Persuasive Calibration*

Yiding Feng † Wei Tang ‡

Abstract. We introduce and study the *persuasive calibration* problem, where a principal aims to provide trustworthy predictions about underlying events to a downstream agent to make desired decisions. We adopt the standard calibration framework that regulates predictions to be unbiased conditional on their own value, and thus, they can reliably be interpreted at face value by the agent. Allowing a small calibration error budget, we aim to answer the following question: what the optimal predictor is and how to compute it under this calibration error budget, especially when there exists incentive misalignment between the principal and the agent? We focus on the standard ℓ_t-norm Expected Calibration Error (ECE) metric.

We develop a general framework by viewing predictors as post-processed versions of perfectly calibrated predictors. Using this framework, we first characterize the structure of the optimal predictor. Specifically, when the principal's utility is outcome-independent and for ℓ_1-norm ECE, we show: (1) the optimal predictor is over-(resp. under-) confident for high (resp. low) true expected outcomes, while remaining perfectly calibrated in the middle; (2) the miscalibrated predictions exhibit a collinearity structure with the principal's utility function. On the algorithmic side, we provide an FPTAS for computing approximately optimal predictor for general principal utility and general ℓ_t-norm ECE. Moreover, for the ℓ_1- and ℓ_∞-norm ECE, we provide polynomial-time algorithms that compute the exact optimal predictor.

1 Introduction Over the past decade, machine learning models have grown remarkably powerful, with recent large-scale models (like LLMs) often containing billions of parameters and providing good predictions in a wide variety of domains. Yet these models/algorithms are frequently regarded as black-box systems: their internal mechanisms remain proprietary, or are considered trade secrets and thus, are not observed by outsiders. Consequently, end users and downstream decision-makers are often left to trust predictions without transparent insight into how those predictions are generated.

A prominent approach to bolstering trust in such predictions is *calibration* (Dawid, 1982; Foster and Vohra, 1998; Hébert-Johnson et al., 2018). A calibrated predictor, in simple terms, regulates that predicted probabilities align with the true (conditional) probability of the outcome. For example, predictions of "70% likelihood" materialize approximately 70% outcome realizations of the time. Calibration thus provides a guarantee that predictions are trustworthy and can be reliably taken at face value for use in downstream decision-making. Indeed, it is well established that agents who naively best respond to calibrated predictions (i.e., those with low calibration error – a measure of how far the predictions deviate from the true conditional probability) achieve low regret (Foster and Vohra, 1998; Kleinberg et al., 2023; Haghtalab et al., 2023; Roth and Shi, 2024; Hu and Wu, 2024), whose rate is even no larger than the inherent calibration error of the predictor. At the same time, however, the principal – the entity designing or deploying the prediction model – may have incentives that differ from those of the downstream agent who relies on these predictions. For example, the principal might wish to skew predictions ever so slightly to influence the agent's actions in a beneficial (to the principal) way.

The above tension naturally leads to the following question: If a small calibration error is permissible, how might the principal design an optimal predictor that balances the principal's objectives while preserving the agent's trust by not miscalibrating so much? To study this question, in this paper, we formulate and analyze the *persuasive calibration* problem. Specifically, we introduce a framework in which the principal aims to provide predictions that must remain acceptably calibrated while ensuring that the downstream agent, who only sees the final predictions, would maintain sufficient trust to best respond to them. We aim to understand the following two questions:

*The full version of the paper (Feng and Tang, 2025a) can be accessed at https://arxiv.org/abs/2504.03211

†Hong Kong University of Science and Technology (ydfeng@ust.hk, https://www.ydfeng.us/).
‡The Chinese University of Hong Kong (weitang@cuhk.edu.hk, https://wtang.org/).

Can we characterize the optimal predictor under a calibration error budget, especially when there exists incentive misalignment between the principal and the agent? In addition, can we compute this optimal predictor or an approximately optimal predictor in polynomial time?

1.1 Our Contributions and Techniques In this paper, we provide compelling answers to both of these questions. In a nutshell, we introduce the *persuasive calibration* problem, which captures the interaction between a principal, who designs an approximately calibrated predictor, and an agent, who takes actions based on the predictions. The calibration constraint serves as an operational guarantee that ensures the trustworthiness of the predictor from the agent's perspective. In the first part of the paper, we focus on a canonical setting where the principal's utility is outcome-independent and the predictor is evaluated based on the most classic ℓ_1-norm calibration error. For this setting, we provide a comprehensive characterization of the optimal predictor. In the second part of the paper, we move to the more general setting where the principal's utility can be arbitrary and the predictor can be evaluated by any ℓ_t-norm calibration error. For this rich setting, we develop an FPTAS (fully polynomial-time approximation scheme) algorithm for all instances of our problem, and an optimal polynomial-time algorithm for ℓ_1-norm or ℓ_∞-norm calibration errors. Below, we discuss our model, results, and techniques in detail.

The Principal-Agent Model (Section 2) The *persuasive calibration* problem models the interaction between a predictor designer (the principal, she) and a downstream decision-maker (the agent, he). In this setting, there is an underlying randomized binary outcome, which is unobserved by both players. However, the principal observes a realized *event* (e.g., a feature or context) that is correlated with the outcome and uses this information to generate a prediction. This prediction is a scalar value representing the probability that the binary outcome equals one.

The agent, whose utility depends on both his action and the realized outcome, *trusts* the prediction provided by the principal and chooses an action that best responds to the provided prediction. In contrast, the principal's utility depends not only on the agent's action but also on the realized event and outcome, and may be misaligned with the agent's utility. The principal's goal is to design a predictor that maximizes her own utility, given the agent's best response.

We adopt the calibration framework to ensure the trustworthiness of the predictors, that is, conditional on a given prediction, the expected true outcome should be close to the predicted probability. In this work, we focus on the classic ℓ_t-norm expected calibration error (ECE), which quantifies the expected ℓ_t-norm difference between the prediction and the true expected outcome. More formally, the *persuasive calibration* problem seeks the optimal predictor that maximizes the principal's payoff, subject to an exogenously specified ECE budget.

Result I: Structural Characterization of the Optimal Predictor (Section 3) We begin our exploration of the persuasive calibration problem under the *ℓ_1-norm ECE* (also known as the K_1 ECE), which is the most standard calibration error metric. We focus on a canonical setting where the principal's *indirect utility*—mapping predictions to her utilities under the assumption that the agent best responds—is outcome-independent.[1]

The structure results of the optimal predictor. As the first main result of this paper, we provide a comprehensive characterization of the optimal predictor under any exogenously specified K_1 ECE budget. In Subsection 3.1, we establish two key properties: (1) the *miscalibration structure* where the optimal predictor is over- (resp. under-) confident for high (resp. low) true expected outcomes while remaining perfectly calibrated in the middle; and (2) the *payoff structure*, where the miscalibrated predictions in the optimal predictor exhibit a collinearity pattern with the principal's indirect utility function.

We illustrate both properties in Figure 1.1a. In this figure, the black solid curve represents the principal's indirect utility, while each blue dot corresponds to a prediction generated by the optimal predictor. Specifically, the x-coordinate of each blue dot indicates the prediction value, and the y-coordinate represents the corresponding payoff for the principal. The brown squares denote the true expected outcomes conditional on the predictions.

Our *miscalibration structure* (in Theorem 3.2) suggests that the prediction space $[0, 1]$ can be partitioned into three sub-intervals: (1) the *under-confidence interval* $[0, p_{\mathsf{L}}]$, where predictions (blue dots) are lower (smaller x-coordinate) than their corresponding true expected outcomes (brown squares); (2) the *over-confidence interval* $[p_{\mathsf{H}}, 1]$, where predictions (blue dots) are higher (larger x-coordinate) than their corresponding true

[1] This assumption is common in various principal-agent models (e.g., Dworczak and Martini, 2019; Lipnowski and Ravid, 2020; Feng et al., 2022; Arieli et al., 2023; Corrao and Dai, 2023; Feng et al., 2024).

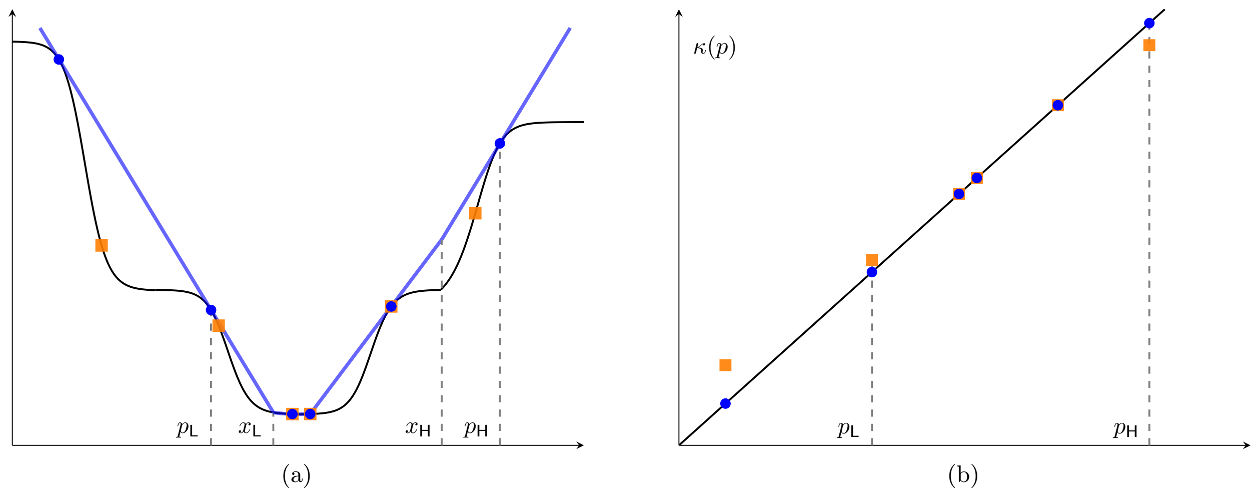

Figure 1.1: In both figures, blue dots are the predictions p generated from the optimal predictor $\tilde{f}^*$, brown squares are the corresponding true expected outcomes $\kappa(p)$ conditional on the prediction p, Figure 1.1a is an illustration of Theorem 3.2 – black solid curve is indirect utility U and blue curve is the convex function Γ in Proposition 3.4. Figure 1.1b is the reliability diagram (a standard visualization of the miscalibration structure) – black solid line is diagonal line.

expected outcomes (brown squares); and (3) the *perfectly calibrated interval* $[p_\mathsf{L}, p_\mathsf{H}]$, where predictions (blue dots) exactly match their corresponding true expected outcomes (brown squares). We remark that this three-interval miscalibration pattern appears to be a well-observed phenomenon in the machine learning community. For example, Guo et al. (2017) show that uncalibrated ResNet tends to be overconfident in its predictions, while Kuleshov et al. (2018) show that neural networks consistently generate both underestimated and overestimated predictions/forecasts at the two extremes of the probability spectrum. In fact, the terms "under-confident" and "over-confident" are originally introduced in the calibrated machine learning literature (e.g., Guo et al., 2017; Wang et al., 2021; Ao et al., 2023). By interpreting the loss function in these machine learning tasks as the principal's utility, our results provide a theoretical explanation for this practical observation.

Our *payoff structure* (in Theorem 3.2 and Proposition 3.4) suggests that all blue dots (representing predictions and their induced payoffs for the principal) lie on a *symmetric linear-tailed convex function* Γ (blue curve), which is pointwise (weakly) greater than the principal's indirect utility (black curve). Moreover, all blue dots in the under-confidence interval $[0, p_\mathsf{L}]$ (resp. over-confidence interval $[p_\mathsf{H}, 1]$) are collinear, with the slopes of these two linear tails having the same absolute value.

Two-step framework to generate optimal predictors. To prove the aforementioned structural results, we develop a two-step framework for the principal to design and analyze the optimal predictor. At a high level, the framework treats the predictor as a post-processed version of a perfectly calibrated predictor (i.e., zero calibration error). Specifically, we introduce the *post-processing plan*, which "miscalibrates" a perfectly calibrated predictor into an imperfectly calibrated one (Definitions 3.5 and 3.6 and Proposition 3.8). Using this concept, identifying the optimal predictor subject to an ECE budget constraint can be reformulated as optimizing a pair consisting of a perfectly calibrated predictor and an "ECE-budget-feasible" miscalibration plan. This reformulation allows us to leverage well-established properties of perfectly calibrated predictors in the analysis.

A particularly interesting subclass of post-processing plans is the event-independent ones. However, not every imperfectly calibrated predictor can be obtained by event-independently post-processing a perfectly calibrated predictor (see Figure 3.1 for an illustration and Example 3.9 for an example). Perhaps interestingly, we find that when the principal's indirect utility is outcome-independent, it suffices to consider event-independent post-processing plans to obtain the optimal predictor (Proposition 3.12). Furthermore, optimizing over perfectly calibrated predictors and event-independent miscalibration plans can be characterized as a linear program (see LP-TwoStep).

Primal-dual analysis of LP-TwoStep. Utilizing the two-step framework, we know that the optimal predictor can be transformed into an optimal solution of LP-TwoStep and vice versa. Therefore, proving the structural results for the optimal predictor is equivalent to proving their analogs for the optimal solution in LP-TwoStep. By developing both primal-based and duality-based arguments, we establish the aforementioned structural results. Notably, the miscalibration structure is primarily proved through a primal-based argument, which leverages the structure (i.e., mean-preserving contraction (MPC)) of perfectly calibrated predictors, while the payoff structure is primarily proved using duality-based arguments. In particular, the symmetric linear-tailed convex function Γ (see Figure 1.1a and Definition 3.3) arises as a consequence of the optimal dual solutions. For example, the slope of its linear tail corresponds to the optimal dual variable associated with the ECE budget constraint in LP-TwoStep.

Result II: FPTAS for Computing Approximately Optimal Predictor (Section 4) Moving forward, we explore the algorithmic aspects of the persuasive calibration problem. We consider a more general setting where the principal's utility can be arbitrary and the calibration metrics can be general ℓ_t-norm ECE with any $t \geq 1$. As the second main result of this paper, we provide an FPTAS (Algorithm 4.1 and Theorem 4.1) for computing an approximately optimal predictor subject to any exogenously specified ECE budget.

To obtain a time-efficient algorithm for computing an approximately optimal predictor, a natural approach is to formulate the principal's problem as a computationally tractable optimization program. However, due to the calibration error constraint defined on the predictor $\tilde{f}$, this approach is not straightforward to ensure an efficient algorithm. In the full version of the paper (Feng and Tang, 2025a), we explain the failure of two natural optimization program formulations that directly use the predictor as the decision variable and optimize over the entire space of feasible predictors. Our FPTAS contains two technical ingredients: it combines a generalized two-step framework, which extends the two-step approach (used in our structural characterization) from the outcome-independent setting to the general setting, with a carefully designed discretization scheme.

Generalized two-step framework. For the general setting where the principal's utility can depend on the realized event, our two-step framework with the *event-independent* post-processing plan established for the structural results may not be sufficient. Motivated by this, we introduce a generalized two-step framework in Subsection 4.2. However, if we consider an event-dependent post-processing plan that *fully decouples* the miscalibration across all events, additional non-linear constraints—seemingly intractable—are required. To bypass this issue, we introduce a generalized two-step framework with an *(event-dependent) bi-event post-processing plan* (Definitions 4.2 and 4.3), which allows us to formulate an infinite-dimensional (but tractable) linear program LP-TwoStep$^+$. Our bi-event post-processing plan is rich enough to ensure that there always exists an optimal predictor that can be generated within our generalized framework (Proposition 4.6).[2]

Instance-dependent two-layer discretization & rounding-based analysis. Equipped with an infinite-dimensional linear program LP-TwoStep$^+$, a natural approach is to consider its discretization. However, common discretization schemes, such as uniform discretization, are not directly applicable. Specifically, due to the ECE budget constraint, standard discretization may either render the discretized LP infeasible or significantly degrade the objective value.

To overcome this challenge, in Subsection 4.3, we introduce a carefully designed discretization scheme (Definition 4.8) that is instance-dependent and has a two-layer structure. First, given a discretization parameter $\delta > 0$, our scheme constructs a non-uniform $\Theta(\delta)$-net, where the discretized points depend on the problem primitives, ensuring feasibility of the discretized program. Then, for each point in this $\Theta(\delta)$-net, we introduce a finer $\Theta(\delta_0)$-net within a small neighborhood. Here, δ_0 depends not only on the discretization precision δ but also on the ECE budget ε and the norm exponent parameter t. This second layer, which adapts to the ECE budget, allows us to carefully control changes in ECE when transitioning from a solution in the infinite-dimensional LP to one in the discretized LP. Crucially, the second-layer discretized points are introduced only locally, ensuring that the overall discretization set remains of polynomial size.

Finally, we develop a rounding scheme (Algorithm 4.2) to analyze both the feasibility and the objective value of the discretized program LP-DiscTwoStep$^+_\delta$. We believe that both our discretization scheme and rounding argument might be of independent interest in the algorithmic information design literature.

[2] The foundational reason why the bi-event post-processing plan suffices is that our current model assumes a binary outcome. In fact, we believe that our FPTAS admits a relatively straightforward generalization to an extended model with a constant number of outcomes, which we leave for future exploration.

Result III: Computing Optimal Predictor under K_1 or K_∞ ECE (Section 5) As the third main result of this paper, we focus on the most standard ECE metrics, namely the K_1 ECE (i.e., ℓ_1-norm ECE). We present a polynomial-time algorithm (Algorithm 5.1 and Theorem 5.1) that computes an optimal predictor while adhering to any exogenously specified K_1 ECE budget. Furthermore, we demonstrate that the same algorithm is also applicable to the K_∞ ECE (i.e., ℓ_∞-norm ECE).

For K_1 or K_∞ ECE, it is not difficult to express the calibration constraint as a linear constraint of the predictor. However, to obtain an optimal algorithm with polynomial running time, we also need to restrict the continuous space of predictions to a discrete polynomial-size set. To bypass this challenge, we build on an idea from the algorithmic information design literature.

Revelation principle and (Bayesian) persuasion with signal-dependent bias. In the classic Bayesian persuasion problem (Kamenica and Gentzkow, 2011), the *revelation principle* assures that it is sufficient to consider signaling schemes that recommend an action. Hence, instead of searching over a (possibly) continuous signal space, it suffices to construct a signaling scheme with a signal space equal to the agent's action space, which is polynomial-sized. However, such an approach seems difficult to apply to our model, as the design space in our model is restricted to predictions rather than arbitrary signals, and the agent in our model follows a much simpler behavior—i.e., always trusting the prediction—rather than engaging in strategic reasoning. Therefore, it is unclear a priori whether a revelation principle can be established.

Although at first glance, our persuasive calibration problem and the classic Bayesian persuasion problem may seem to be different, we demonstrate that the persuasive calibration problem can also be interpreted as a new variant of the Bayesian persuasion problem, which we refer to as *(Bayesian) persuasion with signal-dependent bias*. In this model, the receiver (equivalent to the agent) (1) has a utility function that depends linearly on the payoff-relevant state, and (2) exhibits a signal-dependent bias when updating their belief, rather than updating their belief in a fully Bayesian manner. The sender (equivalent to the principal) now determines both the signaling scheme and the bias assignment for the receiver, subject to an exogenously specified constraints (Eqn. (5.3)) on the total bias.[3] In the special case where the total bias is restricted to zero, this variant reduces to the classic Bayesian persuasion problem.

