relative weight of each feature on the learned assessment rule, where $|\theta_4| > |\theta_3| > |\theta_1| > |\theta_2|$. Because the fourth feature has the largest weight, taking action a_4 will have the largest effect on an individual's prediction. As a result, the decision maker utility is the least sensitive to increases in the cost of taking that action. Similarly, we observe that the degree to which changes in $\Delta\mathbf{x}(a)$ affect the expected utility is more drastic for a_4 compared to other actions (middle row, right).

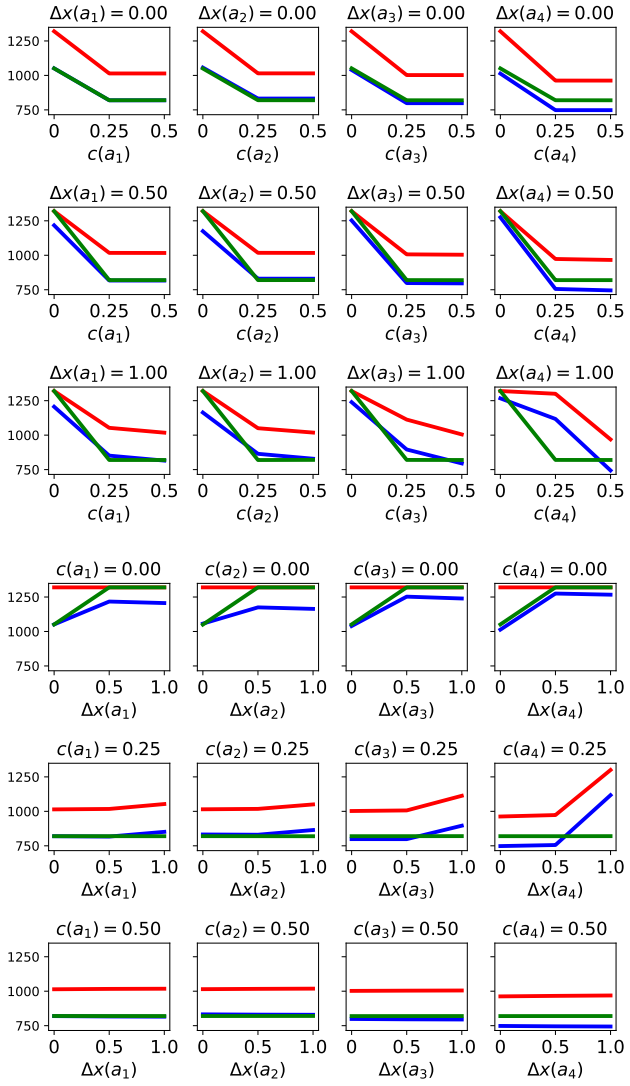


Figure 6: 2-D slices of Figure 5 across $c(a)$ (left) and $\Delta\mathbf{x}(a)$ (right). Across these two axes, the optimal signaling policy (red) dominates the revealing full information (blue) and revealing no information (green), though it may be possible for (blue) and (green) to vary in terms of which provides higher decision maker utility.

E COMPUTING THE OPTIMAL SIGNALING POLICY

$$\max_{p(\sigma=a|R),\forall a\in\mathcal{A},R\in\widetilde{\mathcal{R}}} \quad \sum_{a\in\mathcal{A}}\sum_{R\in\widetilde{\mathcal{R}}} \tilde{p}(R)p(\sigma=a|R)u_{dm}(a)$$

$$\text{s.t.} \quad \sum_{R\in\widetilde{\mathcal{R}}} p(\sigma=a|R)\tilde{p}(R)\left(u_{ds}(a,R)-u_{ds}(a',R)+\epsilon\right)\geq 0,$$

$$\forall a,a'\in\mathcal{A}$$

$$\sum_{a\in\mathcal{A}} p(\sigma=a|R)=1,\ \forall R\in\widetilde{\mathcal{R}},\quad p(\sigma=a|R)\geq 0,\ \forall R\in\widetilde{\mathcal{R}},$$

$$a\in\mathcal{A}.$$

(APPROX-LP)

Proof of Theorem 5.1

Proof. Our proof is similar to the approximation algorithm proof in Dughmi and Xu [14], and follows directly from the following lemmas, whose proofs are in Appendix E. First, since the approximation algorithm solves an approximation LP (APPROX-LP) of polynomial size, it runs in polynomial time.