With this equivalent interpretation, it becomes relatively more straightforward for us to identify a suitable revelation principle (Lemma 5.4) and formulate the problem of optimizing the pair consisting of the signaling scheme and the bias assignment (equivalent to the predictor) as an action-recommendation program P-ActRec (Proposition 5.5). Notably, both the equivalence between two models and P-ActRec holds for a general ℓ_t-norm ECE budget constraint with any $t \geq 1$. For K_1 and K_∞ ECE, the corresponding P-ActRec becomes a linear program. Since the program has polynomial size and the transformation between the signaling scheme (with the bias assignment) and the predictor is also polynomial-time computable, we obtain Algorithm 5.1, which computes the optimal predictor under the K_1 or K_∞ ECE budget in polynomial time.

Given the significance of the Bayesian persuasion and information design literature, we believe that our new variant of persuasion problem with signal-dependent bias, together with its results, may be of independent interest.

1.2 Important Future Directions

The persuasive calibration problem studied in this work opens up several interesting directions for future research, some of which we outline below.

Structural characterizations for general utility. A natural question is whether our structural results characterized in Section 3 can be (at least partially) extended to more general utility functions. We conjecture that the miscalibration structure may still hold under broader conditions, although this likely requires a more fine-grained duality-based analysis.

Deterministic predictor. Throughout the paper, we focus on potentially randomized predictors. While randomization is allowed in our framework, in the full version of the paper (Feng and Tang, 2025a), we show that, conditional on each event, the number of possible predictions is bounded by a constant (at most four). These observations suggest that optimal predictors exhibit very limited randomization. With that being said, it is still

[3]By allowing the sender to design the bias assignment, our persuasion problem with signal-dependent bias shares similarities with the literature on ε-IC mechanism design (Balcan et al., 2005; Hartline and Lucier, 2015; Cai et al., 2012; Balseiro et al., 2024). Notably, even when selling a single item to a single buyer, the optimal ε-IC mechanism lacks a simple structural characterization (Balseiro et al., 2024). In contrast, due to the equivalence between the persuasive calibration and this variant, we obtain a clean structural characterization for this persuasion problem.

interesting to explore the design of the optimal (ε, ℓ_t)-calibrated deterministic predictor, which have been widely adopted in practice.

Extensions with multicalibration, multi-class calibration, and beyond. Beyond the ECE metric considered in this work, there is a rich body of literature on broader notions of calibration. We believe that our persuasive calibration framework naturally extends to these settings. For example, by viewing each event as a data point, and assuming a membership structure among events, multicalibration becomes a more appropriate metric. It would be interesting to extend our results to predictors subject to multicalibration constraints. In fact, the computational results and the connections to Bayesian persuasion established in Theorem 5.1 can be extended to the multicalibration setting. The key difference would be in the receiver's behavior – defined in Definition 5.2 – which, under multicalibration, would depend on the group membership of the events.

Moreover, it is also interesting to explore the setting with multi-outcome (i.e., multi-class) calibration. This problem would be particularly technically interesting as predictions must now lie in a high-dimensional simplex. It closely relates to the high-dimensional information design, a notoriously difficult and unresolved area in the theoretical economics literature.

Lastly, in this work, we focus on a setting where the principal has full knowledge of each event's true expected outcome. While this is already a foundational setup, a more realistic and practical scenario may involve the principal only observing noisy realizations of outcomes. In such a case, one must design a predictor based on empirical observations, which we leave it as an important direction for future work.

1.3 Additional Related Work Our work contributes to a growing line of research on calibration in decision-making settings. Beginning with Dawid (1982); Foster and Vohra (1998), it is well established that agents who best respond to calibrated predictions (i.e., those with low K_1 ECE) achieve diminishing swap regret. Recent works have extended this foundation by introducing new decision-theoretic calibration measures that offer finer-grained guarantees on the regret incurred by downstream agents (Kleinberg et al., 2023; Roth and Shi, 2024; Hu and Wu, 2024; Qiao and Zhao, 2025).

For example, Kleinberg et al. (2023) show that a decision-maker cannot fare worse by trusting well-calibrated predictions – the external regret incurred by best-responding to such predictions is linearly (up to a small factor) bounded by the K_1 ECE. They introduce "U-Calibration" – a measure bounded by a small constant times the K_1 ECE, and show that sublinear U-Calibration is both necessary and sufficient for ensuring sublinear regret for all downstream agents. Roth and Shi (2024); Hu and Wu (2024) consider a similar setting but focusing on swap regret. In particular, Roth and Shi (2024) show that by ensuring unbiasedness relative to a carefully chosen collection of events, one can achieve diminishing swap regret for arbitrary downstream agents with an improved rate compared to using calibrated predictions. Hu and Wu (2024) then introduce "Calibration Decision Loss (CDL)", defined as the maximum improvement (over all decision tasks) in decision payoff obtainable by recalibrating the predictions. They show that CDL is upper bounded by twice the K_1 ECE and that a vanishing CDL guarantees vanishing payoff loss from miscalibration across all decision tasks, which also removes the regret dependence on the number of actions appeared in the results of Roth and Shi (2024). Qiao and Zhao (2025) propose a new calibration measure called "subsampled step calibration", which has both decision-theoretic implications and a truthfulness property. In contrast to these works, which primarily adopt a worst-case regret minimization framework and aim to achieve robustness against all decision tasks, our work focuses on characterizing the structure and algorithmic properties of the optimal predictor. In addition, many of the aforementioned works consider a specific principal's goal – aiming to minimize regret across all downstream agents. By contrast, our framework accommodates arbitrary misaligned incentives between the principal and the agent.

Other conceptually related works include Noarov et al. (2023); Haghtalab et al. (2023) who focus on single decision task, Camara et al. (2020); Collina et al. (2024) who explore repeated principal-agent interactions in a prior-free setting, and Feng and Tang (2025b) who study the confusion matrix design for downstream decision-makers. Zhao et al. (2021) introduce "decision calibration" that requires the predicted distribution and true distribution to be indistinguishable to a set of downstream decision-makers. Gopalan et al. (2022) develop an omniprediction framework which aims to design a single predictor that ensures every downstream decision-maker's loss is no worse than some benchmark family of functions (Gopalan et al., 2023; Garg et al., 2024). Guo and Shmaya (2021) analyze a sender-receiver game where the sender, facing miscalibration costs, provides predictions to the receiver. They show that the game's Nash equilibrium crucially depends on the magnitude of the miscalibration costs. Jain and Perchet (2024) study repeated principal-agent interactions in a Bayesian environment. They establish the equivalence between perfectly calibrated predictors and mean-preserving contractions over the prior, and use this equivalence (along with results from the information design literature) to identify the structure of the optimal perfectly calibrated predictor for the principal. Our work can be seen as a strict generalization of Jain and Perchet (2024), extending from the optimal perfectly calibrated predictor to the optimal predictor subject to a given ECE budget and, in addition, providing efficient algorithms to compute them. Notably, the introduction of the ECE budget complicates the direct application of results from the information design literature, as the equivalent model—persuasion with signal-dependent bias introduced in Subsection 5.1—is novel in that context. This necessitates identifying new structural insights, such as miscalibration patterns, and developing new technical components, such as the two-step framework, which is used in both our structural and algorithmic results. There is also a concurrent work by Tang et al. (2025), which study a similar model to ours but focus on different research questions. In particular, Tang et al. (2025) examine a variant of our model where the principal only has sample access to the event-outcome distribution. They develop a learning algorithm with provable sample complexity and time complexity, assuming the principal has an *oracle* to solve for the optimal predictor under the empirical distribution. Our work and Tang et al. (2025) complement each other, as the algorithmic results in Sections 4 and 5 can serve as the oracle required in their algorithm. (As we illustrated, the construction of such an oracle is highly non-trivial.) Additionally, our work sheds light on the structure of the optimal predictor. It is also worth noting that the algorithm in Tang et al. (2025) can only output a predictor that approximately satisfies the ECE budget constraint, whereas the predictors returned by our algorithms fully satisfy the ECE budget constraint.

In this work, we adopt the ℓ_t-norm ECE as a framework to ensure the predictions to be trustworthy to the agents. This ℓ_t-norm calibration metric is broadly accepted in the forecasting community (Dawid, 1982; Foster and Vohra, 1998; Gneiting et al., 2007; Ranjan and Gneiting, 2010), within theoretical computer science (see, e.g., Qiao and Valiant, 2021; Dagan et al., 2024; Casacuberta et al., 2024), empirical machine learning community (see, e.g., Gneiting and Katzfuss, 2014; Rahaman et al., 2021). Beyond ℓ_t-norm ECE and the above mentioned decision-driven calibration metrics, other calibration measures have been proposed to capture different desired aspects, e.g., the multicalibration (Hébert-Johnson et al., 2018; Garg et al., 2024), multi-class calibration (Gopalan et al., 2024), distance to calibration (Błasiok et al., 2023; Qiao and Zheng, 2024; Arunachaleswaran et al., 2025).

2 Preliminaries

In this work, we introduce and study the *"persuasive calibration"* problem between a principal and a downstream decision-maker. Below, we outline the key components of our model.

Environment. Consider a stochastic environment that randomly generates an *event* from a finite set of n events indexed by $[n] \triangleq \{1, 2, \dots, n\}$.[4] We denote by $\lambda_i \in [0, 1]$ the probability of event i, satisfying $\sum_{i\in[n]} \lambda_i = 1$. Once an event is realized, it further induces a randomized *binary outcome*. Specifically, for each event $i \in [n]$, the binary outcome $y \in \mathcal{Y} \triangleq \{0, 1\}$ is drawn from a Bernoulli distribution with mean $\theta_i \in [0, 1]$. Without loss of generality, we assume that the events are sorted in non-decreasing order of their Bernoulli means: $\theta_1 \le \theta_2 \le \dots \le \theta_n$.

There is a *principal* (she) and a downstream decision-maker (he), referred to as the *agent*. The principal observes the realized event i but not the final binary outcome y, while the agent observes neither the event nor the outcome. We detail the interaction between the principal and the agent below.

Calibrated predictors. The principal provides predictions of the final binary outcome (which is unknown to both herself and the agent) to the agent, who then makes a decision that affects the payoffs of both the principal and himself. A *prediction* $p \in [0, 1]$ is a scalar indicating the probability that the binary outcome is one. The principal's predictions are generated by a (possibly randomized) *predictor* $\tilde{f} = \{\tilde{f}_i\}_{i\in[n]}$, which specifies a profile of n conditional distributions.[5] Specifically, for each event i, the conditional predictor $\tilde{f}_i$ represents the distribution over predictions given that event i occurs. We define the conditional prediction space $\mathsf{supp}(\tilde{f}_i)$ as the support of conditional distribution $\tilde{f}_i$ and (unconditional) prediction space $\mathsf{supp}(\tilde{f}) \triangleq \cup_{i\in[n]}\mathsf{supp}(\tilde{f}_i)$. With slight abuse of notation, we use $\tilde{f}_i(p)$ to denote the probability mass (or probability density if $\mathsf{supp}(\tilde{f}_i)$ is continuous) of generating prediction p when event i is realized.[6] We are also interested in the *marginalized predictor*, denoted by $f \in \Delta([0, 1])$, which marginalizes the predictor $\tilde{f}$ over the events. Specifically, we define $f(p) \triangleq \sum_{i\in[n]} \lambda_i \tilde{f}_i(p)$ to represent the

[4]Equivalently, event distribution λ can be viewed as a continuum population of events.

[5]Since the principal observes the realized event i but not the outcome y, her predictor is a function of the realized event that does not depend on the outcome.

[6]While predictors with continuous prediction space are considered, our results show that there exists optimal predictors with finite prediction space.

marginal probability of generating the prediction p.

The predictor provided by the principal is required to be *trustworthy*. In our model, the trustworthiness of predictors follows the calibration framework: given a fixed predictor $\tilde{f}$, for every prediction $p \in \mathsf{supp}(\tilde{f})$, we denote by $\kappa(p) \triangleq \mathbb{E}[y \mid p]$ the *true expected outcome* conditional on the realized prediction p, where the randomness is taken over event and outcome. Since the outcome is binary, it also represents the true probability that binary outcome is one, conditional on the realized prediction p. As a sanity check, if the prediction space $\mathsf{supp}(\tilde{f})$ is discrete, the true expected outcome $\kappa(p)$ can be expressed according to Bayes' rule as

$$\kappa(p) \triangleq \mathbb{E}[y \mid p] = \frac{\sum_{i\in[n]} \lambda_i \cdot \tilde{f}_i(p) \cdot \theta_i}{\sum_{i\in[n]} \lambda_i \cdot \tilde{f}_i(p)} \tag{2.1}$$

A predictor $\tilde{g}$ is *perfectly calibrated* if, for every prediction $q \in \mathsf{supp}(\tilde{g})$, the true expected outcome $\kappa(q)$ equals the prediction itself, i.e., $\kappa(q) = q$.[7] For illustration, Example 2.1 presents two distinct perfectly calibrated predictors.

Example 2.1 (Perfect calibration). *Suppose there are $n = 2$ events with $\lambda_1 = \lambda_2 = 0.5$ and $\theta_1 = 0.3, \theta_2 = 0.9$. Two perfectly calibrated predictors $\tilde{g}^\dagger$ and $\tilde{g}^\ddagger$ can be constructed as follows:*

- *In the first predictor $\tilde{g}^\dagger$, deterministic prediction $q = 0.6$ is generated regardless of the realized event. Specifically, $\tilde{g}_i^\dagger(q) \triangleq \mathbf{1}\{q = 0.6\}$ for both $i \in [2]$.[8] Note that the ex ante probability of outcome being one is $0.5 \cdot 0.3 + 0.5 \cdot 0.9 = 0.6$. Hence, the construction ensures that $\kappa(0.6) = 0.6$ and thus predictor $\tilde{g}^\dagger$ is perfectly calibrated.*
- *In the second predictor $\tilde{g}^\ddagger$, prediction $q = \theta_i$ is generated when event i is realized. Specifically, $\tilde{g}_i^\ddagger(q) \triangleq \mathbf{1}\{q = \theta_i\}$ for both $i \in [2]$. By construction, predictor $\tilde{g}^\ddagger$ is also perfectly calibrated.*

We adopt the standard ℓ_t-norm expected calibration error metric to quantify the deviation of a predictor from being perfectly calibrated.

Definition 2.2 (Expected calibration error). *Fix any $t \in [1, \infty)$. The ℓ_t-*norm expected calibration error *(*ECE*) $\mathbf{ECE}_t\left[\tilde{f}\right]$ of predictor $\tilde{f}$ is*

$$\mathbf{ECE}_t\left[\tilde{f}\right] \triangleq \left(\mathbb{E}\left[|\kappa(p) - p|^t\right]\right)^{\frac{1}{t}} ,$$

*where the expectation is taken over the randomness of both the event and the prediction. Similarly, the ℓ_∞-*norm ECE *of predictor $\tilde{f}$ is $\mathbf{ECE}_\infty\left[\tilde{f}\right] \triangleq \max_{p\in\mathsf{supp}(\tilde{f})} |\kappa(p) - p|$.*

Given any $\varepsilon \geq 0$ and $t \geq 1$, we say a predictor $\tilde{f}$ is (ε, ℓ_t)*-calibrated* if its ℓ_t-norm ECE is at most ε, i.e., $\mathbf{ECE}_t\left[\tilde{f}\right] \leq \varepsilon$. We refer to ε as the *ECE budget*, and denote by $\mathcal{F}_{(\varepsilon,\ell_t)} \triangleq \{\tilde{f} : \mathbf{ECE}_t\left[\tilde{f}\right] \leq \varepsilon\}$ the space of all (ε, ℓ_t)-calibrated predictors. Note that when the ECE budget is set to zero, the $(0, \ell_t)$-calibration recovers perfect calibration for all ℓ_t norms. Thus, we write $\mathcal{F}_0 \equiv \mathcal{F}_{(0,\ell_t)}$.

Among all ℓ_t-norm ECEs, the most classic and important one is the ℓ_1-norm ECE, also known as the K_1 ECE. Additionally, the ℓ_2-norm ECE and ℓ_∞-norm ECE, referred to as the K_2 and K_∞ ECEs, respectively, are also standard in the literature and commonly used in practice. While our results apply to general ℓ_t-norm ECEs, more refined and improved results are derived for these classic ECEs, particularly the K_1 ECE. When the ℓ_t norm is clear from the context, we sometimes simplify notations and omit them, e.g., writing $\mathbf{ECE}\left[\tilde{f}\right]$, $\mathcal{F}_\varepsilon$ instead of $\mathbf{ECE}_t\left[\tilde{f}\right]$, $\mathcal{F}_{(\varepsilon,\ell_t)}$, and saying ε-calibrated instead of (ε, ℓ_t)-calibrated.

Example 2.3 (ε-calibrated predictor). *Consider the same two-event setting as in Example 2.1 and two predictors, $\tilde{f}^\dagger$ and $\tilde{f}^\ddagger$, constructed as follows:*

[7]Throughout the paper, we use $\tilde{g}$ to denote a perfectly calibrated predictor and $\tilde{f}$ to denote a general (possibly imperfectly calibrated) predictor. Similarly, we use q to denote the prediction generated by perfectly calibrated predictors and p to denote the prediction generated by general (possibly imperfectly calibrated) predictors.

[8]We use $\mathbf{1}\{\cdot\}$ to denote the indicator function.

- *In the first predictor $\tilde{f}^\dagger$, the deterministic prediction $p = 0.4$ is generated regardless of the realized event. Specifically, $\tilde{f}_i^\dagger(p) = \mathbf{1}\{p = 0.4\}$ for both $i \in [2]$. For all $t \geq 1$, its ℓ_t-norm ECEs are identical and equal to $\mathbf{ECE}_t\left[\tilde{f}^\dagger\right] = 0.2$.*
- *In the second predictor $\tilde{f}^\ddagger$, predictions $p = 0.4$ and $p = 0.7$ are generated under events 1 and 2, respectively. Specifically, $\tilde{f}_1^\ddagger(p) = \mathbf{1}\{p = 0.4\}$ and $\tilde{f}_2^\ddagger(p) = \mathbf{1}\{p = 0.7\}$. For this predictor, the K_1, K_2 and K_∞ ECEs are $\mathbf{ECE}_1\left[\tilde{f}^\ddagger\right] = 0.15$, $\mathbf{ECE}_2\left[\tilde{f}^\ddagger\right] = \sqrt{0.025} \approx 0.158$, $\mathbf{ECE}_\infty\left[\tilde{f}^\ddagger\right] = 0.2$, respectively.*

Agent's problem. The agent (downstream decision-maker) has a set of actions, denoted by $\mathcal{A}$. His utility function is given by $v(\cdot,\cdot) : \mathcal{A} \times \mathcal{Y} \to \mathbb{R}$, which depends on both the chosen action and the binary outcome of the event.[9] We use $v(a, y)$ to denote the agent's utility when he takes action $a \in \mathcal{A}$ and the outcome $y \in \mathcal{Y}$ is realized.