Lemma E.1. *Algorithm 1 runs in poly$(m, \frac{1}{\epsilon})$ time.*

By bounding the approximation error in the BIC constraints of (APPROX-LP), we show that the resulting policy satisfies approximate BIC.

Lemma E.2. *Algorithm 1 implements an ϵ-BIC signaling policy.*

Next, we show that a feasible solution to (APPROX-LP) exists which achieves expected decision maker utility at least OPT - ϵ with probability at least $1-\delta$. In order to do so, we first show that there exists an approximately optimal solution $\widetilde{\mathcal{S}}$ to (OPT-LP) such that each signal is either (i) *large* (i.e., output with probability above a certain threshold), or (ii) *honest* (i.e., the signal recommends the action the decision subject would take, had they known the true assessment rule $\boldsymbol{\theta}$). Next, we show that $\widetilde{\mathcal{S}}$ is a feasible solution to (APPROX-LP) with high probability by applying McDiarmid's inequality [32] and a union bound.

Lemma E.3. *There exists an $\frac{\epsilon}{2}$-optimal signaling policy $\widetilde{\mathcal{S}}$ that is large or honest.*

Lemma E.4. *With probability at least $1-\delta$, $\widetilde{\mathcal{S}}$ is a feasible solution to (APPROX-LP) and the expected decision maker utility from playing $\widetilde{\mathcal{S}}$ is at least OPT - ϵ.*

By Lemmas E.3 and E.4, the decision maker's expected utility will be at least OPT - ϵ with probability at least $1-\delta$. □

Proof of Lemma E.1.

Proof. Lines 1-3 trivially run in poly$(m, \frac{1}{\epsilon})$ time. $\tilde{p}(R), \forall R \in \widetilde{\mathcal{R}}$ can be computed in poly$(m, \frac{1}{\epsilon})$ time in an online manner as follows: for each $k \in \{1,\ldots,K\}$, check if $\boldsymbol{\theta}_k$ belongs to an existing region. (Note that this can be done in $O(m)$ time for each region.) If $\boldsymbol{\theta}_k$ belongs to an existing region, update the existing empirical probabilities. Otherwise, create a new region. Finally, note that LP (APPROX-LP) has poly$(m, \frac{1}{\epsilon})$ variables and constraints, and can therefore be solved in poly$(m, \frac{1}{\epsilon})$ time using, e.g., the Ellipsoid Algorithm. □

Proof of Lemma E.2.

Proof. By the principle of deferred decisions, $\boldsymbol{\theta} \sim \Pi'$, where Π' is the uniform distribution over $\widetilde{\boldsymbol{\Theta}}$. (APPROX-LP) implements an ϵ-BIC signaling policy for Π' by definition, so

$$\mathbb{E}_{\boldsymbol{\theta}\sim\Pi'}[u_{ds}(a,R) - u_{ds}(a',R)|\sigma = a] \geq -\epsilon, \forall a, a' \in \mathcal{A}.$$

Finally, apply the law of iterated expectation with respect to $\widetilde{\boldsymbol{\Theta}}$ to obtain the desired result. □

Proof of Lemma E.3. In order to prove Lemma E.3, we make use of the following definitions.

Definition E.5 (Large signal). *A signal $\sigma = a$ is large if $p(\sigma = a) = \sum_{R\in\mathcal{R}} p(\sigma = a|R)p(R) > \frac{\epsilon}{2m}$.*

Definition E.6 (Honest signal). *A signal $\sigma = a$ is honest if $a \in \arg\max_{a'\in\mathcal{A}} u_{ds}(a', R)$.*

Proof. We proceed via proof by construction. Let $\mathcal{S}^*$ be the optimal BIC signaling policy. Define the signaling policy $\widetilde{\mathcal{S}}$ as follows: for a given $\boldsymbol{\theta}$, it first samples a signal $a \sim \mathcal{S}^*(\boldsymbol{\theta})$. If the signal is *large*, output signal $\sigma = a$. Otherwise, output signal $\sigma = a \in \arg\max_{a'\in\mathcal{A}} u_{ds}(a', R_{\boldsymbol{\theta}})$. Every signal of $\widetilde{\mathcal{S}}$ is trivially large or honest. $\widetilde{\mathcal{S}}$ is BIC since $\mathcal{S}^*$ is BIC and $\widetilde{\mathcal{S}}$ only replaces recommendations of $\mathcal{S}^*$ with honest recommendations. Finally, since the total probability of signals that are not large is at most $\frac{\epsilon}{2}$, and the decision maker's utilities are in $[0, 1]$, their expected utility is no worse than $\frac{\epsilon}{2}$ smaller than their expected utility from $\mathcal{S}^*$. □

Proof of Lemma E.4.
The following claim will be useful when proving Lemma E.4.

Claim E.7. *The expected decision maker utility from playing $\widehat{\mathcal{S}}$ is $\sum_{a\in\mathcal{A}}\sum_{R\in\mathcal{R}} p(R)p(\sigma = a|R)u_{dm}(a)$.*

Proof. The expected decision maker utility from playing $\widehat{\mathcal{S}}$ is

$$\mathbb{E}_{\boldsymbol{\theta}\sim\Pi}[\sum_{a\in\mathcal{A}}\sum_{R\in\widetilde{\mathcal{R}}} \tilde{p}(R)p(\sigma = a|R)u_{dm}(a)]$$
$$= \mathbb{E}_{\boldsymbol{\theta}\sim\Pi}[\sum_{a\in\mathcal{A}}\sum_{R\in\mathcal{R}} \tilde{p}(R)p(\sigma = a|R)u_{dm}(a)],$$

by the principle of deferred decisions. Apply the law of iterated expectation with respect to $\widetilde{\boldsymbol{\Theta}}$ to obtain the desired result. □

Additionally, we will make use of McDiarmid's inequality [32], stated for completeness below.

Lemma E.8 (McDiarmid's Inequality [32]). *Let $X_1, \ldots, X_n$ be independent random variables, with X_k taking values in a set A_k for each k. Suppose that the (measurable) function $f : \Pi A_k \to \mathbb{R}$ satisfies*

$$|f(\mathbf{x}) - f(\mathbf{x}')| \leq c_k$$

whenever the vectors $\mathbf{x}$ and $\mathbf{x}'$ differ only in the kth coordinate. Let Y be the random variable $f(X_1, \ldots, X_n)$. Then for any $t > 0$.

$$\mathbb{P}(|Y - \mathbb{E}[Y]| \geq t) \leq 2\exp\left(-2t^2/\sum_k c_k^2\right).$$

We are now ready to prove Lemma E.4.

Proof. First, note that the ϵ-BIC constraints can be rewritten using the observed decision rules as

$$\frac{1}{K}\sum_{k=1}^{K} p(\sigma = a|R_k)(u_{ds}(a, R_k) - u_{ds}(a', R_k)) \geq -\epsilon, \quad \forall a, a' \in \mathcal{A},$$

where $\boldsymbol{\theta}_k \in R_k$. Note that this is a bounded function of $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K$. Let $Y(a, a') = \frac{1}{K}\sum_{k=1}^{K} p(\sigma = a|R_k)(u_{ds}(a, R_k) - u_{ds}(a', R_k))$. Note that $\mathbb{E}[Y(a, a')] = \sum_{R\in\mathcal{R}} p(\sigma = a|R)p(R)(u_{ds}(a, R) - u_{ds}(a', R))$.