The agent has no knowledge of the environment, the realized outcome, or the details of the predictor. He can only use the realized prediction p generated by the principal's predictor to inform his decision. Knowing that the predictions are generated by a predictor with small calibration error, the agent *trusts* the principal's prediction p and chooses the action that maximizes his expected utility assuming the outcome is equal to one with probability p.[10] More formally, for any prediction p generated by an ε-calibrated predictor $\tilde{f}$, the agent takes the action $a^*(p) \in \mathcal{A}$ that satisfies[11]

$$a^*(p) \in \arg\max_{a\in\mathcal{A}} \; \mathbb{E}_{y\sim\mathsf{Bern}(p)}[v(a, y)] = \arg\max_{a\in\mathcal{A}} \; p \cdot v(a, 1) + (1 - p) \cdot v(a, 0) \tag{2.2}$$

Principal's problem. The principal's utility depends on the agent's action, the realized outcome, and the realized event. Specifically, for each event $i \in [n]$, agent's action $a \in \mathcal{A}$, and outcome $y \in \mathcal{Y}$, we denote the principal's utility by $u_i(a, y) \in \mathbb{R}^+$.[12] To simplify the presentation, we also introduce the *indirect utility function* U_i, which maps a given prediction p to the principal's expected utility when the agent selects his best action $a^*(p)$ upon the realization of event i.

Definition 2.4 (Indirect utility). *For each event $i \in [n]$ and prediction $p \in \mathsf{supp}(\tilde{f})$, the principal's* indirect utility *is $U_i(p) \triangleq \theta_i \cdot u_i(a^*(p), 1) + (1 - \theta_i) \cdot u_i(a^*(p), 0)$, where $a^*(p)$ is defined in Eqn. (2.2).*[13]

Given a predictor $\tilde{f}$, the principal's expected utility is expressed as[14]

$$\mathbf{Payoff}\left[\tilde{f}\right] \triangleq \mathbb{E}_{i\sim\lambda, p\sim\tilde{f}_i, y\sim\mathsf{Bern}(\theta_i)}[u_i(a^*(p), y)] = \sum_{i\in[n]} \sum_{p\in\mathsf{supp}(\tilde{f}_i)} \lambda_i \cdot \tilde{f}_i(p) \cdot U_i(p)$$

where the second equality holds due to the definition of the indirect utility.

Fixing an exogenously given ECE budget ε and ℓ_t-norm calibration error, we say an (ε, ℓ_t)-calibrated predictor $\tilde{f}^*$ is *optimal* if it maximizes the principal's expected utility among all (ε, ℓ_t)-calibrated predictors, i.e.,

$$\tilde{f}^* \in \arg\max_{\tilde{f}\in\mathcal{F}_{(\varepsilon,\ell_t)}} \; \mathbf{Payoff}\left[\tilde{f}\right]$$

We use the tuple $(\Theta, \lambda, u, v, \mathcal{A}, (\varepsilon, \ell_t))$ to refer to as a persuasive calibration instance, where $\Theta = \{\theta_i\}_{i\in[n]}$ represents the profile of expected outcomes for all events.

[9]Notably, the agent's utility does not depend on the event, while the principal's utility may depend on the event, besides the agent's action and the binary outcome.

[10]The behavior of agent directly best responding to the provided predictions that are generated from a low-ECE predictor has been widely adopted and considered in calibration literature, see, e.g., Haghtalab et al. (2023); Garg et al. (2024); Kleinberg et al. (2023); Hu and Wu (2024); Roth and Shi (2024); Qiao and Zhao (2025).

[11]Following the literature, we assume the agent breaks ties in favor of the principal when he faces indifferent actions. As a result, the principal's indirect utility function under each event preserves the upper semicontinuity. This property guarantees the principal's problem always feasible.

[12]Throughout the paper, we consider non-negative principal's utility, and we do not make any other assumptions unless explicitly stated.

[13]Notably, when the principal evaluates her indirect utility $U_i(p)$, the outcome follows the Bernoulli distribution with mean θ_i instead of prediction p.

[14]When $\tilde{f}_i$ is continuous, the inner summation in the expression of $\mathbf{Payoff}\left[\tilde{f}\right]$ becomes integral, i.e., $\int_0^1 \lambda_i \cdot \tilde{f}_i(p) \cdot U_i(p)\, \mathrm{d}p$. To avoid redundancy, in the remainder of the paper, we present formulas for either the continuous or discrete distribution, as the analogous version can be easily derived.

Timeline. We formalize the timeline of the model as follows:

1. The principal designs a (ε, ℓ_t)-calibrated predictor $\tilde{f}$.[15]

2. A randomized event $i \in [n]$ is realized with probability λ_i, and a (randomized) prediction p is sent to the agent according to the conditional distribution $\tilde{f}_i(\cdot)$.

3. The agent selects his best action $a^*(p)$ as defined in Eqn. (2.2).

4. The binary outcome y is realized from a Bernoulli distribution with mean θ_i. The principal and the agent receive utilities $u_i(a^*(p), y)$ and $v(a^*(p), y)$, respectively.

3 Structural Characterization for the Outcome-Independent Setting

In this section, we focus on the *outcome-independent setting*, where the principal's indirect utility does not depend on the realized event. Specifically, we assume that $U_i(p) \equiv U_j(p)$ for all events $i, j \in [n]$ and predictions $p \in [0, 1]$.[16] Within this section, we omit the subscript i in the principal's indirect utility function, i.e., write $U(p)$ instead of $U_i(p)$.

In Subsection 3.1, we focus on the ℓ_1-norm ECE (a.k.a., K_1 calibration error) and present the structural characterization of the optimal ε-calibrated predictor. In Subsection 3.2, we introduce a two-step framework for generating (ε, ℓ_t)-calibrated predictors for general $t \geq 1$. We utilize this two-step framework to prove our structural characterization of the optimal ε-calibrated predictor.

3.1 Structural Results for the Optimal (ε, ℓ_1)-Calibrated Predictor

In this section, we focus on the ℓ_1-norm ECE, which is the most classic and important calibration error definition, and we present the structural characterization of the optimal (ε, ℓ_1)-calibrated predictor. Except for the theorem statements, we omit ℓ_1 for the presentation simplicity in this subsection.

We first introduce the following concepts of under-confidence and over-confidence, which are introduced and observed in the machine learning literature (e.g., Guo et al., 2017; Kuleshov et al., 2018; Fan et al., 2023).

Definition 3.1 (Under-/over-confidence). *Fix any predictor $\tilde{f}$. A prediction $p \in \mathsf{supp}(\tilde{f})$ is* under-confident *(resp.* over-confident, perfectly calibrated*) if $p \leq \kappa(p)$ (resp. $p \geq \kappa(p)$, $p = \kappa(p)$), where κ is the true expected outcome function of predictor $\tilde{f}$ defined in Eqn.* (2.1).

We now present the main structural characterization of the optimal predictor.[17]

Theorem 3.2 (Structure of optimal ε-calibrated predictor). *For every persuasive calibration instance in the outcome-independent setting, there exists $0 \leq p_{\mathsf{L}} \leq p_{\mathsf{H}} \leq 1$ that partitions the prediction space $[0, 1]$ into* under-confidence interval $[0, p_{\mathsf{L}}]$, perfectly calibrated internal $[p_{\mathsf{L}}, p_{\mathsf{H}}]$, *and* over-confidence *interval $[p_{\mathsf{H}}, 1]$. Then, there exists an optimal (ε, ℓ_1)-calibrated predictor $\tilde{f}^*$ that satisfies the following three properties:*

(1) (Miscalibration Structure) *Every prediction p in the under-confidence interval (resp. over-confidence interval, perfectly calibrated interval) is under-confident (resp. over-confident, perfectly calibrated), i.e.,*

$$
\begin{aligned}
\forall p \in \mathsf{supp}(\tilde{f}^*) \cap [0, p_{\mathsf{L}}]: &\qquad p \leq \kappa(p) \\
\forall p \in \mathsf{supp}(\tilde{f}^*) \cap [p_{\mathsf{L}}, p_{\mathsf{H}}]: &\qquad p = \kappa(p) \\
\forall p \in \mathsf{supp}(\tilde{f}^*) \cap [p_{\mathsf{H}}, 1]: &\qquad p \geq \kappa(p)
\end{aligned}
$$

(2) (Payoff Structure) *All points $(p, U(p))_{p \in \mathsf{supp}(\tilde{f}^*)}$ form a convex function. Moreover, there exists $\alpha \geq 0$ such that points $(p, U(p))$ for all predictions in over-confidence interval are collinear (with slope α), and points $(p, U(p))$ for all predictions in under-confidence interval are collinear (with slope $-\alpha$).*

---

[15] Note the ECE of a predictor is evaluated ex ante, thus, it does not matter when the principal determines the predictor details.

[16] Such outcome-independent indirect utility assumptions are standard and widely studied in the algorithmic game theory and economics literature (e.g., Dworczak and Martini, 2019; Lipnowski and Ravid, 2020; Feng et al., 2022; Arieli et al., 2023; Corrao and Dai, 2023; Feng et al., 2024). In our model, it is equivalent to assume the principal's utility u does not depend on the event and outcome.

[17] Theorem 3.2 and its analysis also holds when there is continuum population of events.

Interpretation of the payoff structure. Theorem 3.2 also provides insights into the principal's indirect utility for predictions generated by the optimal predictor. Specifically, the marginal indirect utility derived from predictions in the under-confidence interval $[0, p_{\mathsf{L}}]$ and over-confidence interval $[p_{\mathsf{H}}, 1]$ is higher than the marginal indirect utility derived from predictions in the perfectly calibrated interval $[p_{\mathsf{L}}, p_{\mathsf{H}}]$. This aligns with the intuition that the optimal predictor is willing to incur ECE to generate predictions with higher payoff. Furthermore, our result suggests that the marginal indirect utility is the same for all predictions in the under-confidence interval $[0, p_{\mathsf{L}}]$ and the over-confidence interval $[p_{\mathsf{H}}, 1]$, and is larger than that for predictions in the perfectly calibrated interval $[p_{\mathsf{L}}, p_{\mathsf{H}}]$.

Graphical characterization of the optimal predictor. Theorem 3.2 and its analysis also admit a graphical interpretation as follows.

Definition 3.3 (Symmetric linear-tailed convex function). *A function $\Gamma : [0, 1] \to \mathbb{R}$ is* symmetric linear-tailed convex *if it is convex, and there exists $0 \le x_{\mathsf{L}} \le x_{\mathsf{H}} \le 1$ such that (i) Γ is linear on both intervals $[0, x_{\mathsf{L}}]$ and $[x_{\mathsf{H}}, 1]$; and (ii) the absolute values of the slopes of these linear parts are equal. We refer to the intervals $[0, x_{\mathsf{L}}]$ and $[x_{\mathsf{H}}, 1]$ as the* linear tails *of the function Γ.*

Proposition 3.4. *For every persuasive calibration instance in the outcome-independent setting, fix an optimal (ε, ℓ_1)-calibrated predictor $\tilde{f}^*$ satisfying the properties in Theorem 3.2. There exists a symmetric linear-tailed convex function $\Gamma : [0, 1] \to \mathbb{R}$ such that (i) $\Gamma(p) \ge U(p)$ for all $p \in [0, 1]$; (ii) $\mathsf{supp}(\tilde{f}^*) \subseteq \{p \in [0, 1] : \Gamma(p) = U(p)\}$; (iii) it is linear over $[0, p_{\mathsf{L}}]$ and $[p_{\mathsf{H}}, 1]$.*

In Figure 1.1, we illustrate the optimal predictor characterized in Theorem 3.2 along with the corresponding symmetric linear-tailed convex function Γ from Proposition 3.4. When the principal's indirect utility function is general, as shown in Figure 1.1, we observe the (non-degenerate) under-confidence, perfectly calibrated, and over-confidence intervals structure. The symmetric linear-tailed convex function Γ exhibits linear tails at both ends, covering the under-confidence and over-confidence intervals.

3.2 Two-Step Framework to Generate ε-Calibrated Predictors

In this section, we provide a systematic way for the principal to design and analyze the predictor. At a high level, the process consists of two steps. First, a perfectly calibrated predictor is constructed. Second, we apply a simple transformation from this perfectly calibrated predictor to obtain an ε-calibrated predictor.[18] As we will illustrate in later sections, this two-step framework and its generalization play an important role in computing and analyzing the optimal ε-calibrated predictor.

Our two-step framework relies on the *post-processing plan* defined as follows:

Definition 3.5 (Event-independent post-processing plan). *An* (event-independent) post-processing plan *χ is a joint distribution over $[0, 1] \times [0, 1]$. We denote by $\chi(q, p)$ the probability density (or probability mass for discrete distribution) of $(q, p) \in [0, 1]^2$.*

Fix any perfectly calibrated predictor $\tilde{g}$ and any $t \ge 1, \varepsilon \ge 0$. A post-processing plan χ is feasible *under perfectly calibrated predictor $\tilde{g}$ and ℓ_t-norm ECE budget ε if it satisfies*

$$\text{(Calibration-Feasibility)} \qquad \left(\int_0^1 \int_0^1 \chi(q, p) \cdot |q - p|^t \, \mathrm{d}p\mathrm{d}q\right)^{\frac{1}{t}} \le \varepsilon$$

$$\text{(Supply-Feasibility)} \qquad \forall q \in [0, 1] : \quad \int_0^1 \chi(q, p) \, \mathrm{d}p = g(q)$$

where g is marginalized predictor induced by $\tilde{g}$, defined as $g(q) \triangleq \sum_{i \in [n]} \lambda_i \tilde{g}_i(q)$ for every $q \in \mathsf{supp}(\tilde{g})$. When $\tilde{g}, t, \varepsilon$ are clear from the context, we simply refer to χ as a feasible post-processing plan.

Intuitively, one can view the post-processing plan χ as a joint distribution over the prediction $q \in \mathsf{supp}(\tilde{g})$ and the prediction p that *"miscalibrates"* the true expected outcome q to be p. We say that χ is event-independent because

[18] All results and discussions in Subsection 3.2 apply to all ℓ_t-norm ECEs. Thus, except for the theorem statements, we omit ℓ_t for simplicity in this subsection.

this joint distribution does not depend on the realized event. We formalize this intuition through the following concept of a *post-processing procedure*. We also illustrate it using Example 3.7.[19]

DEFINITION 3.6 (Event-independent post-processing procedure). *Fix any perfectly calibrated predictor $\tilde{g}$ and any $t \geq 1, \varepsilon \geq 0$. Given any feasible post-processing plan χ, a new (possibly imperfectly calibrated) predictor $\tilde{f}$ is* generated *as follows: for every event $i \in [n]$, and prediction $p \in [0,1]$,*

$$\tilde{f}_i(p) \triangleq \int_0^1 \frac{\tilde{g}_i(q)}{g(q)} \cdot \chi(q,p)\,\mathrm{d}q \ . \tag{3.1}$$

To simplify the presentation, we also say predictor $\tilde{f}$ is generated by *$\tilde{g}$ and χ.*

EXAMPLE 3.7 (Predictors generated from post-processing procedure). *Consider the same two-event setting and four predictors $\tilde{g}^\dagger, \tilde{g}^\ddagger, \tilde{f}^\dagger, \tilde{f}^\ddagger$ as in Examples 2.1 and 2.3. Predictors $\tilde{f}^\dagger, \tilde{f}^\ddagger$ can be generated by the perfectly calibrated predictors $\tilde{g}^\dagger, \tilde{g}^\ddagger$ via Definition 3.6 with the following post-processing plans:*

- *Construct post-processing plan $\chi^\dagger$ with $\chi^\dagger(q,p) = \mathbf{1}\{q = 0.6, p = 0.4\}$. Then, predictor $\tilde{f}^\dagger$ is generated by perfectly calibrated predictor $\tilde{g}^\dagger$ and post-processing plan $\chi^\dagger$.*
- *Construct post-processing plan $\chi^\ddagger$ with $\chi^\ddagger(q,p) = 0.5 \cdot \mathbf{1}\{q = \theta_1, p = 0.4 \text{ or } q = \theta_2, p = 0.7\}$. Then, predictor $\tilde{f}^\ddagger$ is generated by perfectly calibrated predictor $\tilde{g}^\ddagger$ and post-processing plan $\chi^\ddagger$.*

With SUPPLY-FEASIBILITY of the feasible post-processing plan χ, the predictor $\tilde{f} = (\tilde{f}_i)_{i\in[n]}$ generated in Eqn. (3.1) is indeed well-defined as each $\tilde{f}_i$ is a valid distribution over $[0,1]$. In addition, CALIBRATION-FEASIBILITY ensures that predictor $\tilde{f}$ is ε-calibrated.

PROPOSITION 3.8. *Fix any perfectly calibrated predictor $\tilde{g}$, any $t \geq 1, \varepsilon \geq 0$, and any feasible post-processing plan χ. The predictor $\tilde{f}$ generated by $\tilde{g}$ and χ (in Definition 3.6) is well-defined and (ε, ℓ_t)-calibrated, i.e., $\mathbf{ECE}_t\left[\tilde{f}\right] \leq \varepsilon$.*

Proof. We first verify that for each event $i \in [n]$, $\tilde{f}_i$ is a valid distribution on $[0,1]$:

$$\int_0^1 \tilde{f}_i(p)\,\mathrm{d}p \overset{(a)}{=} \int_0^1 \int_0^1 \frac{\tilde{g}_i(q)}{g(q)} \cdot \chi(q,p)\,\mathrm{d}q\mathrm{d}p = \int_0^1 \frac{\tilde{g}_i(q)}{g(q)} \int_0^1 \chi(q,p)\,\mathrm{d}p\mathrm{d}q \overset{(b)}{=} \int_0^1 \tilde{g}_i(q)\,\mathrm{d}q = 1$$

where equality (a) holds due to the construction of predictor $\tilde{f}$ in Definition 3.6, and equality (b) holds due to SUPPLY-FEASIBILITY of post-processing plan χ.

We next bound the ℓ_t-norm ECE of predictor $\tilde{f}$:

$$\begin{aligned}
\left(\mathbf{ECE}_t\left[\tilde{f}\right]\right)^t &= \mathbb{E}\left[|\kappa(p) - p|^t\right] \\
&= \int_0^1 \left|\sum\nolimits_{i\in[n]} \lambda_i \tilde{f}_i(p) \cdot (p - \theta_i)\right|^t \mathrm{d}p \\
&\overset{(a)}{=} \int_0^1 \left|\sum\nolimits_{i\in[n]} \lambda_i \cdot (p-\theta_i) \int_0^1 \frac{\tilde{g}_i(q)}{g(q)} \cdot \chi(q,p)\,\mathrm{d}q\right|^t \mathrm{d}p \\
&\overset{(b)}{=} \int_0^1 \left|\int_0^1 p \cdot \chi(q,p)\,\mathrm{d}q - \int_0^1 \sum\nolimits_{i\in[n]} \lambda_i\theta_i \cdot \frac{\tilde{g}_i(q)}{g(q)} \cdot \chi(q,p)\,\mathrm{d}q\right|^t \mathrm{d}p \\
&\overset{(c)}{=} \int_0^1 \left|\int_0^1 p \cdot \chi(q,p)\,\mathrm{d}q - \int_0^1 q \cdot \chi(q,p)\,\mathrm{d}q\right|^t \mathrm{d}p \\
&\overset{(d)}{\leq} \int_0^1 \int_0^1 \chi(q,p) \cdot |p-q|^t\,\mathrm{d}q\mathrm{d}p \overset{(e)}{\leq} \varepsilon^t \ ,
\end{aligned}$$

[19]In fact, due to SUPPLY-FEASIBILITY and Lemma 3.14, a perfectly calibrated predictor $\tilde{g}$ can be determined by a feasible post-processing plan χ. Hence, such a post-processing plan χ is sufficient to generate an imperfectly calibrated predictor via Definition 3.6. Nevertheless, we choose to explicitly express the perfectly calibrated predictor $\tilde{g}$ (or its marginalized version, g), as it reveals useful structural properties–such as the MPC representation in Definition 3.13 and Lemma 3.14–which play an important role in our analysis.