Applying Lemma E.8, we see that $\forall a, a' \in \mathcal{A}$,

$$\mathbb{P}(|Y(a, a') - \mathbb{E}[Y(a, a')]| \geq \epsilon) \leq 2\exp\left(-K\epsilon^2/2\right).$$

Similarly, let

$$Z = \sum_{a\in\mathcal{A}}\sum_{R\in\widetilde{\mathcal{R}}} \tilde{p}(R)p(\sigma = a|R)u_{dm}(a)$$
$$= \frac{1}{K}\sum_{k=1}^{K}\sum_{a\in\mathcal{A}} p(\sigma = a|R_k)u_{dm}(a)$$

(where R_k contains $\boldsymbol{\theta}'_k$). By Claim E.7, $\mathbb{E}[Z] = \sum_{a\in\mathcal{A}}\sum_{R\in\mathcal{R}} p(R)p(\sigma = a|R)u_{dm}(a)$. Applying Lemma E.8,

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq \epsilon/2) \leq 2\exp\left(-K\epsilon^2/2\right).$$

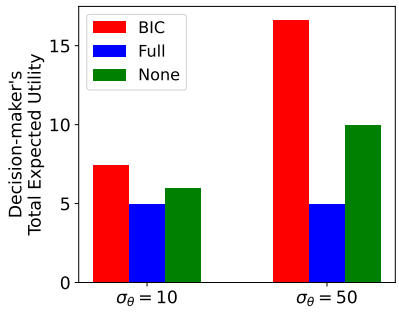
Applying the union bound, we see that the probability that *all* m^2+1 above inequalities hold is at least $2(m^2+1)\exp\left(-K\epsilon^2/2\right)$. By inverting the tail bound and picking $K = \frac{2}{\epsilon^2}\log\left(\frac{2(m^2+1)}{\delta}\right)$, we get that $|Z - \mathbb{E}[Z]| \leq \epsilon/2$ *and* $|Y(a, a') - \mathbb{E}[Y(a, a')]| \leq \epsilon, \forall a, a' \in \mathcal{A}$, with probability at least $1 - \delta$. Therefore, with probability at least $1 - \delta$, $\widetilde{\mathcal{S}}$ is a feasible solution for LP (APPROX-LP) and the objective value is at most OPT $-\frac{\epsilon}{2} - \frac{\epsilon}{2} =$ OPT $-\epsilon$. □

F INSTANTIATING 1-DIMENSIONAL SCENARIO

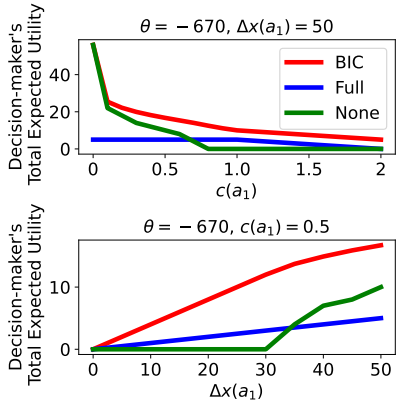
In this section we instantiate the example introduced in Section 3 and demonstrate the decision maker's gain in utility from the optimal signaling policy over other baselines. To contextualize this simple synthetic setup, consider a banking institution deciding whether approve a loan application from an applicant based on credit score $\mathbf{x}_0 \in [300, 850]$ with a simple threshold classifier. The bank approves the application ($\hat{y} = 1$) if $\mathbf{x}_0 + \theta > 0$ and rejects ($\hat{y} = -1$) otherwise. Here, we assume the ground-truth threshold value used by the decision maker to be 670 (i.e. $\theta = -670$), which is typically considered as a decent credit score. Recall that $a_\emptyset$ = "do nothing" and a_1 = "pay off existing debt" and set the utility of the decision maker to be $u_{dm}(a_1) = 1, u_{dm}(a_\emptyset) = 0$, as, for the sake of our illustration, we assume credit score to be a good measure of credit-worthiness. Finally, we assume the prior to be $\pi(\theta) \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$.