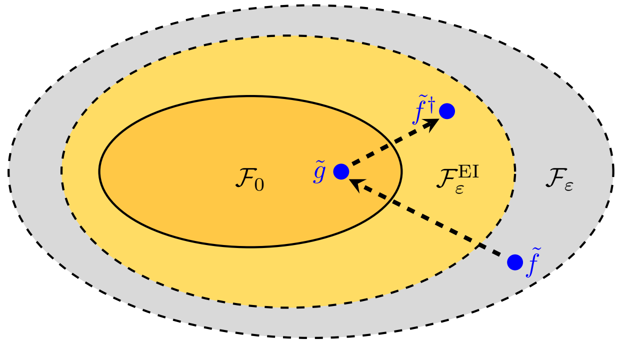

Figure 3.1: An illustration of hierarchical relation between $\mathcal{F}_\varepsilon$, $\mathcal{F}_\varepsilon^{\mathrm{EI}}$ and $\mathcal{F}_0$. The dashed arrow line from predictor $\tilde{f}$ to perfectly calibrated predictor $\tilde{g}$ is established by Proposition 3.10, and the dashed arrow line from $\tilde{g}$ to the predictor $\tilde{f}^{\dagger}$ is established in Definition 3.6 and Proposition 3.8.

where equality (a) holds due to the construction of predictor $\tilde{f}$ in Definition 3.6, equality (b) holds due to the definition of marginalized perfectly calibrated predictor g, equality (c) holds due to the fact that perfectly calibrated predictor $\tilde{g}$ has zero ECE, inequality (d) holds due to the Jensen's inequality and the convexity of $|\cdot|^t$, and inequality (e) holds due to Calibration-Feasibility of post-processing plan χ. This completes the proof of Proposition 3.8. ∎

We remark that various post-processing procedures have been introduced in the calibration literature. However, the majority of these methods focus on transforming a miscalibrated predictor into a perfectly calibrated one or reducing its ECE (e.g., Kim et al., 2019; Fan et al., 2023; Blasiok et al., 2023; Gopalan et al., 2024). That said, in the full version of the paper (Feng and Tang, 2025a), we demonstrate that when the principal and agent have misaligned incentives, allowing the principal to use a predictor with higher ECE may improve both her and the agent's utility–creating a "win-win" scenario. Hence, we believe our post-processing procedure is not only valuable as an analytical tool but may also have practical implications.

Predictor space $\mathcal{F}_\varepsilon^{\mathrm{EI}}$. We denote by $\mathcal{F}_\varepsilon^{\mathrm{EI}}$ the space of all predictors generated by perfectly calibrated predictors and feasible miscalibration plans via the event-independent post-processing procedure (Definition 3.6). Proposition 3.8 guarantees that every predictor $\tilde{f} \in \mathcal{F}_\varepsilon^{\mathrm{EI}}$ is ε-calibrated, and thus

$$\mathcal{F}_\varepsilon^{\mathrm{EI}} \subseteq \mathcal{F}_\varepsilon \ .$$

Example 3.9 shows that the inclusion relation between the two spaces is strict, i.e., there exists a predictor $\tilde{f} \in \mathcal{F}_\varepsilon$ such that $\tilde{f} \notin \mathcal{F}_\varepsilon^{\mathrm{EI}}$, showing a strict inclusion relation between the two spaces.[20] (See Figure 3.1 for the hierarch illustration.) However, when the principal's indirect utility is outcome-independent, we argue that restricting to predictors in $\mathcal{F}_\varepsilon^{\mathrm{EI}}$ is without loss of generality.

Example 3.9 (Strict inclusion relation between $\mathcal{F}_\varepsilon^{\mathrm{EI}}$ and $\mathcal{F}_\varepsilon$). *There are two events and let $\lambda_1 = \lambda_2 = 0.5$ with $\theta_1 = \theta_2 = 0.5$. Consider predictor $\tilde{f}$ with $\tilde{f}_1(0.125) = \tilde{f}_2(0.875) = 1$. It can be verified that such predictor $\tilde{f} \in \mathcal{F}_\varepsilon \backslash \mathcal{F}_\varepsilon^{\mathrm{EI}}$ for every $\varepsilon \geq 0.375$ and $t \geq 1$.*

Proposition 3.10. *For every persuasive calibration instance in the outcome-independent setting, for every (ε, ℓ_t)-calibrated predictor $\tilde{f} \in \mathcal{F}_{(\varepsilon,\ell_t)}$, there exists a pair of perfectly calibrated predictor $\tilde{g}$ and feasible miscalibration plan χ that can generate a (possibly different) predictor $\tilde{f}^{\dagger} \in \mathcal{F}_{(\varepsilon,\ell_t)}^{\mathrm{EI}}$ with the same expected utility for the principal, i.e., $\textbf{Payoff}\left[\tilde{f}^{\dagger}\right] = \textbf{Payoff}\left[\tilde{f}\right]$.*

Specifically, such a pair of perfectly calibrated predictor $\tilde{g}$ and feasible miscalibration plan χ can be constructed

[20]In Section 4, we extend the event-independent post-processing procedure to the event-dependent post-processing procedure, which generates every ε-calibrated predictors from perfectly calibrated predictor.

as follows: for every event $i \in [n]$, and predictions $q, p \in [0,1]$,

$$\tilde{g}_i(q) \triangleq \sum\nolimits_{p\in\mathsf{supp}(\tilde{f}_i):\kappa(p)=q} \tilde{f}_i(p) \;\; \textit{and} \;\; \chi(q,p) \triangleq \sum\nolimits_{i\in[n]} \lambda_i \tilde{f}_i(p)\cdot \mathbf{1}\{\kappa(p)=q\} \tag{3.2}$$

where $\kappa(p)$ is the true expected outcome defined in Eqn. (2.1).

Corollary 3.11. *For every persuasive calibration instance in the outcome-independent setting, there exists an optimal predictor $\tilde{f}^* \in \mathcal{F}^{\mathrm{EI}}_{(\varepsilon,\ell_t)}$, i.e., $\mathbf{Payoff}\Big[\tilde{f}^*\Big] \geq \mathbf{Payoff}\Big[\tilde{f}\Big]$ for every $\tilde{f} \in \mathcal{F}_{(\varepsilon,\ell_t)}$.*

Proof of Proposition 3.10. Fix any ε-calibrated predictor $\tilde{f} \in \mathcal{F}_\varepsilon$. Consider predictor $\tilde{g}$ and post-processing plan χ constructed as in Eqn. (3.2).

We first verify that the constructed predictor $\tilde{g}$ has zero ECE and thus is perfectly calibrated. Let κ and $\kappa^\dagger$ be the true expected outcome function induced by original predictor $\tilde{f}$ and constructed predictor $\tilde{g}$, respectively. Note that for every prediction $q \in \mathsf{supp}(\tilde{g})$, its true expected outcome $\kappa^\dagger$ can be computed as

$$\kappa^\dagger(q) = \frac{\sum_{i\in[n]}\lambda_i\tilde{g}_i(q)\cdot\theta_i}{\sum_{i\in[n]}\lambda_i\tilde{g}_i(q)} = \frac{\sum_{i\in[n]}\lambda_i\sum_{p\in\mathsf{supp}(\tilde{f}_i):\kappa(p)=q}\tilde{f}_i(p)\cdot\theta_i}{\sum_{i\in[n]}\lambda_i\sum_{p\in\mathsf{supp}(\tilde{f}_i):\kappa(p)=q}\tilde{f}_i(p)} = q$$

where the first and third equalities hold due to the definition of true expected outcome function $\kappa^\dagger$ and κ, while the second equality holds due to the construction of predictor $\tilde{g}$.

We next verify that the constructed post-processing plan χ is feasible. Note that for each prediction $q \in [0,1]$,

$$\begin{aligned}\int_0^1 \chi(q,p)\;\mathrm{d}p &= \int_0^1 \sum\nolimits_{i\in[n]}\lambda_i\tilde{f}_i(p)\cdot\mathbf{1}\{\kappa(p)=q\}\;\;\mathrm{d}p \\ &= \sum\nolimits_{i\in[n]}\lambda_i\sum\nolimits_{p\in\mathsf{supp}(\tilde{f}_i):\kappa(p)=q}\tilde{f}_i(p) = \sum\nolimits_{i\in[n]}\lambda_i\cdot\tilde{g}_i(q)\;,\end{aligned}$$

where the first and last equalities follow the construction in Eqn. (3.2). Thus, the Supply-Feasibility (in Definition 3.5) of post-processing plan χ is satisfied. In addition,

$$\begin{aligned}\int_0^1\int_0^1 \chi(q,p)\cdot|q-p|^t\;\mathrm{d}p\mathrm{d}q &= \int_0^1\int_0^1\sum\nolimits_{i\in[n]}\lambda_i\tilde{f}_i(p)\cdot\mathbf{1}\{\kappa(p)=q\}\cdot|q-p|^t\;\;\mathrm{d}p\mathrm{d}q \\ &= \int_0^1\sum\nolimits_{i\in[n]}\lambda_i\tilde{f}_i(p)\cdot|\kappa(p)-p|^t\;\;\mathrm{d}p = \left(\mathbf{ECE}_t\Big[\tilde{f}\Big]\right)^t \leq \varepsilon^t\end{aligned}$$

which implies the Calibration-Feasibility of post-processing plan χ. Thus, χ is feasible.

Invoking Proposition 3.10, we know predictor $\tilde{f}^\dagger$ generated by $\tilde{g}$ and χ is well-defined and (ε,ℓ_t)-calibrated. Finally, we compute the principal's expected utility given predictor $\tilde{f}^\dagger$,

$$\begin{aligned}\mathbf{Payoff}\Big[\tilde{f}^\dagger\Big] &= \sum\nolimits_{i\in[n]}\int_0^1\lambda_i\tilde{f}^\dagger_i(p)\cdot U(p)\,\mathrm{d}p \overset{(a)}{=} \int_0^1\int_0^1\sum\nolimits_{i\in[n]}\lambda_i\frac{\tilde{g}_i(q)}{g(q)}\cdot\chi(q,p)\cdot U(p)\,\mathrm{d}q\,\mathrm{d}p \\ &= \int_0^1\int_0^1\chi(q,p)\cdot U(p)\,\mathrm{d}q\,\mathrm{d}p \overset{(b)}{=} \int_0^1\int_0^1\sum\nolimits_{i\in[n]}\lambda_i\tilde{f}_i(p)\cdot\mathbf{1}\{\kappa(p)=q\}\cdot U\left(p\right)\;\;\mathrm{d}p\mathrm{d}q \\ &= \int_0^1\sum\nolimits_{i\in[n]}\lambda_i\tilde{f}_i(p)\cdot U\left(p\right)\;\;\mathrm{d}p = \mathbf{Payoff}\Big[\tilde{f}\Big]\end{aligned}$$

where equality (a) follows by definition of $\tilde{f}^\dagger$ (as it is constructed according to Eqn. (3.1)), equality (b) follows by the construction of χ defined in Eqn. (3.2). This finishes the proof of Proposition 3.10. ∎

A two-step framework for the principal. Proposition 3.10 (along with Corollary 3.11) establishes an equivalence between ε-calibrated predictors $\tilde{f}$ and pairs consisting of perfectly calibrated predictor $\tilde{g}$ and feasible miscalibration plan χ. This equivalence allows us to adopt a conceptually *two-step approach*, where the principal first designs a perfectly calibrated predictor and then selects a feasible miscalibration plan. The following proposition formalizes this idea and shows that the optimal ε-calibrated predictor can be characterized by LP-TwoStep, an infinite-dimensional linear program, which plays a key role for proving all our structure results in Subsection 3.1.

PROPOSITION 3.12. *For every persuasive calibration instance in the outcome-independent setting, the principal's expected utility under the optimal (ε, ℓ_t)-calibrated predictor $\tilde{f}^*$ is equal to the optimal objective value of the following linear program with variables $\{\chi(q,p), g(q)\}_{q,p\in[0,1]}$:*

$$
\text{(LP-TwoStep)} \qquad
\begin{aligned}
& \max_{\chi\geq 0, g} && \int_0^1\int_0^1 \chi(q,p)\cdot U(p)\ \mathrm{d}p\mathrm{d}q && s.t. \\
& (\textsc{Calibration-Feasibility}) && \int_0^1\int_0^1 \chi(q,p)\cdot |p-q|^t\ \mathrm{d}p\mathrm{d}q \leq \varepsilon^t, && \\
& (\textsc{Supply-Feasibility}) && \int_0^1 \chi(q,p)\ \mathrm{d}p = g(q), && q\in[0,1] \\
& (\textsc{MPC-Feasibility}) && g \in \mathsf{MPC}(\lambda)\ , &&
\end{aligned}
$$

where $g \in \mathsf{MPC}(\lambda)$ is the mean-preserving constraint (see Definition 3.13).

Moreover, every optimal solution of LP-TwoStep can be converted into an optimal (ε, ℓ_t)-calibrated predictor $\tilde{f}^*$ (see Lemma 3.14 and Proposition 3.8).

DEFINITION 3.13 (Mean-preserving contraction). *Fix any distribution λ with support $\mathsf{supp}(\lambda) \subseteq [0,1]$. For any distribution g with support $\mathsf{supp}(g) \subseteq [0,1]$, we say distribution g is a* mean-preserving contraction *of distribution λ if for all $s \in [0,1]$,*

$$\int_0^s G(q)\,\mathrm{d}q \leq \int_0^s \Lambda(q)\,\mathrm{d}q$$

with the equality at $s = 1$, where G and Λ are the cumulative density function of distributions g and λ, respectively. We define $\mathsf{MPC}(\lambda)$ as the set of all distributions that are mean-preserving contractions of the distribution λ.

LEMMA 3.14. *A distribution $g \in \Delta([0,1])$ is a marginalized perfectly calibrated predictor if and only if $g \in \mathsf{MPC}(\lambda)$.*

The above necessary and sufficient condition follows from the well-known equivalence between the distribution of posterior means induced by a signal and the set of mean-preserving contractions of the prior (Blackwell, 1953; Rothschild and Stiglitz, 1978; Gentzkow and Kamenica, 2016). Its complete proof is deferred to the full version of the paper (Feng and Tang, 2025a). We conclude this subsection by proving Proposition 3.12.

Proof of Proposition 3.12. We first prove the direction that for any ε-calibrated predictor $\tilde{f}$, there exists a pair of perfectly calibrated predictor $\tilde{g}$ (let g be its corresponding marginalized perfectly calibrated predictor) and a miscalibration plan χ such that g, χ is a feasible solution to LP-TwoStep, and meanwhile its objective in LP-TwoStep exactly equals to $\textbf{Payoff}\left[\tilde{f}\right]$.

To see this, given the predictor $\tilde{f}$, we consider the predictor $\tilde{g}$, and the miscalibration plan constructed as in Eqn. (3.2). Proposition 3.10 shows that the constructed predictor $\tilde{g}$ is perfectly calibrated and post-processing plan χ is feasible. Let g be the marginalized predictor of $\tilde{g}$. Invoking Lemma 3.14, we know that $g \in \mathsf{MPC}(\lambda)$. Thus, (g, χ) forms a feasible solution to LP-TwoStep. Moreover, the objective value of g, χ in LP-TwoStep is

$$
\begin{aligned}
\int_0^1\int_0^1 \chi(q,p)\cdot U(p)\ \mathrm{d}p\mathrm{d}q &= \int_0^1\int_0^1 \sum\nolimits_{i\in[n]} \lambda_i \tilde{f}_i(p)\cdot \mathbf{1}\{\kappa(p) = q\}\cdot U(p)\ \mathrm{d}p\mathrm{d}q \\
&= \int_0^1 \sum\nolimits_{i\in[n]} \lambda_i \tilde{f}_i(p)\cdot U(p)\ \mathrm{d}p = \textbf{Payoff}\left[\tilde{f}\right]\ .
\end{aligned}
$$

We next prove the reverse direction: given any feasible solution g, χ to LP-TwoStep, there exists an ε-calibrated predictor $\tilde{f}$ such that its payoff exactly equals to the objective value of the solution g, χ in LP-TwoStep.

To see this, we note that from Lemma 3.14, we know for any $g \in \mathsf{MPC}(\lambda)$, there must exist a perfectly calibrated predictor $\tilde{g}$ whose marginalized perfectly calibrated predictor exactly equals to g. Now given such a perfectly calibrated predictor $\tilde{g}$, and together with χ, we consider a predictor $\tilde{f}$ generated according to Eqn. (3.1). Invoking Proposition 3.8, predictor $\tilde{f}$ is well-defined and ε-calibrated. Finally, the principal's expected utility under predictor $\tilde{f}$ is

$$\begin{aligned}\textbf{Payoff}\Big[\tilde{f}\Big] &= \sum\nolimits_{i\in[n]} \lambda_i \int_0^1 \tilde{f}_i(p)\cdot U(p)\ \mathrm{d}p \\ &= \int_0^1\int_0^1 \sum\nolimits_{i\in[n]} \lambda_i \frac{\tilde{g}_i(q)}{g(q)}\cdot \chi(q,p)\cdot U(p)\,\mathrm{d}q\,\mathrm{d}p = \int_0^1\int_0^1 \chi(q,p)\cdot U(p)\,\mathrm{d}q\,\mathrm{d}p\end{aligned}$$

which exactly equals to the objective value of the solution g, χ in LP-TwoStep. This finishes the proof of Proposition 3.12. ∎

Utilizing the above two-step framework, we know that the optimal ε-calibrated predictor can be transformed into an optimal solution of LP-TwoStep and vice versa. Therefore, proving the structural results for the optimal predictor is equivalent to proving their analogs for the optimal solution in LP-TwoStep. However, characterizing the structural properties of the optimal solution in LP-TwoStep is rather technical, and so we defer its proof, together with the complete proofs of Theorem 3.2 and Proposition 3.4, to the full version of the paper (Feng and Tang, 2025a).

4 FPTAS for General Setting In this section, we study the time complexity of computing an optimal (ε, ℓ_t)-calibrated predictor. As the main result of this section, we present an FPTAS (Algorithm 4.1) that works for the general setting (e.g., the sender's indirect utility can be outcome-dependent) and for all ℓ_t-norm ECEs.[21]

THEOREM 4.1 (FPTAS for general setting). *For every persuasive calibration instance, given any $\delta \in (0, 1)$, there exists a linear programming (see LP-DiscTwoStep$^+_\delta$) based algorithm (Algorithm 4.1) that computes a (ε, ℓ_t)-calibrated predictor $\tilde{f}$ achieving a $(1-\delta)$-approximation to the optimal (ε, ℓ_t)-calibrated predictor $\tilde{f}^*$, i.e., $\textbf{Payoff}\Big[\tilde{f}\Big] \geq (1-\delta)\cdot\textbf{Payoff}\Big[\tilde{f}^*\Big]$. The running time of the algorithm is $\mathsf{poly}(1/\delta, n, m)$, where $n = |\Theta|, m = |\mathcal{A}|$ are the number of events and the number of agent's actions, respectively.*

We remark that the running time of Algorithm 4.1 in the above theorem does not depend on the ECE budget ε or the exponent parameter t in the ECE metric. In Section 5, we also present a polynomial-time algorithm to compute the optimal predictor for ℓ_1-norm ECE and ℓ_∞-norm ECE. Determining whether a polynomial-time algorithm for computing the optimal predictor exists for ℓ_t-norm ECE with other exponent parameters $t \in (1, \infty)$, or proving a computational hardness result, is left as an interesting future direction.