In Figure 7, we verify that our optimal signaling policy (**BIC**, red) yields higher decision maker utility compared to the two baselines: revealing full information (**Full**, blue) and revealing no information (**None**, green)[7]. To measure the total amount of decision maker's

[7]We set the decision subject cost of taking action a_1 to $c(a_1) = 0.5$, and $\Delta\mathbf{x}(a_1) = 40$ (i.e., action a_1 improves an applicant's credit score by 40 points).



(a) Expected decision maker utility summed up across different $\mathbf{x}_0$ values under different standard deviation σ_θ of the prior.



(b) Total expected decision maker utility under different $c(a_1)$ (top) and Δx (bottom)

Figure 7: (a) Total expected decision maker utility summed across difference $\mathbf{x}_0$ for our optimal signaling policy (BIC, red), against the two baselines: revealing full information about the assessment rule (Full, blue), and revealing no information (None, green). As the decision subject's uncertainty about the true threshold θ (measured by σ_θ) increases, the advantage of the optimal signaling policy becomes more visible. (c) When taking action a_1 becomes cost-prohibitive (high $c(a_1)$) or less effective (small $\Delta\mathbf{x}(a_1)$), the decision maker's utility decreases as there is less incentive for the decision subject to take the action. Nevertheless, the optimal signaling policy yields consistently higher decision maker utility compared to the baselines.

expected utility yielded by each policy, we assume a uniform distribution of the decision subjects' credit scores $\mathbf{x}_0$ in the population and take the sum of expected decision maker utility values across different scores. We plot these total utility values in Figure 7a, and as expected, the larger the σ_θ is, the more comparative advantage our method has over the baselines. As the decision subjects' uncertainty about the true θ increases (i.e., the standard deviation of the prior distribution increases from 10 to 50), the decision maker benefits from our optimal signaling policy even more.

When action a_1 becomes more cost-prohibitive (or less effective), as there is less incentive for the decision subjects to take the action, we expect the decision maker's utility to decrease[8]. As shown in Figure 7b, we indeed observe such a trend as $c(a_1)$ increases (top) and $\Delta\mathbf{x}(a_1)$ decreases (bottom). Nevertheless, our optimal signaling policy yields consistently higher total decision maker utility compared to the baselines across all conditions.

G EXPERIMENT DETAILS AND ADDITIONAL RESULTS

G.1 Remark on the decision maker's assessment rule for HELOC dataset

To simulate a setting in which the decision maker employs a machine learning model for making decisions about the decision subjects, we train a simple logistic regression model on the subset of HELOC dataset. We specifically work on four features selected in Table 2, and split the dataset into train/test set (7425, 1857 data points respectively). The test accuracy of the model was 71.08 percent, and the corresponding model coefficients were $\boldsymbol{\theta} = [-0.22974527, 0.15633134, 0.52023116, -0.61600619]$ with the bias term -0.08242841. Note that each coefficient term has the sign that is aligned with how the desired action was defined in Table 2 (i.e., for features where increasing the value is desirable, the sign is positive and vice-versa). To further make sure that the defined actions correctly align with the model, we select the test samples that the trained model made no mistakes on. This resulted in a total of 1,320 samples from the test set on which each policy was optimized.