In Subsection 4.1, we introduce the FPTAS (Algorithm 4.1) and provide an overview of its two main technical ingredients and their analysis. In Subsections 4.2 and 4.3, we present the full details of these two technical ingredients, respectively. Finally, we prove Theorem 4.1 in Subsection 4.4.

4.1 Algorithm Description and Analysis Overview In this section, we first sketch the main algorithmic ingredients of Algorithm 4.1, along with the intuitions and challenges behind them. We then provide details of each ingredient and their corresponding analysis.

To obtain a time-efficient algorithm for computing an approximately optimal predictor, a natural approach is to formulate the principal's problem as a computationally tractable optimization program (e.g., a linear or convex program). However, due to the calibration error constraint defined on the predictor $\tilde{f}$, this approach is not straightforward. In the full version of the paper (Feng and Tang, 2025a), we explain the failure of some natural optimization program formulations that directly use the predictor as the decision variable and optimize over the entire feasible predictor space $\mathcal{F}_{(\varepsilon,\ell_t)}$.

We overcome the challenge and develop Algorithm 4.1, which contains two main technical ingredients: it combines a generalized two-step framework, which extends the two-step approach in Subsection 3.2 from the outcome-independent setting to the general setting, with a carefully designed discretization scheme.

Ingredient 1: generalized two-step framework. For the general setting where the principal's utility may simultaneously depend on the realized event, the agent's action, and the realized outcome, the two-step framework with the *event-independent* post-processing plan established in Subsection 3.2 may not be sufficient. Motivated by this, we introduce a generalized two-step framework in Subsection 4.2. However, if we consider an event-dependent

[21]In Sections 4 and 5, we assume the agent's action set $\mathcal{A}$ is finite, and denote by $m = |\mathcal{A}|$ the number of actions. Notably, finite action set is commonly assumed in the algorithmic information design literature (Dughmi and Xu, 2017; Babichenko et al., 2024).

Algorithm 4.1 The FPTAS for the persuasive calibration with general $t \geq 1$

Input: persuasive calibration instance $(\Theta, \lambda, u, v, \mathcal{A}, (\varepsilon, \ell_t))$, approximation precision $\delta \in (0, 1)$.
Solve LP-DiscTwoStep$^{+}_{(\delta/3)}$ and let $\{\chi^*_{i,j}\}_{i,j:i\leq j}$ be its optimal solution.
Use the optimal solution $\{\chi^*_{i,j}\}_{i,j:i\leq j}$ to construct the predictor $\tilde{f}$ as follows:

$$\forall i \in [n],\ p \in [0,1]: \quad \tilde{f}_i(p) \leftarrow \frac{1}{\lambda_i} \cdot \int_0^1 \left(\sum_{k\in[i:n]} \chi^*_{i,k}(q,p) \cdot \frac{\theta_k - q}{\theta_k - \theta_i} + \sum_{k\in[i-1]} \chi^*_{k,i}(q,p) \cdot \frac{q - \theta_k}{\theta_i - \theta_k} \right) \mathrm{d}q$$

Output: predictor $\tilde{f}$.

post-processing plan that *fully decouples* the miscalibration across all events, additional non-linear constraints—seemingly intractable—are required. To bypass this issue, we introduce a generalized two-step framework with an *(event-dependent) bi-event post-processing plan* (Definitions 4.2 and 4.3), which allows us to formulate an infinite-dimensional (but tractable) linear program LP-TwoStep^{+}. Our bi-event post-processing plan is rich enough to ensure that there always exists an optimal predictor that can be generated within our generalized framework (Proposition 4.6).

Ingredient 2: Instance-dependent two-layer discretization & rounding-based analysis. Our FPTAS (Algorithm 4.1) solves a discretized version of LP-TwoStep^{+}. Common discretization schemes, such as uniform discretization, are not directly applicable, as they may render the discretized LP infeasible or significantly degrade the objective value due to the ECE budget constraint. To overcome this challenge, in Subsection 4.3, we introduce an instance-dependent discretization scheme (Definition 4.8) with a two-layer structure. First, our scheme constructs a non-uniform $\Theta(\delta)$-net that includes the expected outcomes $\Theta = \{\theta_1, \ldots, \theta_n\}$ and the discontinuity points $\mathcal{Z}$ of the principal's indirect utility function. Second, for each point in this $\Theta(\delta)$-net, it introduces a finer $\Theta(\delta_0)$-net within a small neighborhood, where δ_0 depends not only on the discretization precision δ but also on the ECE budget ε and the norm exponent parameter t. This two-layer structure of our discretization scheme ensures that the discretized LP LP-DiscTwoStep$^{+}_{\delta}$ remains feasible, closely approximates the original objective value, and maintains a polynomial size. Finally, we develop a rounding scheme (Algorithm 4.2) to analyze both the feasibility and the objective value of LP-DiscTwoStep$^{+}_{\delta}$. We believe that both our discretization scheme and our rounding argument are of independent interest in the algorithmic information design literature.

4.2 Generalized Two-step Framework: Bi-Event Post-Processing Plan

In this section, we generalize the two-step framework developed in Subsection 3.2 from the outcome-independent setting to the general setting.

In the outcome-independent setting, where the principal has an outcome-independent indirect utility, Proposition 3.10 shows that it suffices to consider predictors from $\mathcal{F}^{\mathrm{EI}}_{(\varepsilon,\ell_t)}$, which can be implemented via an event-independent post-processing plan (Definition 3.5). However, in the general setting where the principal's utility depends on the realized event, this sufficiency is no longer guaranteed.

Perhaps the most natural generalization is to allow event-dependent post-processing, by extending $\chi(q,p)$ to $\chi_i(q,p)$, which depends on the realized event i. However, *fully decoupling* the post-processing plan across events introduces a significant challenge: it requires us to impose a *non-linear* feasibility constraint that ensures $\kappa(p) = q$ after post-processing according to $\{\chi_i(q,p)\}_{i\in[n]}$.

To address this challenge, rather than fully decoupling the post-processing plans, we propose a more tractable alternative: an *(event-dependent) bi-event post-processing plan*, which captures pairwise miscalibration between events. This formulation offers greater flexibility than the event-independent approach, while avoiding the complexity of full decoupling. The formal definition is as follows.

Definition 4.2 ((Event-dependent) Bi-event post-processing plan). *An* (event-dependent) bi-event post-processing plan χ *is a joint distribution over* $[n] \times [n] \times [0,1] \times [0,1]$. *We denote by* $\chi_{i,j}(q,p)$ *the probability density (or probability mass for discrete distribution) of* $(q,p) \in [0,1]^2$ *under a pair of events* (i,j) *with* $i \leq j$.

Fix any $t \geq 1, \varepsilon \geq 0$. A bi-event post-processing plan χ is feasible *under ℓ_t-norm ECE budget ε if it satisfies*

(BiE-Calibration-Feasibility)

$$\left(\int_0^1 \int_0^1 \sum\nolimits_{(i,j):i\leq j} \chi_{i,j}(q,p) \cdot |q-p|^t \,\mathrm{d}p\mathrm{d}q\right)^{\frac{1}{t}} \leq \varepsilon,$$

and for every event $i \in [n]$,

(BiE-Supply-Feasibility)

$$\int_0^1 \int_0^1 \left(\sum\nolimits_{k\in[i:n]} \chi_{i,k}(q,p) \cdot \omega_{i,k}(q) + \sum\nolimits_{k\in[i-1]} \chi_{k,i}(q,p) \cdot (1-\omega_{k,i}(q))\right) \mathrm{d}p\mathrm{d}q = \lambda_i$$

where auxiliary contribution ratio $\omega_{i,j}(q) \triangleq \frac{\theta_j - q}{\theta_j - \theta_i}$ for $\theta_i < \theta_j$, and $\omega_{i,j}(q) \triangleq 1$ for $\theta_i = \theta_j$. When (t,ε) are clear from the context, we simply refer to χ as a feasible bi-event post-processing plan.

The two feasibility constraints for the bi-event post-processing plan in Definition 4.2 can be viewed as a generalization of the analogous constraints for the event-independent post-processing plan in Definition 3.5. Additionally, it bypasses the challenge that arises when fully decoupling the post-processing plan across events by introducing auxiliary contribution ratios $\{\omega_{i,j}(q)\}$. The intuition is as follows: consider any pair of events (i,j) and prediction q. Events i and j are combined to form a true expected outcome q *if and only if* the probability mass contributed by event i is proportional to $\omega_{i,j}(q)$, while the contribution from event j is proportional to $1-\omega_{i,j}(q)$. We formalize our intuition in Definition 4.3 and Proposition 4.5 below. We also illustrate it using Example 4.4.[22]

Definition 4.3 (Bi-event post-processing procedure). *Fix any $t \geq 1, \varepsilon \geq 0$. Given any feasible bi-event post-processing plan χ, a new (possibly imperfectly calibrated) predictor $\tilde{f}$ is* generated *as follows: for every event $i \in [n]$, and prediction $p \in [0,1]$,*

$$\tilde{f}_i(p) \leftarrow \frac{1}{\lambda_i} \cdot \int_0^1 \left(\sum\nolimits_{k\in[i:n]} \chi_{i,k}(q,p) \cdot \omega_{i,k}(q) + \sum\nolimits_{k\in[i-1]} \chi_{k,i}(q,p) \cdot (1-\omega_{k,i}(q))\right) \mathrm{d}q \tag{4.1}$$

To simplify the presentation, we also say predictor $\tilde{f}$ is generated by *bi-event post-processing plan χ.*

Example 4.4 (Predictors generated from bi-event post-processing procedure). *Consider the same two-event setting and predictor $\tilde{f}$ as in Example 3.9. Predictor $\tilde{f}$ can be generated by the following bi-event post-processing plan χ: $\chi_{1,1}(0.5, 0.125) = 0.5$ and $\chi_{2,2}(0.5, 0.875) = 0.5$. It can be verified that for every $t \geq 1$ and $\varepsilon \geq \mathbf{ECE}_t\left[\tilde{f}\right]$, bi-event post-processing plan χ is feasible under ℓ_t-norm ECE budget ε.*

Proposition 4.5. *Fix any $t \geq 1, \varepsilon \geq 0$, and any feasible bi-event post-processing plan χ. The predictor $\tilde{f}$ generated by χ (in Definition 4.3) is well-defined and (ε, ℓ_t)-calibrated, i.e., $\mathbf{ECE}_t\left[\tilde{f}\right] \leq \varepsilon$.*

Proof of Proposition 4.5. Fix any feasible bi-event post-processing plan χ. Consider predictor $\tilde{f}$ generated by χ (according to Eqn. (4.1) in Definition 4.3). Note that predictor $\tilde{f}$ is well-defined since $\int_0^1 \tilde{f}_i(p)\mathrm{d}p = 1$ for every event $i \in [n]$, which is implied by BiE-Supply-Feasibility (in Definition 4.2) of bi-event post-processing plan χ.

We next bound the ℓ_t-ECE of predictor $\tilde{f}$. Note that

$$\tilde{f}(p) = \sum\nolimits_{i\in[n]} \lambda_i \tilde{f}_i(p) = \int_0^1 \sum\nolimits_{(i,j):i\leq j} \chi_{i,j}(q,p)\,\mathrm{d}q\ .$$

We next argue that for any prediction $p \in [0,1]$, if $\tilde{f}(p) > 0$, we have

$$|p - \kappa(p)|^t \leq \int_0^1 \frac{1}{\tilde{f}(p)} \sum\nolimits_{(i,j):i\leq j} \chi_{i,j}(q,p) \cdot |p-q|^t \,\mathrm{d}q \tag{4.2}$$

[22] Unlike our (event-independent) post-processing plan, in Definition 4.2, we do not explicitly consider the perfectly calibrated predictor, as our current goal is to develop an LP for computation rather than characterizing its structure. See also Footnote 19.

To see the above inequality, we notice that the true expected outcome $\kappa(p)$ of prediction p under predictor $\tilde{f}$ is

$$\begin{aligned}\kappa(p) &= \frac{1}{\tilde{f}(p)} \sum\nolimits_{i\in[n]} \theta_i \cdot \left(\int_0^1 \sum\nolimits_{k\in[i:n]} \chi_{i,k}(q,p)\omega_{i,k}(q)\,\mathrm{d}q + \int_0^1 \sum\nolimits_{k\in[i-1]} \chi_{k,i}(q,p)(1-\omega_{k,i}(q))\,\mathrm{d}q\right)\\ &= \frac{1}{\tilde{f}(p)} \sum\nolimits_{(i,j):i\le j} \int_0^1 \chi_{i,j}(q,p)\cdot(\omega_{i,j}(q)\cdot\theta_i + (1-\omega_{i,j}(q))\cdot\theta_j)\,\mathrm{d}q = \frac{1}{\tilde{f}(p)} \sum\nolimits_{(i,j):i\le j} \int_0^1 \chi_{i,j}(q,p)\cdot q\,\mathrm{d}q\end{aligned}$$

Thus, as function $x \mapsto |p-x|^t$ is convex for $t \ge 1$, the inequality (4.2) holds by Jensen's inequality. Consequently, we have

$$\left(\mathbf{ECE}_t\left[\tilde{f}\right]\right)^t = \int_0^1 \sum\nolimits_{i\in[n]} \lambda_i \tilde{f}_i(p)\cdot|p-\kappa(p)|^t\,\mathrm{d}p \le \int_0^1\int_0^1 \sum\nolimits_{(i,j):i\le j} \chi_{i,j}(q,p)\cdot|p-q|^t\,\mathrm{d}q\mathrm{d}p \le \varepsilon^t$$

where the last inequality holds due to BiE-Calibration-Feasibility (in Definition 4.2) of bi-event post-processing plan χ. This finishes the proof of Proposition 4.5. ∎

So far, we have extended the two-step framework from the outcome-independent setting, which uses the event-independent post-processing plan, to the general setting, which uses the bi-event post-processing plan. While the event-independent post-processing plan may not be sufficient to generate the optimal predictor in the general setting, the key question now is whether the more general bi-event post-processing plan is sufficient. The following proposition answers this question affirmatively—considering the bi-event post-processing plan is indeed sufficient to generate the optimal predictor.

Proposition 4.6. *For every persuasive calibration instance, the principal's expected utility under the optimal (ε, ℓ_t)-calibrated predictor $\tilde{f}^*$ is equal to the optimal objective value of the following linear program with variables $\{\chi_{i,j}(q,p)\}_{i,j\in[n]:i\le j,q,p\in[0,1]}$:*

(LP-TwoStep$^+$)

$$\max_{\chi\ge 0} \quad \sum_{(i,j):i\le j} \int_0^1\int_0^1 \chi_{i,j}(q,p)\cdot U_{i,j}(q,p)\,\mathrm{d}p\mathrm{d}q \qquad s.t.$$

$$(\textit{BiE-Calibration-Feasibility}) \quad \left(\int_0^1\int_0^1 \sum\nolimits_{(i,j):i\le j} \chi_{i,j}(q,p)\cdot|q-p|^t\,\mathrm{d}p\mathrm{d}q\right)^{\frac{1}{t}} \le \varepsilon,$$

$$(\textit{BiE-Supply-Feasibility})$$
$$\int_0^1\int_0^1 \left(\sum_{k\in[i:n]} \chi_{i,k}(q,p)\cdot\omega_{i,k}(q) + \sum_{k\in[i-1]} \chi_{k,i}(q,p)\cdot(1-\omega_{k,i}(q))\right)\mathrm{d}p\mathrm{d}q = \lambda_i \qquad i\in[n]$$

where, with a slight abuse of notation, we expand the definition of indirect utility function as: for every pair of events (i,j) with $i\le j$ and predictions $q,p\in[0,1]$,

$$U_{i,j}(q,p) \triangleq \omega_{i,j}(q)\cdot U_i(p) + (1-\omega_{i,j}(q))\cdot U_j(p)$$

Moreover, every optimal solution of LP-TwoStep$^+$ can be converted into an optimal ε-calibrated predictor $\tilde{f}^*$ using Eqn. (4.1).

We remark that LP-TwoStep$^+$ is an infinite-dimensional linear program. In the next subsection, we introduce an instance-dependent discretization scheme for this program, providing a provable guarantee on the discretization error. The proof of Proposition 4.6 utilizes a technical result developed in Feng et al. (2022) that decomposes a random variable to a set of random variables where each of them is binary-supported and has the same mean. We defer its proof to the full version of the paper (Feng and Tang, 2025a).

4.3 Instance-Dependent Discretization and Rounding Scheme In this section, we introduce the second key technical ingredient of Algorithm 4.1: a discretization scheme for the infinite-dimensional linear program LP-TwoStep$^+$. Our goal is to develop this scheme with a provable guarantee on the discretization error. To achieve this, an ideal discretization scheme and its induced discretized linear program (LP) should to satisfy the following two properties:

1. The discretized LP has a feasible solution.

2. Given an optimal solution to original LP LP-TwoStep^{+}, we can identify a feasible solution in the discretized LP whose objective value is close and the ECE is weakly smaller.

Both properties introduce challenges in applying common discretization schemes, such as uniform discretization, to our problem. (See Example 4.7 where the discretized LP becomes infeasible under uniform discretization.) While similar challenges arising from the first property has been observed in the algorithmic information design literature (e.g., Agrawal et al., 2023), the second challenge is, to the best of our knowledge, unique due to the ECE budget constraint. In particular, suppose the discretized LP under uniform discretization with precision δ happens to be feasible. A natural rounding approach would then convert a feasible solution into a discretized feasible solution by rounding each $\chi_{i,j}(q,p)$ to $\chi_{i,j}(q',p')$, where (q',p') is the closest discretized point to (q,p). However, after this naive rounding, the ECE may increase by $\Theta(\delta)$, which could be significantly larger than the ECE budget ε. To ensure that the ECE increment remains controlled under this rounding procedure, we would need to set the discretization precision as $\delta \leftarrow \Theta(\varepsilon)$, i.e., on the same order as the ECE budget ε. However, with such a discretization precision, the size of the discretized set and the corresponding discretized LP would have a polynomial dependence on ε, which prevents us from obtaining the FPTAS stated in Theorem 4.1.

Example 4.7 (The failure of uniform discretization). *Consider a three-event setting with $\{\theta_i\}_{i\in[3]}$ and a zero ECE budget $\varepsilon = 0$, i.e., predictor is feasible if and only if it is perfectly calibrated.*

Now, consider a uniform-grid-only discretization scheme $\mathcal{S}_\delta = \{\delta, 2\delta, \ldots\} \cap [0,1]$. Suppose the event distribution λ assigns negligible probabilities to the first and third events, and the true expected outcome for the second event, θ_2, is not in the the discretized set $\mathcal{S}_\delta$ (i.e., $\theta_2 \notin \mathcal{S}_\delta$), then it can be verified with Lemma 3.14 that it is impossible to construct any perfectly calibrated predictor whose support lies entirely on the grid points in $\mathcal{S}_\delta$. Therefore, no feasible predictor exists if we restrict its support to discretized set $\mathcal{S}_\delta$.