G.2 Computing different costs for HELOC dataset using Bradley-Terry model

While exact action costs may be unknown, it is often reasonable for the decision maker to know an *ordering* over possible actions in terms of their cost for decision subjects. For example, it may be common knowledge that opening a new credit card is easier than paying off some existing amount of debt, but exactly *how much* easier may be unclear. The Bradley-Terry model uses exponential score functions to model the probability that feature x_i is more costly for a decision subject to take compared to feature x_j. Specifically, it assumes

$$\mathbb{P}(a_i \succ a_j) = \frac{e^{c(a_i)}}{e^{c(a_i)} + e^{c(a_j)}}.$$

Given pairwise cost comparisons (generated from common knowledge or gathered from experts) we can estimate $\mathbb{P}(a_i \succ a_j)$ and solve for the parameters $\{c(a_i)\}_{i=1}^4$ using maximum likelihood estimation. In order to gain more insight into how different action cost orderings affect the decision maker utility, we consider several different ground-truth cost orderings over actions and simulate expert advice in order to estimate $\mathbb{P}(a_i \succ a_j), \forall a_i, a_j \in \mathcal{A}$. While the expert advice is purely synthetic in our setting, this method provides a principled way to estimate action costs whenever input

[8] In this setting, we set $\mu_\theta = -650$ and $\sigma_\theta = 50$ so that the decision subjects are considered to have a reasonable estimate of the true threshold $\theta = -670$, to make the situation more favorable to the baselines.

Feature A	Feature B	# (A > B)	# (A < B)
$\mathbf{x}_1$	$\mathbf{x}_2$	8	2
$\mathbf{x}_1$	$\mathbf{x}_3$	9	1
$\mathbf{x}_1$	$\mathbf{x}_4$	7	3
$\mathbf{x}_2$	$\mathbf{x}_3$	2	8
$\mathbf{x}_2$	$\mathbf{x}_4$	0	10
$\mathbf{x}_3$	$\mathbf{x}_4$	1	9

(a) $c(a_1) > c(a_4) > c(a_3) > c(a_2)$

Feature A	Feature B	# (A > B)	# (A < B)
$\mathbf{x}_1$	$\mathbf{x}_2$	2	8
$\mathbf{x}_1$	$\mathbf{x}_3$	3	7
$\mathbf{x}_1$	$\mathbf{x}_4$	4	6
$\mathbf{x}_2$	$\mathbf{x}_3$	6	4
$\mathbf{x}_2$	$\mathbf{x}_4$	7	3
$\mathbf{x}_3$	$\mathbf{x}_4$	6	4

(b) $c(a_2) > c(a_3) > c(a_4) > c(a_1)$

Feature A	Feature B	# (A > B)	# (A < B)
$\mathbf{x}_1$	$\mathbf{x}_2$	2	8
$\mathbf{x}_1$	$\mathbf{x}_3$	1	9
$\mathbf{x}_1$	$\mathbf{x}_4$	4	6
$\mathbf{x}_2$	$\mathbf{x}_3$	3	7
$\mathbf{x}_2$	$\mathbf{x}_4$	7	3
$\mathbf{x}_3$	$\mathbf{x}_4$	7	3

(c) $c(a_3) > c(a_2) > c(a_4) > c(a_1)$

Feature A	Feature B	# (A > B)	# (A < B)
$\mathbf{x}_1$	$\mathbf{x}_2$	8	2
$\mathbf{x}_1$	$\mathbf{x}_3$	9	1
$\mathbf{x}_1$	$\mathbf{x}_4$	2	3
$\mathbf{x}_2$	$\mathbf{x}_3$	7	8
$\mathbf{x}_2$	$\mathbf{x}_4$	0	10
$\mathbf{x}_3$	$\mathbf{x}_4$	1	9

(d) $c(a_4) > c(a_1) > c(a_3) > c(a_2)$

Configuration	$c(a_1)$	$c(a_2)$	$c(a_3)$	$c(a_4)$
(i)	0.5151	0.0282	0.0723	0.3844
(ii)	0.1159	0.428	0.2758	0.1803
(iii)	0.07640764	0.27692769	0.50635064	0.14031403
(iv)	0.2987	0.0428	0.0476	0.6109

(e) Cost values learned by the Bradley-Terry model from the pair-wise comparison inputs above.

Table 3: Comparison inputs used by the Bradley-Terry model to generate different cost configurations.

from domain experts (e.g., financial advisors) is available to the decision maker.

We use the following set of comparison inputs (manually generated) in Table 3a-3d to generate cost values with the relative ordering presented in Section D. While these comparison inputs are generated arbitrarily for the simulations, these can be obtained by querying several domain experts and aggregating their answers regarding which feature is more difficult to change. The resulting cost values are shown in Table 3e.

G.3 Additional results for different cost and $\Delta\mathbf{x}(a)$ configurations

Figure 8 shows more exhaustive results on different cost configurations (i)-(iv) as defined in Table 3e and $\Delta\mathbf{x}(a_i) \in 0, 0.25, 0.5, 0.75, 1.0$ for $i = 1, 2, 3, 4$ on HELOC datset. For all configurations considered, our optimal signaling policy (red) consistently yields utility no less than both baselines: revealing full information about the assessment rule (blue), and revealing no information (green).

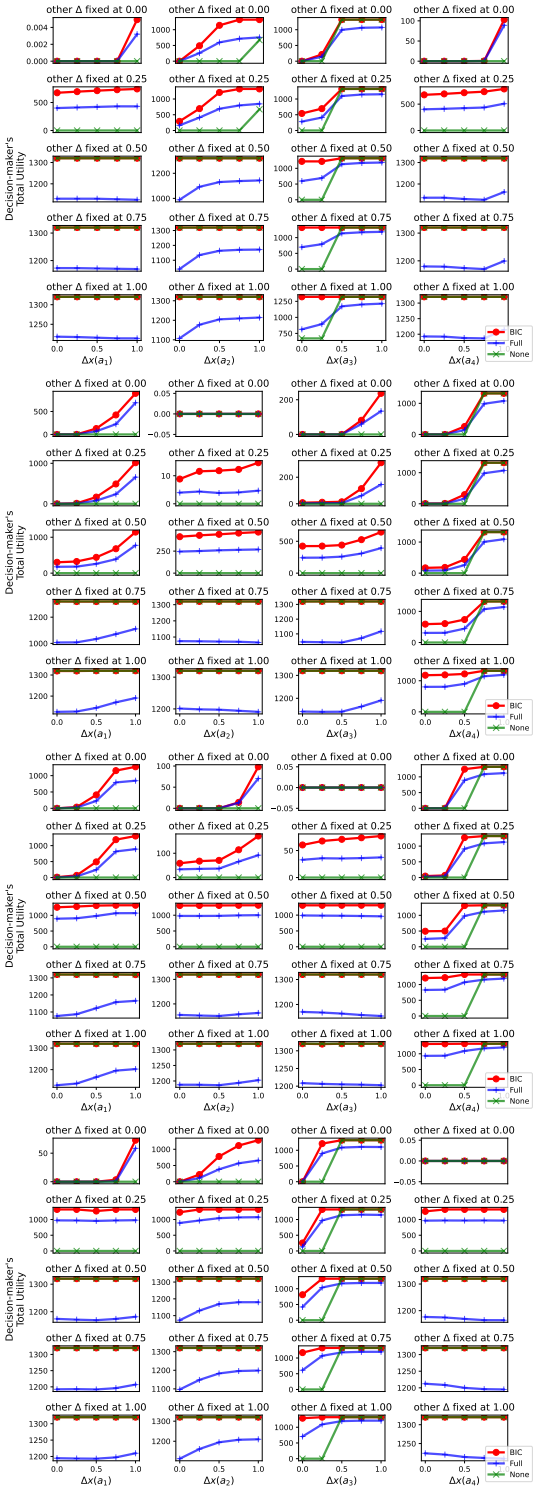


Figure 8: More detailed view on decision maker utility for different $\Delta \mathbf{x}(a)$ values and cost configurations (i)-(iv) (from upper right to bottom right quadrants). Our optimal signaling policy (red) consistently achieves utility no less than revealing full information (blue) and revealing no information (green) in all settings.