We overcome the two aforementioned challenges by introducing an instance-dependent two-layer discretization scheme (Definition 4.8). Loosely speaking, to ensure that the discretized LP remains feasible, our scheme first constructs a non-uniform $\Theta(\delta)$-net that includes discretized points corresponding to the expected outcomes $\Theta = \{\theta_1, \ldots, \theta_n\}$ and the discontinuity points $\mathcal{Z}$ of the principal's indirect utility function. Additionally, for every point in this $\Theta(\delta)$-net, we introduce a finer "multiplicative $\Theta(\delta\varepsilon^t)$-net" within a tiny neighborhood of length $\Theta(\varepsilon^t)$ (see formal construction in Definition 4.8). By analyzing the optimal solution of the original program LP-TwoStep^{+}, we argue that locally refining the grid within these small neighborhoods is sufficient to approximately preserve the optimal objective value while keeping ECE variations under control (Lemmas 4.10 and 4.11). With this approach, the size of the discretized set remains polynomial in the problem input and the discretization parameter δ (Claim 4.9), enabling the construction of an FPTAS.

Definition 4.8 (Instance-dependent two-layer discretization scheme). *Given any discretization parameter $\delta \in (0, 1/3)$, define* discretized set $\mathcal{S}_\delta$ *of the space $[0,1]$ for both true expected outcomes and the predictions as follows:*

$$\mathcal{S}_\delta \triangleq \underbrace{(\{0,\delta,2\delta,\ldots\} \cap [0,1]) \cup \Theta \cup \mathcal{Z}}_{\text{“first global layer”}} \cup \underbrace{\left\{ z \pm \left(\delta_0 \cdot (1+\delta)^s\right)^{1/t} \right\}_{z\in\mathcal{Z}\cup\Theta, s\in[0:S]}}_{\text{“second local layer”}}$$

where $\mathcal{Z}$ is the set of all discontinuity points of the sender's indirect utility function $\{U_i\}_{i\in[n]}$, $\delta_0 \triangleq \varepsilon^t \cdot \delta$, and $S \triangleq \lceil \frac{2}{\ln(1+\delta)} \ln \frac{1}{\delta} \rceil$.

Although our discretization scheme above depends on the ECE budget ε and the norm exponent parameter t, the number of discretized points in the set only depends on the problem instance and discretization parameter δ.

Claim 4.9. *Given any discretization parameter $\delta \in (0, 1/3)$, the size of the discretized set $\mathcal{S}_\delta$ defined in Definition 4.8 satisfies $|\mathcal{S}_\delta| = O(\frac{1}{\delta} + \frac{m+n}{\ln(1+\delta)} \ln \frac{1}{\delta})$, where $n = |\Theta|, m = |\mathcal{A}|$ are the number of events and the number of agent's actions, respectively.*

Proof. Since the agent has m actions, the discontinuity points of the sender's indirect utility function $\{U_i\}_{i\in[n]}$ is at most $m+1$. Thus, the size of the discretized set $\mathcal{S}_\delta$ in the statement holds by construction. ∎

Given the discretization scheme, we are ready to present the discretized LP LP-DiscTwoStep$^+_\delta$ and its discretization error guarantee.

Lemma 4.10 (Discretized LP and discretization error guarantee). *Given any discretization parameter $\delta \in (0, 1/3)$, consider the following discretized version of LP-TwoStep$^+$:*

(LP-DiscTwoStep$^+_\delta$)

$$\begin{aligned}
\max_{\chi \geq 0} \quad & \sum_{(i,j):i\leq j} \sum_{p,q\in\mathcal{S}_\delta} \chi_{i,j}(q,p)\cdot U_{i,j}(q,p) \qquad\qquad s.t. \\
& \sum_{p,q\in\mathcal{S}_\delta} \sum_{(i,j):i\leq j} \chi_{i,j}(q,p)\cdot |q-p|^t \leq \varepsilon^t, \\
& \sum_{p,q\in\mathcal{S}_\delta} \left(\sum_{k\in[i:n]} \chi_{i,k}(q,p)\cdot\omega_{i,k}(q) + \sum_{k\in[i-1]} \chi_{k,i}(q,p)\cdot(1-\omega_{k,i}(q)) \right) \mathrm{d}p\mathrm{d}q = \lambda_i, \quad i\in[n]
\end{aligned}$$

where discretized set $\mathcal{S}_\delta$ is defined in Definition 4.8. Then, the following three properties hold:

(i) LP-DiscTwoStep$^+_\delta$ is a feasible linear program with $\mathsf{poly}(n, m, 1/\delta)$ size, where $n = |\Theta|, m = |\mathcal{A}|$ are the number of events and the number of agent's actions, respectively.

(ii) Every feasible solution of LP-DiscTwoStep$^+_\delta$ is also a feasible solution of LP-TwoStep$^+$.

(iii) The optimal objective value of LP-DiscTwoStep$^+_\delta$ is a $(1-3\delta)$-approximation to the optimal objective value of LP-TwoStep$^+$, i.e.,

$$\mathbf{OBJ}\left[\text{LP-DiscTwoStep}^+_\delta\right] \geq (1-3\delta)\cdot\mathbf{OBJ}\left[\text{LP-TwoStep}^+\right] .$$

In Lemma 4.10, the second property follows directly from the construction of LP-DiscTwoStep$^+_\delta$. In the remainder of this section, we develop a rounding argument to establish feasibility and bound the discretization error of LP-DiscTwoStep$^+_\delta$.

A rounding scheme for analyzing LP-DiscTwoStep$^+_\delta$. To show Lemma 4.10, we argue that (1) an optimal solution to LP-TwoStep$^+$ can, via a rounding scheme (see Algorithm 4.2), be converted to a feasible solution to LP-DiscTwoStep$^+_\delta$; (2) the converted solution has a close objective value to the original solution.

Before introducing the rounding scheme, we first observe that it is without loss to consider an optimal solution χ to LP-TwoStep$^+$ such that any generated prediction lies in the set $\mathcal{S}_\delta$.

Lemma 4.11. *In LP-TwoStep$^+$, there exists an optimal solution χ such that for any $\chi_{i,j}(q,p) > 0$, it satisfies that $p \in \mathcal{Z} \cup \Theta$.*

The intuition behind this result is that the principal's indirect utility function, U_i, is piecewise constant. Consequently, if the true expected outcome q and the prediction p fall into different constant segments, it is always possible—without loss—to modify the predictor so that it "miscalibrates" true expected outcome q to a discontinuity point. Conversely, if they lie within the same constant segment, we can distribute true expected outcome q across a pair of events or discontinuity points and then generate perfectly calibrated predictions for these points.

Proof of Lemma 4.11. Recall that $\mathcal{Z} \triangleq \{z_k\}_{k\in[m]}$ denotes the set of all discontinuity points in the principal's indirect utility function. Fix any feasible solution χ to LP-TwoStep$^+$, for any pair of (q,p) with $\chi_{i,j}(q,p) > 0$ and $p \notin \mathcal{Z}$, we consider two possible cases:

Suppose $p \in (z_k, z_{k+1})$ and $q \in (z_l, z_{l+1})$ where $k < l$ (or $k > l$). Consider a new solution that, instead of miscalibrating q to p, it miscalibrates q to z_{k+1} (or z_k). This adjustment weakly decreases the calibration error while weakly improving the expected utility of the principal, due to the upper semi-continuity of the principal's indirect utility function (see Footnote 11).

Suppose both $p, q \in (z_k, z_{k+1})$. In this case, it is without loss to assume $q = p$, since otherwise the solution incurs additional calibration error without any improvement in expected utility of the principal. Suppose further that $q \in (\theta_i, \theta_j)$. Define $q_L \triangleq \theta_i \vee z_k$ and let $q_R \triangleq \theta_j \wedge z_{k+1}$. Now consider a modified solution that, instead of contracting the true expected outcomes θ_i, θ_j to q, it contracts them to q_L, q_R. Notice that this is always feasible as $\theta_i \le q_L \le q_R \le \theta_j$. Now this new solution generates perfectly calibrated predictions from these true expected outcomes q_L, q_R. Since the calibration error does not increase and the expected utility of the principal weakly improves (again due to upper semi-continuity – see Footnote 11), the modified solution is at least as good as the original.

Combing the two cases, we complete the proof of Lemma 4.11. ∎

Algorithm 4.2 Rounding scheme of χ for the discretization error analysis

1: **Input:** a feasible solution χ to LP-TwoStep^{+};
2: **Input:** $\delta > 0, \mathcal{S}_\delta$ and $\delta_0 = \delta \cdot \varepsilon^t$, and let $\delta^\dagger \leftarrow 1 - \frac{1}{1+2\delta}$
3: **Output:** a feasible solution $\chi^\dagger$ to LP-DiscTwoStep$^{+}_{\delta}$.
4: **Initialize:** $\chi^\dagger_{i,j}(q,p) \leftarrow 0$ for all $i, j \in [n]$ and all $q, p \in \mathcal{S}_\delta$.
5: **for** every $(q,p) \in [0,1]^2$ with $\chi_{i,j}(q,p) > 0$ **do** /* Without loss consider $p \le q$ */
6: /* Without loss consider $p \in \mathcal{Z} \cup \Theta$. */
7: /* Below we use $\delta^\dagger$ fraction of $\chi_{i,j}(q,p)$ to induce perfectly calibrated predictions. */
8: $\chi^\dagger_{i,j}(\theta_i,\theta_i) \leftarrow \chi^\dagger_{i,j}(\theta_i,\theta_i) + \delta^\dagger \cdot \chi_{i,j}(q,p) \cdot \omega_{i,j}(q)$.
9: $\chi^\dagger_{i,j}(\theta_j,\theta_j) \leftarrow \chi^\dagger_{i,j}(\theta_j,\theta_j) + \delta^\dagger \cdot \chi_{i,j}(q,p) \cdot (1 - \omega_{i,j}(q))$.
10: /* Below we round q to be in $\mathcal{S}_\delta$. */
11: **if** $q - p < \delta_0^{1/t}$ **then** /* Case I */
12: Set $q_L \leftarrow \theta_i \vee p$, $q_R \leftarrow (p + \delta_0^{1/t}) \wedge \theta_j$. /* When $i = j$, we must have $q_L = \theta_i = q_R$ */
13: Round q to be q_L, q_R according to Eqn. (4.3). /* We set $\frac{0}{0} = 1$ when $i = j$ happens */
14: **else if** $\delta_0^{1/t} \le q - p \le (\delta_0 \cdot (1+\delta)^{S-1})^{1/t}$ **then** /* Case II */
15: Set $q_L \leftarrow \theta_i \vee p$, and set q_R as defined in Claim 4.12.
16: Round q to be q_L, q_R according to Eqn. (4.3).
17: **else** /* Case III */
18: $\chi^\dagger_{i,j}(\theta_i,\theta_i) \leftarrow \chi^\dagger_{i,j}(\theta_i,\theta_i) + (1-\delta^\dagger) \cdot \chi_{i,j}(q,p) \cdot \omega_{i,j}(q)$;
19: $\chi^\dagger_{i,j}(\theta_j,\theta_j) \leftarrow \chi^\dagger_{i,j}(\theta_j,\theta_j) + (1-\delta^\dagger) \cdot \chi_{i,j}(q,p) \cdot (1 - \omega_{i,j}(q))$.
20: **end if**
21: **end for**

By Lemma 4.11, we only need to round the true expected outcome q so that it lies in the set $\mathcal{S}_\delta$. At a high-level, our rounding scheme proceeds in two steps:

Step 1 (Lines 8-9). In this step, for each $\chi_{i,j}(q,p) > 0$, we use $\delta^\dagger$ fraction of it to generate perfectly calibrated predictions. By doing this, we save some calibration error so that we can have enough buffer for the potential increase of calibration error in the later rounding steps.

Step 2 (Lines 11-20). In this step, given a pair (q,p) with $\chi_{i,j}(q,p) > 0$, we round the true expected outcome q, using the remaining fraction $1 - \delta^\dagger$ of $\chi_{i,j}(q,p)$, to two grid points $q_L, q_R \in \mathcal{S}_\delta$. We then consider three possible cases, depending on the distance $|q-p|$.

In **Case I and Case II** where the distance $|q-p|$ is not too large, we round q to q_L, q_R and then miscalibrate these points to the prediction p as follows:

$$\begin{aligned} \chi^\dagger_{i,j}(q_L,p) &\leftarrow \chi^\dagger_{i,j}(q_L,p) + (1-\delta^\dagger)\chi_{i,j}(q,p)\frac{q_R - q}{q_R - q_L} \\ \chi^\dagger_{i,j}(q_R,p) &\leftarrow \chi^\dagger_{i,j}(q_R,p) + (1-\delta^\dagger)\chi_{i,j}(q,p)\left(1 - \frac{q_R - q}{q_R - q_L}\right) \end{aligned} \tag{4.3}$$

where the ratio $\frac{q_R - q}{q_R - q_L}$ ensures that, after rounding, the constructed $\chi^\dagger$ satisfies the second constraint in LP-DiscTwoStep$^{+}_{\delta}$. We carefully choose the grid points $q_L, q_R \in \mathcal{S}_\delta$ so that, after rounding, the calibration error does not increase significantly. Note that for these two cases, the expected utility of the principal remains unchanged after rounding.

In **Case III** where the distance $|q-p|$ is very large, we use the remaining fraction $1-\delta^\dagger$ of $\chi_{i,j}(q,p)$ to generate perfectly calibrated predictions. In this case, there is no increase in calibration error. Furthermore, by the condition defining this case, we can also argue that any loss of the expected utility of the principal is upper-bounded by a factor of $\delta(1+\delta)$.

With the rounding scheme (Algorithm 4.2), we are now ready to prove Lemma 4.10.

Proof of Lemma 4.10. The property (ii) follows directly from the construction of LP-DiscTwoStep$^+_\delta$. We next prove the statements (i) and (iii).

For the simplicity of the presentation, we define $\overline{\mathbf{ECE}}_t[\chi] \triangleq \int_0^1\int_0^1 \sum_{i,j:i\le j}\chi_{i,j}(q,p)\cdot|q-p|^t\,\mathrm{d}q\mathrm{d}p$ and $\overline{\mathbf{ECE}}_t[\chi\mid(q,p)] \triangleq \sum_{i,j:i\le j}\chi_{i,j}(q,p)\cdot|q-p|^t$. By definition, we have $\overline{\mathbf{ECE}}_t\left[\chi^\dagger\right] = \sum_{q,p}\overline{\mathbf{ECE}}_t[\chi\mid(q,p)]$.

Let us fix a optimal solution χ to LP-TwoStep$^+$ that satisfying Lemma 4.11. Below we show that the rounding scheme detailed in Algorithm 4.2 which takes the solution χ as input can output a feasible solution $\chi^\dagger$ to LP-DiscTwoStep$^+_\delta$, namely, we have $\overline{\mathbf{ECE}}_t\left[\chi^\dagger\right]\le\varepsilon^t$, and its support are restricted to be in the set $\mathcal{S}_\delta$; and meanwhile, the output $\chi^\dagger$ has $(1-3\delta)$-approximate expected utility of the principal to the solution χ.

We analyze the incurred calibration error of the constructed $\chi^\dagger$ step by step:

- In Step 1, the constructed $\chi^\dagger$ has 0 calibration error as it always generates perfectly calibrated predictions.
- In Step 2 with Case I, the calibration error incurred in $\chi^\dagger$ can be bounded as follows:

$$
\begin{aligned}
&\sum\nolimits_{(q_L,q_R,p)\text{ in Case I}}\overline{\mathbf{ECE}}_t\left[\chi^\dagger\mid(q_L,p)\right]+\overline{\mathbf{ECE}}_t\left[\chi^\dagger\mid(q_R,p)\right]\\
&=\sum\nolimits_{(q_L,q_R,p)\text{ in Case I}}\sum\nolimits_{(i,j):i\le j}\left(\chi^\dagger_{i,j}(q_L,p)\cdot|q_L-p|^t+\chi^\dagger_{i,j}(q_R,p)\cdot|q_R-p|^t\right)\\
&\text{(as } q_L\le q_R\le p+\delta_0^{1/t}\text{)}\\
&\le(1-\delta^\dagger)\cdot\delta_0\sum\nolimits_{(q,p)\text{ in Case I}}\sum\nolimits_{(i,j):i\le j}\left(\chi_{i,j}(q,p)\frac{q_R-q}{q_R-q_L}+\chi_{i,j}(q,p)\left(1-\frac{q_R-q}{q_R-q_L}\right)\right)\\
&\le(1-\delta^\dagger)\cdot\delta_0\ .
\end{aligned}
$$

- In Step 2 with Case II, we know that, by construction, we have

$$
\begin{aligned}
\overline{\mathbf{ECE}}_t\left[\chi^\dagger\mid(q_L,p)\right]&=\sum\nolimits_{(i,j):i\le j}\chi^\dagger_{i,j}(q_L,p)\cdot|q_L-p|^t\\
\text{(as } p\le q_L\le q\text{)}\qquad&\le\sum\nolimits_{(i,j):i\le j}(1-\delta^\dagger)\chi_{i,j}(q,p)\frac{q_R-q}{q_R-q_L}\cdot|q-p|^t\\
\overline{\mathbf{ECE}}_t\left[\chi^\dagger\mid(q_R,p)\right]&=\sum\nolimits_{(i,j):i\le j}\chi^\dagger_{i,j}(q_R,p)\cdot|q_R-p|^t\\
&=(1-\delta^\dagger)\sum\nolimits_{(i,j):i\le j}\chi_{i,j}(q,p)\left(1-\frac{q_R-q}{q_R-q_L}\right)\cdot|q_R-p|^t\\
\text{(as } p\le q_R\le q_R'\text{)}\qquad&\le(1-\delta^\dagger)\sum\nolimits_{(i,j):i\le j}\chi_{i,j}(q,p)\left(1-\frac{q_R-q}{q_R-q_L}\right)\cdot|q_R'-p|^t\\
&\le(1+\delta)(1-\delta^\dagger)\sum\nolimits_{(i,j):i\le j}\chi_{i,j}(q,p)\left(1-\frac{q_R-q}{q_R-q_L}\right)\cdot|q-p|^t
\end{aligned}
$$

  Thus, the total calibration error of $\chi^\dagger$ contributed from this case can be upper bounded by

$$
\begin{aligned}
&\sum\nolimits_{(q_L,q_R,p)\text{ in Case II}}\left(\overline{\mathbf{ECE}}_t\left[\chi^\dagger\mid(q_L,p)\right]+\overline{\mathbf{ECE}}_t\left[\chi^\dagger\mid(q_R,p)\right]\right)\\
&\le(1+\delta)(1-\delta^\dagger)\cdot\sum\nolimits_{(q,p)\text{ in Case II}}\sum\nolimits_{(i,j):i\le j}\left(\chi_{i,j}(q,p)\cdot|q-p|^t+\chi_{i,j}(q,p)\cdot|q-p|^t\right)\\
&=(1+\delta)(1-\delta^\dagger)\cdot\sum\nolimits_{(q,p)\text{ in Case II}}\cdot\overline{\mathbf{ECE}}_t[\chi\mid(q,p)]\\
&\text{(by definition } \delta_0=\delta\cdot\varepsilon^t,\ \overline{\mathbf{ECE}}_t[\chi]\le\varepsilon^t\text{)}\\
&\le(1+\delta)(1-\delta^\dagger)\overline{\mathbf{ECE}}_t[\chi]\le(1+\delta)(1-\delta^\dagger)\cdot\frac{\delta_0}{\delta}\ .
\end{aligned}
$$

- In Step 2 with Case III, notice that in this case, the constructed $\chi^\dagger$ always generates perfectly calibrated predictions. Moreover, the total probability mass for this case cannot be large, as

$$
\begin{aligned}
\overline{\mathbf{ECE}}_t[\chi] &\geq \sum\nolimits_{(q,p)\text{ in Case III}} \overline{\mathbf{ECE}}_t[\chi \mid (q,p)] \\
&= \sum\nolimits_{(q,p)\text{ in Case III}} \sum\nolimits_{(i,j):i\leq j} \chi_{i,j}(q,p)|q-p|^t \\
\text{(by condition of Case III)} \qquad &\geq \sum\nolimits_{(q,p)\text{ in Case III}} \sum\nolimits_{(i,j):i\leq j} \chi_{i,j}(q,p)\delta_0(1+\delta)^{S-1} \\
\text{(by definition of } S\text{)} \qquad &\geq \sum\nolimits_{(q,p)\text{ in Case III}} \sum\nolimits_{(i,j):i\leq j} \chi_{i,j}(q,p) \cdot \frac{\overline{\mathbf{ECE}}_t[\chi]}{\delta(1+\delta)} .
\end{aligned}
$$

  Thus, in Case III, we must have $\sum_{(q,p)\text{ in Case III}} \sum_{(i,j):i\leq j} \chi_{i,j}(q,p) \leq \delta$. Consequently, the loss of the expected utility of the principal in constructed $\chi^\dagger$ at this case is upper bounded by $\delta(1+\delta)\cdot \mathbf{OBJ}\left[\text{LP-TwoStep}^+ \mid \chi\right]$.

Putting all pieces together, we have the following guarantee of the ECE of constructed χ:

$$
\begin{aligned}
\overline{\mathbf{ECE}}_t[\chi] - \overline{\mathbf{ECE}}_t\left[\chi^\dagger\right] &\geq \overline{\mathbf{ECE}}_t[\chi] - \delta_0(1-\delta^\dagger) - (1+\delta)(1-\delta^\dagger)\cdot\frac{\delta_0}{\delta} \\
\text{(by the choice of } \delta^\dagger\text{)} \qquad &= \delta_0\left(\frac{1}{\delta} - (1-\delta^\dagger) - \frac{(1+\delta)(1-\delta^\dagger)}{\delta}\right) = 0
\end{aligned}
$$

In addition, we have the following guarantee of the expected utility of the principal of constructed χ:

$$
\begin{aligned}
&\mathbf{OBJ}\left[\text{LP-DiscTwoStep}^+_\delta \mid \chi^\dagger\right] \\
&\geq \mathbf{OBJ}\left[\text{LP-TwoStep}^+ \mid \chi\right] - \underbrace{\delta^\dagger\mathbf{OBJ}\left[\text{LP-TwoStep}^+ \mid \chi\right]}_{\text{payoff loss in Step 1}} - \underbrace{\delta(1+\delta)\mathbf{OBJ}\left[\text{LP-TwoStep}^+ \mid \chi\right]}_{\text{payoff loss in Step 2 with Case III}} \\
&\text{(by the choice of } \delta^\dagger\text{)} \\
&\geq (1-3\delta)\mathbf{OBJ}\left[\text{LP-TwoStep}^+ \mid \chi\right]
\end{aligned}
$$

We thus finish the proof of Lemma 4.10 as desired. □

Finally, we include Claim 4.12 (and its proof) mentioned in our rounding scheme (Algorithm 4.2).

CLAIM 4.12. *In Step 2 with Case II of Algorithm 4.2 (line 15), we define $q'_R \leftarrow p + (q-p)(1+\delta)^{1/t}$. When $q'_R \geq \theta_j$, set $q_R \leftarrow \theta_j$. When $q'_R < \theta_j$, there must exist $s \in [S]$ such that $p + (\delta_0(1+\delta)^s)^{1/t} \in [q, q'_R]$, set $q_R \leftarrow p + (\delta_0(1+\delta)^s)^{1/t}$.*

Proof of Claim 4.12. To show that when $q'_R < \theta_j$, there must exist $s \in [S]$ such that $p + (\delta_0(1+\delta)^s)^{1/t} \in [q, q'_R]$, it suffices to show that there exists $s \in [S]$:

$$
q \leq p + (\delta_0(1+\delta)^s)^{1/t} \leq q'_R \;\Leftrightarrow\; \frac{\ln\frac{(q-p)^t}{\delta_0}}{\ln(1+\delta)} \leq s \leq \frac{\ln\frac{(q'_R-p)^t}{\delta_0}}{\ln(1+\delta)} = 1 + \frac{\ln\frac{(q-p)^t}{\delta_0}}{\ln(1+\delta)} .
$$

By the condition $q - p \geq \delta_0^{1/t}$, we observe that we must have $\frac{\ln\frac{(q-p)^t}{\delta_0}}{\ln(1+\delta)} \geq 0$. We next argue that we must have $q'_R \leq p + (\delta_0(1+\delta)^S)^{1/t}$. To see this, we observe

$$
\begin{aligned}
p + (\delta_0(1+\delta)^S)^{1/t} - q'_R &= (\delta_0(1+\delta)^S)^{1/t} - (q-p)(1+\delta)^{1/t} \\
&\geq (\delta_0(1+\delta)^S)^{1/t} - (\delta_0(1+\delta)^{S-1})^{1/t}(1+\delta)^{1/t} = 0
\end{aligned}
$$

where the last equality holds due to the condition of Case II. Thus, there must exist $s \in [S]$ such that $p + (\delta_0(1+\delta)^s)^{1/t} \in [q, q'_R]$. This finishes the proof of Claim 4.12. □

4.4 Analysis of Algorithm 4.1

We are ready to analyze Algorithm 4.1 and prove Theorem 4.1.

Proof of Theorem 4.1. In Algorithm 4.1, it first solves LP-DiscTwoStep$^+_\delta$, which is a linear program with size $\mathsf{poly}(1/\delta, n, m)$, and then constructs the predictor based on the optimal solution of LP-DiscTwoStep$^+_\delta$. Both steps require $\mathsf{poly}(1/\delta, n, m)$ running time and thus the algorithm is polynomial-time. Finally, the approximation of the algorithm follows Proposition 4.6 and Lemma 4.10. This finishes the proof of Theorem 4.1. □

5 Polytime Algorithm for ℓ_1-Norm ECE and ℓ_∞-Norm ECE In this section, we show that for the most standard ECE metrics, the K_1 ECE, and also the K_∞ ECE, there exists a polynomial-time algorithm that can compute the optimal (ε, ℓ_t)-calibrated predictor with $t \in \{1, \infty\}$.

THEOREM 5.1 (Polynomial-time algorithm for ℓ_1-norm and ℓ_∞-norm ECE). *For every persuasive calibration instance with ℓ_1-norm ECE (resp. ℓ_∞-norm) ECE, there exists a linear programming (see P-ActRec) based algorithm (Algorithm 5.1) that computes an optimal (ε, ℓ_1)-calibrated (resp. $(\varepsilon, \ell_\infty)$-calibrated) predictor. The running time of the algorithm is $\mathsf{poly}(n, m)$, where $n = |\Theta|, m = |\mathcal{A}|$ are the number of events and the number of agent's actions, respectively.*

For ℓ_1-norm or ℓ_∞-norm ECE, it is not difficult to express the (ε, ℓ_t)-calibration as a linear constraint on $\tilde{f} = \{\tilde{f}_i(p)\}_{i\in[n],p\in[0,1]}$. However, to obtain an optimal algorithm with polynomial running time, we also need to restrict the continuous space of predictions to a discrete polynomial-size set. To bypass this challenge, we build on an idea from the algorithmic information design literature.

As an example, in the classic Bayesian persuasion problem (Kamenica and Gentzkow, 2011), the *revelation principle* assures that it is sufficient to consider signaling schemes that recommend an action. Hence, instead of searching over a (possibly) continuous signal space, it suffices to construct a signaling scheme with a signal space equal to the action space, which is polynomial-sized. However, such an approach seems difficult to apply to our model, as the design space in our model is restricted to predictions rather than arbitrary signals, and the agent in our model follows a much simpler behavior—i.e., always trusting the prediction—rather than engaging in strategic reasoning. Therefore, it is unclear a priori whether a revelation principle can be established.

To prove Theorem 5.1, we show that the above idea from the algorithmic information design literature indeed applies. To achieve this, we first introduce a new variant of the Bayesian persuasion problem, in which the receiver (agent in our model): (1) has a utility function that depends linearly on the payoff-relevant state, and (2) instead of updating the belief in a fully Bayesian manner, exhibits a signal-dependent bias when updating the belief.

In Subsection 5.1, we formally define this variant of the Bayesian persuasion problem, establish the revelation principle, and provide a solution for this problem, which will be used to solve our persuasive calibration problem and is of independent interest. In Subsection 5.2, we show the instance equivalence between our problem and this Bayesian persuasion variant (see Theorem 5.6), provide Algorithm 5.1.

5.1 (Bayesian) Persuasion with Signal-Dependent Bias Bayesian persuasion (Kamenica and Gentzkow, 2011) is a classic and important model that has been studied extensively in the literature. In this section, we introduce a new variant—*(Bayesian) persuasion with signal-dependent bias*—and provide a characterization of its optimal solution, which serves as a key ingredient for developing Algorithm 5.1 and proving Theorem 5.1. Additionally, given the significance of the Bayesian persuasion and information design literature, we believe that our new variant and its results may be of independent interest. Below, we first revisit the classic Bayesian persuasion model and then introduce the new variant with signal-dependent bias.

A Bayesian persuasion instance is a game for two players, a sender and a receiver, and it is specified by a tuple $(\Theta^{\mathsf{BP}}, \lambda^{\mathsf{BP}}, u^{\mathsf{BP}}, v^{\mathsf{BP}}, \mathcal{A}^{\mathsf{BP}})$: $\Theta^{\mathsf{BP}} \subseteq [0, 1]$ denotes the payoff-relevant state space; $\lambda^{\mathsf{BP}} \in \Delta(\Theta^{\mathsf{BP}})$ represents the common prior for both players, and $\lambda^{\mathsf{BP}}_\theta$ denotes the prior probability of state θ; $u^{\mathsf{BP}}(\cdot, \cdot) : \mathcal{A}^{\mathsf{BP}} \times \Theta \to \mathbb{R}$ (resp. $v^{\mathsf{BP}}(\cdot, \cdot) : \mathcal{A}^{\mathsf{BP}} \times \Theta \to \mathbb{R}$) is the sender's (resp. receiver's) utility function that depends on the receiver's action and the realized state. The sender observes the realized state, while the receiver cannot.

The sender's goal is to design a signaling scheme, given by conditional distributions $(\pi_\theta(\cdot))_{\theta\in\Theta^{\mathsf{BP}}}$ where each $\pi_\theta(\cdot) \in \Delta(\Sigma)$ specifies a distribution over a measurable signal space Σ, to maximize her expected payoff. In classic Bayesian persuasion, upon observing a realized signal, the receiver updates his posterior belief about the underlying state in a *Bayesian* manner and takes an action that maximizes his expected utility (subject to updated posterior belief).

(Bayesian) persuasion with signal-dependent bias. In our considered variant, the receiver has a utility function $v(a, \cdot)$ that linearly depends on the state. Thus, only the mean of the belief matters for his optimal action. In addition, when updating his belief, the receiver exhibits a signal-dependent bias defined as follows:

DEFINITION 5.2 (Signal-dependent bias in belief updates). *Upon observing a signal realization $\sigma \sim \pi$, the* *receiver computes the following mean of the updated belief, denoted by $p(\sigma)$:*

$$p(\sigma) \triangleq \frac{b(\sigma) + \sum_{\theta \in \Theta^{\mathsf{BP}}} \lambda_\theta^{\mathsf{BP}} \pi_\theta(\sigma) \cdot \theta}{\sum_{\theta \in \Theta^{\mathsf{BP}}} \lambda_\theta^{\mathsf{BP}} \pi_\theta(\sigma)} \in [0,1] \tag{5.1}$$

where $b(\sigma) \in \mathbb{R}$ represents receiver's bias when updating his belief (relative to an exact Bayesian update). The receiver then takes an action maximizing his expected utility according to the mean $p(\sigma)$ of his belief,

$$a^*(\sigma) \triangleq \arg\max_{a \in \mathcal{A}^{\mathsf{BP}}} v^{\mathsf{BP}}(a, p(\sigma)) \ . \tag{5.2}$$

Intuitively, $b(\sigma)$ captures the receiver's (possibly) irrational bias when updating their belief. As a sanity check, when $b(\sigma) \equiv 0$ for all $\sigma \in \Sigma$, the receiver's belief update in Eqn. (5.1) recovers the standard Bayesian posterior mean. In our variant, such biases are chosen to be favorable to the sender but must satisfy the following *bounded rationality* constraint:

$$\left(\mathbb{E}_{\sigma \sim \pi}\left[\left|\frac{b(\sigma)}{\sum_{\theta \in \Theta^{\mathsf{BP}}} \lambda_\theta^{\mathsf{BP}} \cdot \pi_\theta(\sigma)}\right|^t\right]\right)^{1/t} \leq \varepsilon \ . \tag{5.3}$$

The above constraint essentially regulates that the aggregated biases with being weighted by the subjective belief are upper bounded by a budget $\varepsilon \geq 0$. As the biases are chosen to be favorable to the sender, the sender can also optimize over the bias assignments subject to the above constraint. We refer to the pair of signaling scheme π and a bias assignment b as the sender's strategy.

Our formulation above, in particular the favorable bias assignments, shares similarity to the literature on ε-incentive compatible (IC) mechanism design (see, e.g., Balcan et al., 2005; Hartline and Lucier, 2010, 2015; Cai et al., 2012, 2021; Gonczarowski and Weinberg, 2021; Balseiro et al., 2024) where the ties are also usually chosen to be favorable to the mechanism designer.[23] Notably, even when selling a single item to a single buyer, the optimal ε-IC mechanism admits no simple structural characterization (Balseiro et al., 2024). In contrast, by using the equivalence between the persuasive calibration and Bayesian persuasion with signal-dependent bias (Theorem 5.6), the structural characterizations developed for the former calibration problem in Section 3 can be translated into structural results for the latter persuasion problem as well.

We also remark that many previous works have explored relaxing the Bayesian rationality assumptions in the classic Bayesian persuasion problem, either from the action-taking perspective (see, e.g., Feng et al., 2024; Yang and Zhang, 2024) or from the belief-updating perspective (see, e.g., Alonso and Câmara, 2016; Galperti, 2019; De Clippel and Zhang, 2022). Our new variant, Bayesian persuasion with signal-dependent bias, falls into the latter category. In general, the revelation principle – a commonly-used principle to analyze the classic Bayesian persuasion problem – need not hold when the receiver is subject to biases in probabilistic inference, where the main reason is that biased beliefs typically break linearity, it is no longer true that the convex combinations of implementable outcomes stay implementable in the same way (De Clippel and Zhang, 2022). However, as we show below, the revelation principle still holds in our considered persuasion problem.

Definition 5.3 (Direct and IC strategy). *We say a sender's strategy (π, b), i.e., a pair of signaling scheme π and bias assignment b, is* direct *and* incentive compatible (IC)*, if it satisfies (i) its signal space equals to the receiver's action space; (ii) upon receiving a recommended action, it is indeed receiver's best response (defined as in Eqn. (5.2)) to follow this action recommendation.*

Lemma 5.4 (Revelation principle). *For the persuasion problem with signal-dependent bias, the revelation principle holds: for every sender's strategy (π, b), there exists a direct and IC strategy $(\pi^\dagger, b^\dagger)$ that achieves the same outcome, i.e., conditional on the same state realization, the distributions of actions selected by the receiver are the same under both the original strategy (π, b) and the new direct and IC strategy $(\pi^\dagger, b^\dagger)$.*

Therefore, there always exists a direct and IC sender's strategy that maximizes the sender's expected utility.

[23] There is also another line of research that studies robust mechanism design when agents select the ε-best response, which minimizes the designer's utility (e.g., Gan et al., 2023; Chen and Lin, 2023; Yang and Zhang, 2024).

The proof of Lemma 5.4 is rather straightforward, and thus we defer it to the full version of the paper (Feng and Tang, 2025a). With Lemma 5.4, we can formulate the sender's problem as finding an optimal direct and IC strategy, which can be further characterized by a maximization program with polynomial numbers of constraints and variables.

PROPOSITION 5.5 (Optimization program characterization). *For the persuasion problem with signal-dependent bias, the optimal sender's strategy (π^*, b^*) and her optimal expected utility is the solution of the following maximization program:*

(P-ActRec)

$$\max_{\pi\geq 0;b} \quad \sum_{\theta\in\Theta^{\mathsf{BP}}}\sum_{a\in\mathcal{A}} \lambda_\theta^{\mathsf{BP}}\pi_\theta(a)\cdot u^{\mathsf{BP}}(a,\theta) \qquad \textit{s.t.}$$

$$\left(\sum_{a\in\mathcal{A}^{\mathsf{BP}}}\sum_{\theta\in\Theta^{\mathsf{BP}}} \lambda_\theta^{\mathsf{BP}}\pi_\theta(a)\cdot\left|\frac{b(a)}{\sum_{\theta\in\Theta^{\mathsf{BP}}}\lambda_\theta^{\mathsf{BP}}\pi_\theta(a)}\right|^t\right)^{\frac{1}{t}} \leq \varepsilon$$

$$v^{\mathsf{BP}}\left(a, b(a)+\sum_{\theta\in\Theta^{\mathsf{BP}}}\lambda_\theta^{\mathsf{BP}}\pi_\theta(a)\cdot\theta\right) \geq v^{\mathsf{BP}}\left(a', b(a)+\sum_{\theta\in\Theta^{\mathsf{BP}}}\lambda_\theta^{\mathsf{BP}}\pi_\theta(a)\cdot\theta\right) \qquad a, a'\in\mathcal{A}^{\mathsf{BP}}$$

$$\sum_{a\in\mathcal{A}^{\mathsf{BP}}}\pi_\theta(a) = 1 \qquad \theta\in\Theta^{\mathsf{BP}}$$

$$b(a)+\sum_{\theta\in\Theta^{\mathsf{BP}}}\lambda_\theta^{\mathsf{BP}}\pi_\theta(a)\cdot\theta \geq 0 \qquad a\in\mathcal{A}^{\mathsf{BP}}$$

$$b(a)+\sum_{\theta\in\Theta^{\mathsf{BP}}}\lambda_\theta^{\mathsf{BP}}\pi_\theta(a)\cdot\theta \leq \sum_{\theta\in\Theta^{\mathsf{BP}}}\lambda_\theta^{\mathsf{BP}}\pi_\theta(a) \qquad a\in\mathcal{A}^{\mathsf{BP}}$$

which has $\mathsf{poly}(|\Theta^{\mathsf{BP}}|, |\mathcal{A}^{\mathsf{BP}}|)$ *variables,* $\{\pi_\theta(a), b(a)\}_{\theta\in\Theta^{\mathsf{BP}}, a\in\mathcal{A}^{\mathsf{BP}}}$*, and* $\mathsf{poly}(|\Theta^{\mathsf{BP}}|, |\mathcal{A}^{\mathsf{BP}}|)$ *constraints. Both the objective function and all constraints, except for the first constraint, are linear.*[24]

In the special case of $t = 1$ or $t = \infty$, the first constraint in P-ActRec is equivalent to

$$\textit{when } t=1: \quad \sum_{a\in\mathcal{A}}|b(a)| \leq \varepsilon \;;$$
$$\textit{when } t=\infty: \quad |b(a)| \leq \varepsilon \;, \quad a\in\mathcal{A}^{\mathsf{BP}} \;.$$

which is essentially a linear constraint. Therefore, P-ActRec becomes a linear program for $t = 1$ or $t = \infty$.

As mentioned in Proposition 5.5, P-ActRec is a polynomial-size linear program when $t \in \{1, \infty\}$ and can thus be solved in polynomial time. For other values of $t \in (1, \infty)$, the first constraint in P-ActRec becomes non-linear. We leave it as an interesting future direction to explore whether P-ActRec can be solved in polynomial time for general t, e.g., by constructing a time-efficient separation oracle.

Proof of Proposition 5.5. Invoking Lemma 5.4, there exists an optimal strategy (π^*, b^*) for the sender that is direct and IC. We first verify that it is a feasible solution of P-ActRec with the same objective value. Since strategy (π^*, b^*) is direct and IC, the expected utility of the sender is equal to the objective value of (π^*, b^*) in P-ActRec. The first constraint is satisfied since strategy (π^*, b^*) satisfies the bounded rationality constraint defined in Eqn. (5.3). The second constraint is satisfied since strategy (π^*, b^*) is IC and utility function $v^{\mathsf{BP}}(a, \theta)$ is linear over the states $\theta \in \Theta^{\mathsf{BP}}$. The third, fourth and fifth constraints are satisfied due to the feasibility of strategy (π^*, b^*).

Finally, following the same argument, every feasible solution of P-ActRec represents a sender's strategy that is well-defined, feasible, direct and IC and has the expected utility of the sender equal to the objective of the solution. Therefore, we finish the proof of Proposition 5.5 as desired. ∎

5.2 Instances Equivalence and Analysis of Algorithm 5.1 In this section, we formally establish the equivalence between the persuasive calibration problem and the persuasion problem with signal-dependent bias (Subsection 5.1). We then use the equivalence to develop Algorithm 5.1 and prove Theorem 5.1.

At first glance, the two problems might seem different, as the agent in the persuasive calibration model always trusts the prediction, while the agent in the persuasion model performs a biased Bayesian update to form

[24] Recall that function $v^{\mathsf{BP}}(a, \theta)$ is linear over the states $\theta \in \Theta^{\mathsf{BP}}$.

their posterior belief. However, the ECE budget constraint in the former and the bounded rationality constraint (Eqn. (5.3)) in the latter model enable us to establish an equivalence between the two problems. The proof of Theorem 5.6 is at the end of this section.

THEOREM 5.6 (Instance equivalence). *Fix any instance of the persuasive calibration problem, specified by a tuple* $(\Theta, \lambda, u, v, \mathcal{A}, (\varepsilon, \ell_t))$*. Consider the following instance* $(\Theta, \lambda, u^{\mathsf{BP}}, v^{\mathsf{BP}}, \mathcal{A}, (\varepsilon, \ell_t))$ *of the (Bayesian) persuasion problem with signal-dependent bias:*

(i) the sender's/receiver's utility functions are given as:

$$u^{\mathsf{BP}}(a, \theta_i) \leftarrow (1-\theta_i)\cdot u_i(a, y=0) + \theta_i \cdot u_i(a, y=1), \quad a \in \mathcal{A}, i \in [n] \ ;$$
$$v^{\mathsf{BP}}(a, \theta_i) \leftarrow (1-\theta_i)\cdot v(a, y=0) + \theta_i \cdot v(a, y=1), \quad a \in \mathcal{A}, i \in [n] \ ;$$

(ii) the sender's strategy follows the bounded rationality constraint defined in Eqn. (5.3).

Then an optimal (ε, ℓ_t)*-calibrated predictor* $\tilde{f}^*$ *in the persuasive calibration instance can be converted as an optimal sender's strategy* (π^*, b^*) *in the above defined instance of the (Bayesian) persuasion problem with signal-dependent bias, and vice versa. Specifically, given an optimal sender's strategy* (π^*, b^*) *that is direct and IC, an optimal* (ε, ℓ_t)*-calibrated predictor* $\tilde{f}^*$ *can be constructed as*

$$\tilde{f}_i^*(p) \leftarrow \sum\nolimits_{a\in\mathcal{A}} \pi^*_{\theta_i}(a)\cdot \mathbf{1}\left\{\frac{\sum_{i\in[n]}\lambda_i\pi^*_{\theta_i}(a)\cdot\theta_i + b^*(a)}{\sum_{i\in[n]}\lambda_i\pi^*_{\theta_i}(a)} = p\right\}, \quad i\in[n], \quad p\in[0,1] \ . \tag{5.4}$$

Given an optimal (ε, ℓ_t)*-calibrated predictor* $\tilde{f}^*$*, an optimal sender's strategy* (π^*, b^*) *that is direct and IC can be constructed as*

$$\pi^*_{\theta_i}(a) \leftarrow \sum\nolimits_{p\in[0,1]:a^*(p)=a} \tilde{f}_i^*(p), \qquad a\in\mathcal{A}, i\in[n]$$
$$b^*(a) \leftarrow \sum\nolimits_{p\in[0,1]:a^*(p)=a}\sum\nolimits_{i\in[n]} \lambda_i \tilde{f}_i^*(p)(p-\theta_i), \qquad a\in\mathcal{A}$$

where $a^*(p)$ *is defined in Eqn.* (2.2).

We remark that in the corresponding Bayesian persuasion instance constructed in Theorem 5.6, the receiver's utility $v(a,\theta)$ is a linear function over the states $\theta \in \Theta^{\mathsf{BP}}$, while the sender's utility could be general. Combining Theorem 5.6 and Proposition 5.5, we are ready to provide our Algorithm 5.1. The proofs of Theorem 5.1 and Theorem 5.6 are deferred to the full version of the paper (Feng and Tang, 2025a).

Algorithm 5.1 LP-based polynomial-time algorithm for solving optimal predictor with $t \in \{1, \infty\}$

1: **Input:** persuasive calibration instance $(\Theta, \lambda, u, v, \mathcal{A}, (\varepsilon, \ell_t))$ with $t \in \{1, \infty\}$.
2: Solve linear program P-ActRec using the persuasion instance defined in Theorem 5.6 and let (π^*, b^*) be its solution.
3: Use the optimal solution (π^*, b^*) to construct the predictor $\tilde{f}^*$ based on Eqn. (5.4).
4: **Output:** predictor $\tilde{f}^*$.

Acknowledgments. We thank the reviewers of SODA'26 for the helpful comments and suggestions.

References

Shipra Agrawal, Yiding Feng, and Wei Tang. Dynamic pricing and learning with bayesian persuasion. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems*, 2023.

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

Shuang Ao, Stefan Rueger, and Advaith Siddharthan. Two sides of miscalibration: identifying over and under-confidence prediction for network calibration. In *Uncertainty in artificial intelligence*, pages 77–87. PMLR, 2023.

Itai Arieli, Yakov Babichenko, Rann Smorodinsky, and Takuro Yamashita. Optimal persuasion via bi-pooling. *Theoretical Economics*, 18(1):15–36, 2023.

Eshwar Ram Arunachaleswaran, Natalie Collina, Aaron Roth, and Mirah Shi. An elementary predictor obtaining distance to calibration. In *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1366–1370. SIAM, 2025.

Yakov Babichenko, Inbal Talgam-Cohen, Haifeng Xu, and Konstantin Zabarnyi. Algorithmic cheap talk. In *Proceedings of the 25th ACM Conference on Economics and Computation*, pages 5–6, 2024.

Maria-Florina Balcan, Avrim Blum, Jason D. Hartline, and Yishay Mansour. Mechanism design via machine learning. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005), 23-25 October 2005, Pittsburgh, PA, USA, Proceedings*, pages 605–614. IEEE Computer Society, 2005. doi: 10.1109/SFCS.2005.50. URL https://doi.org/10.1109/SFCS.2005.50.

Santiago R Balseiro, Omar Besbes, and Francisco Castro. Mechanism design under approximate incentive compatibility. *Operations Research*, 72(1):355–372, 2024.

David Blackwell. Equivalent comparisons of experiments. *The annals of mathematical statistics*, pages 265–272, 1953.

Jarosław Błasiok, Parikshit Gopalan, Lunjia Hu, and Preetum Nakkiran. A unifying theory of distance from calibration. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 1727–1740, 2023.

Jaroslaw Blasiok, Parikshit Gopalan, Lunjia Hu, and Preetum Nakkiran. When does optimizing a proper loss yield calibration? *Advances in Neural Information Processing Systems*, 36:72071–72095, 2023.

Yang Cai, Constantinos Daskalakis, and S. Matthew Weinberg. Optimal multi-dimensional mechanism design: Reducing revenue to welfare maximization. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 130–139. IEEE Computer Society, 2012. doi: 10.1109/FOCS.2012.88. URL https://doi.org/10.1109/FOCS.2012.88.

Yang Cai, Argyris Oikonomou, Grigoris Velegkas, and Mingfei Zhao. An efficient ε-bic to bic transformation and its application to black-box reduction in revenue maximization. In *Proceedings of the 2021 acm-siam symposium on discrete algorithms (soda)*, pages 1337–1356. SIAM, 2021.

Modibo K. Camara, Jason D. Hartline, and Aleck C. Johnsen. Mechanisms for a no-regret agent: Beyond the common prior. In Sandy Irani, editor, *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 259–270. IEEE, 2020. doi: 10.1109/FOCS46700.2020.00033. URL https://doi.org/10.1109/FOCS46700.2020.00033.

Sílvia Casacuberta, Cynthia Dwork, and Salil Vadhan. Complexity-theoretic implications of multicalibration. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1071–1082, 2024.

Yiling Chen and Tao Lin. Persuading a behavioral agent: Approximately best responding and learning. *CoRR*, abs/2302.03719, 2023.

Natalie Collina, Aaron Roth, and Han Shao. Efficient prior-free mechanisms for no-regret agents. In *Proceedings of the 25th ACM Conference on Economics and Computation*, pages 511–541, 2024.

Roberto Corrao and Yifan Dai. Mediated communication with transparent motives. In Kevin Leyton-Brown, Jason D. Hartline, and Larry Samuelson, editors, *Proceedings of the 24th ACM Conference on Economics and Computation, EC 2023, London, United Kingdom, July 9-12, 2023*, page 489. ACM, 2023. doi: 10.1145/3580507.3597808. URL https://doi.org/10.1145/3580507.3597808.

Yuval Dagan, Constantinos Daskalakis, Maxwell Fishelson, Noah Golowich, Robert Kleinberg, and Princewill Okoroafor. Breaking the $T^{2/3}$ barrier for sequential calibration. *arXiv preprint arXiv:2406.13668*, 2024.

A Philip Dawid. The well-calibrated bayesian. *Journal of the American statistical Association*, 77(379):605–610, 1982.

Geoffroy De Clippel and Xu Zhang. Non-bayesian persuasion. *Journal of Political Economy*, 130(10):2594–2642, 2022.

Shaddin Dughmi and Haifeng Xu. Algorithmic persuasion with no externalities. In *Proceedings of the 18th ACM Conference on Economics and Computation*, pages 351–368, 2017.

Piotr Dworczak and Giorgio Martini. The simple economics of optimal persuasion. *Journal of Political Economy*, 127(5):1993–2048, 2019.

Yewen Fan, Nian Si, and Kun Zhang. Calibration matters: Tackling maximization bias in large-scale advertising recommendation systems. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL https://openreview.net/forum?id=wzlWiO_WY4.

Yiding Feng and Wei Tang. Persuasive calibration. *arXiv preprint arXiv:2504.03211*, 2025a.

Yiding Feng and Wei Tang. Confusion matrix design for downstream decision-making. In *16th Innovations in Theoretical Computer Science Conference (ITCS 2025)*, pages 49–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2025b.

Yiding Feng, Wei Tang, and Haifeng Xu. Online bayesian recommendation with no regret. In David M. Pennock, Ilya Segal, and Sven Seuken, editors, *EC '22: The 23rd ACM Conference on Economics and Computation, Boulder, CO, USA, July 11 - 15, 2022*, pages 818–819. ACM, 2022. doi: 10.1145/3490486.3538327. URL https://doi.org/10.1145/3490486.3538327.

Yiding Feng, Chien-Ju Ho, and Wei Tang. Rationality-robust information design: Bayesian persuasion under quantal response. In David P. Woodruff, editor, *Proceedings of the 2024 ACM-SIAM Symposium on Discrete Algorithms, SODA 2024, Alexandria, VA, USA, January 7-10, 2024*, pages 501–546. SIAM, 2024. doi: 10.1137/1.9781611977912.19. URL https://doi.org/10.1137/1.9781611977912.19.

Dean P Foster and Rakesh V Vohra. Asymptotic calibration. *Biometrika*, 85(2):379–390, 1998.

Simone Galperti. Persuasion: The art of changing worldviews. *American Economic Review*, 109(3):996–1031, 2019.

Jiarui Gan, Minbiao Han, Jibang Wu, and Haifeng Xu. Robust stackelberg equilibria. In Kevin Leyton-Brown, Jason D. Hartline, and Larry Samuelson, editors, *Proceedings of the 24th ACM Conference on Economics and Computation, EC 2023, London, United Kingdom, July 9-12, 2023*, page 735. ACM, 2023. doi: 10.1145/3580507.3597680. URL https://doi.org/10.1145/3580507.3597680.

Sumegha Garg, Christopher Jung, Omer Reingold, and Aaron Roth. Oracle efficient online multicalibration and omniprediction. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2725–2792. SIAM, 2024.

Matthew Gentzkow and Emir Kamenica. A rothschild-stiglitz approach to bayesian persuasion. *American Economic Review*, 106(5):597–601, 2016.

Tilmann Gneiting and Matthias Katzfuss. Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1(1):125–151, 2014.

Tilmann Gneiting, Fadoua Balabdaoui, and Adrian E Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 69(2):243–268, 2007.

Yannai A Gonczarowski and S Matthew Weinberg. The sample complexity of up-to-ε multi-dimensional revenue maximization. *Journal of the ACM (JACM)*, 68(3):1–28, 2021.

Parikshit Gopalan, Adam Tauman Kalai, Omer Reingold, Vatsal Sharan, and Udi Wieder. Omnipredictors. In *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*, pages 79–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2022.

Parikshit Gopalan, Lunjia Hu, Michael P Kim, Omer Reingold, and Udi Wieder. Loss minimization through the lens of outcome indistinguishability. In *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*, pages 60–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2023.

Parikshit Gopalan, Lunjia Hu, and Guy N. Rothblum. On computationally efficient multi-class calibration. In *Proceedings of Thirty Seventh Conference on Learning Theory*, volume 247, pages 1983–2026. PMLR, 2024.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In *International conference on machine learning*, pages 1321–1330. PMLR, 2017.

Yingni Guo and Eran Shmaya. Costly miscalibration. *Theoretical Economics*, 16(2):477–506, 2021.

Nika Haghtalab, Chara Podimata, and Kunhe Yang. Calibrated stackelberg games: Learning optimal commitments against calibrated agents. *Advances in Neural Information Processing Systems*, 36:61645–61677, 2023.

Jason D. Hartline and Brendan Lucier. Bayesian algorithmic mechanism design. In Moshe Dror and Greys Sosic, editors, *Proceedings of the Behavioral and Quantitative Game Theory - Conference on Future Directions, BQGT '10, Newport Beach, California, USA, May 14-16, 2010*, page 19:1. ACM, 2010. doi: 10.1145/1807406.1807425. URL https://doi.org/10.1145/1807406.1807425.

Jason D Hartline and Brendan Lucier. Non-optimal mechanism design. *American Economic Review*, 105(10): 3102–3124, 2015.

Ursula Hébert-Johnson, Michael Kim, Omer Reingold, and Guy Rothblum. Multicalibration: Calibration for the (computationally-identifiable) masses. In *International Conference on Machine Learning*, pages 1939–1948. PMLR, 2018.

Lunjia Hu and Yifan Wu. Predict to minimize swap regret for all payoff-bounded tasks. In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 244–263. IEEE, 2024.

Atulya Jain and Vianney Perchet. Calibrated forecasting and persuasion. In *The Twenty-Fifth ACM Conference on Economics and Computation (EC'24)*, 2024.

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

Michael P Kim, Amirata Ghorbani, and James Zou. Multiaccuracy: Black-box post-processing for fairness in classification. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pages 247–254, 2019.

Bobby Kleinberg, Renato Paes Leme, Jon Schneider, and Yifeng Teng. U-calibration: Forecasting for an unknown agent. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 5143–5145. PMLR, 2023.

Volodymyr Kuleshov, Nathan Fenner, and Stefano Ermon. Accurate uncertainties for deep learning using calibrated regression. In *International conference on machine learning*, pages 2796–2804. PMLR, 2018.

Elliot Lipnowski and Doron Ravid. Cheap talk with transparent motives. *Econometrica*, 88(4):1631–1660, 2020.

Georgy Noarov, Ramya Ramalingam, Aaron Roth, and Stephan Xie. High-dimensional prediction for sequential decision making. *arXiv preprint arXiv:2310.17651*, 2023.

Mingda Qiao and Gregory Valiant. Stronger calibration lower bounds via sidestepping. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 456–466, 2021.

Mingda Qiao and Eric Zhao. Truthfulness of decision-theoretic calibration measures. *arXiv preprint arXiv:2503.02384*, 2025.

Mingda Qiao and Letian Zheng. On the distance from calibration in sequential prediction. In *The Thirty Seventh Annual Conference on Learning Theory*, pages 4307–4357. PMLR, 2024.

Rahul Rahaman et al. Uncertainty quantification and deep ensembles. *Advances in neural information processing systems*, 34:20063–20075, 2021.

Roopesh Ranjan and Tilmann Gneiting. Combining probability forecasts. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 72(1):71–91, 2010.

Aaron Roth and Mirah Shi. Forecasting for swap regret for all downstream agents. In *Proceedings of the 25th ACM Conference on Economics and Computation*, pages 466–488, 2024.

Michael Rothschild and Joseph E Stiglitz. Increasing risk: I. a definition. In *Uncertainty in economics*, pages 99–121. Elsevier, 1978.

Jingwu Tang, Jiahao Zhang, Fei Fang, and Zhiwei Steven Wu. Persuasive prediction via decision calibration. *arXiv preprint arXiv:2505.16141*, 2025.

Deng-Bao Wang, Lei Feng, and Min-Ling Zhang. Rethinking calibration of deep neural networks: Do not be afraid of overconfidence. *Advances in Neural Information Processing Systems*, 34:11809–11820, 2021.

Kunhe Yang and Hanrui Zhang. Computational aspects of bayesian persuasion under approximate best response. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

Shengjia Zhao, Michael Kim, Roshni Sahoo, Tengyu Ma, and Stefano Ermon. Calibrating predictions to decisions: A novel approach to multi-class calibration. *Advances in Neural Information Processing Systems*, 34:22313–22324, 2021.