Flexible Information Acquisition in the Kyle Model*

S. Viswanathan
Duke University & NBER

Hao Xing
Boston University

July 2026

Abstract

We study an information acquisition problem in which an informed trader acquires costly information prior to trading in the Kyle equilibrium, where the cost of information acquisition is represented by an entropy cost. For any prior distribution of the asset payoff, continuous signals are optimal. Any continuously distributed signal, together with an associated logit-type posterior distribution of the payoff, yields the same ex-ante value for the informed trader, the same distribution of posterior expected payoff, and the same unconditional distribution of the informed trader's trading strategy. Hence, a normally distributed signal can be chosen without loss of generality. Further, we show that when the information acquisition cost increases or the volatility of noise trades decreases, the variance of the posterior expected payoff declines, the profit potential from trading diminishes: meanwhile the posterior expected payoff increasingly resembles a normal distribution, and the information leakage cost from trading decreases.

Keywords: Kyle model, information acquisition, mutual information, entropy regularized optimal transport

*First draft March 2026. We are grateful to Umut Cetin, Pete Kyle, and conference and seminar participants at Boston University, NUS Quantitative Finance workshop for very helpful comments. Emails: S. Viswanathan (viswanat@duke.edu) and Hao Xing (haoxing@bu.edu).

1 Introduction

Information acquisition in the pioneering Kyle (1985) model is typically restricted to a decision problem in which the informed trader acquires costly normally distributed signals about the asset payoff. Specifically, the informed trader begins with a normal prior on the asset payoff and pays a cost to obtain a normal posterior, after which the Kyle trading game is played; see, for example, Kyle (1984), Admati and Pfleiderer (1988), Back and Pedersen (1998), Vives (2014), and Khorrami and Thereze (2025).

Rather than restricting attention to normal signals about a normally distributed payoff, we study information acquisition in the Kyle model when the informed trader has an arbitrary prior distribution on the asset payoff — continuous or discrete, and possibly non-normal — and can acquire signals with arbitrary marginal and joint distributions with the payoff. We follow Sims (1998, 2003) to model the information acquisition cost using a mutual information cost. The informed trader designs the optimal signal structure by balancing the expected profit from trading using the observed signal and the ex-ante information acquisition cost. For this flexible information acquisition problem in the Kyle model, we provide sharp results on the optimal signal structure and the welfare of the informed trader.

Our analysis is motivated by a recent development on characterizing Kyle's equilibrium from an optimal transport point of view. In an important paper, Back (1992) extends the continuous-time Kyle (1985) model by considering non-normal informed trader information. Recently, using an optimal transport approach, Back et al. (2021) provided a representation for the ex-ante expected profit of the informed trader in Back (1992)'s extension of the Kyle model. We use this representation as a starting point for our analysis. This representation suggests a tradeoff between the profit potential from trading and the information leakage cost associated with trading.

On the one hand, the informed trader desires a precise signal so that the posterior distribution of the asset payoff has a high variance, which indicates a greater informational advantage and a higher profit potential in the subsequent Kyle trading game. However, the informed trader must also consider the inferences made by the market makers in this trading game. More aggressive trading leads to more precise inferences by market makers, increasing the cost of information leakage for the informed trader. Consider the case where the asset payoff follows a normal distribution. Extreme payoffs in the tails have small probabilities, but they generate large trading profits for the informed trader. Overweighting tail scenarios

in the signal structure relative to the normal distribution and trading aggressively in those scenarios increases the informed trader's expected profit. However, such aggressive trading may be detected by the market makers, who adjust the price accordingly and thereby reduce the informed trader's profit. In the representation of Back et al. (2021), the profit potential from trading is measured by the variance of the asset payoff, and the information leakage cost is represented by the Wasserstein distance between the payoff distribution and the normal distribution of noise trades. In the absence of information acquisition cost, the former effect always dominates — the informed trader maximizes the variance of the asset payoff using her precise information about the payoff.

We consider a model in which the informed trader acquires a costly signal about the asset payoff prior to trading. Three competing forces shape the informed trader's information acquisition decision. First, the informed trader desires a precise signal that generates a high variance of the posterior conditional expectation of the asset payoff after signal observation, thereby increasing the profit potential from trading. Second, market makers' inference imposes a cost on the informed trader's information strategy. This cost is captured by the Wasserstein distance between the distribution of the posterior expected payoff and the distribution of noise trades. Third, the informed trader incurs an entropy cost that increases with the informational content of the signal. These three forces interact in shaping the informed trader's information acquisition decision.

We first consider the case in which the asset payoff has a normal prior distribution. In the rational inattention literature, following Sims (1998, 2003), it is well known that a decision maker with an entropy cost of information and a normal prior acquires information that leads to a normal posterior. However, in our problem, we must also consider the impact from the Wasserstein distance. To this end, the Gelbrich (1990) bound implies that the distribution that minimizes the Wasserstein distance to the normal noise trade distribution is also normal. Therefore, the normal signal is optimal for the flexible information acquisition decision given a normal prior.

With a normal prior, we show that the optimal signal precision decreases with the information acquisition cost parameter, but increases with the standard deviation of the noise trades. When the information acquisition becomes more costly, the informed trader chooses a less precise signal, which reduces the variance of the posterior expected payoff and the profit potential from trading. When the noise trades become more volatile, the informed trader can better hide her informed orders among noise trades, hence the concern about information leakage is reduced. Consequently, the informed trader increases the

signal precision to take advantage of the market environment.

We then consider general payoff distributions. We show that the informed trader, no matter whether her prior distribution on the payoff is discrete or continuous, always chooses a continuous signal, which generates a continuously-distributed posterior expected payoff. At first glance, this result seems surprising for a discretely distributed prior. In the discrete choice literature, the recommendation lemma (see e.g. (Matějka and McKay, 2015, Lemma 1)) implies that it is suboptimal to acquire distinct costly signals which lead to the same action, hence the same utility for the decision maker. However, because the noise trades in the Kyle model have a continuous distribution, price depends continuously on the informed trader's expectation of the asset payoff after observing the signal. Therefore, there is an infinite number of potential actions for the informed trader, each corresponding to a conditional expected payoff given signal realizations. Moreover, the informed trader's value is convex in the partition of the signal space, incentivizing the informed trader to use a continuously distributed signal. As a result, even when the prior distribution of the asset payoff is a two-point distribution on $\{L, H\}$, the optimal posterior expected payoff has a continuous distribution on the closed interval $[L, H]$.

Regardless of the prior distribution on the asset payoff, we show that any continuous signal, which implies to a higher conditional expected payoff from a higher observed signal, leads to the same ex-ante value for the informed trader. Moreover, the distribution of the posterior expected payoff is invariant of the signal choice. This, perhaps surprising, indeterminacy in the signal choice resembles the independence of trading strategy with respect to the payoff distribution in Back (1992). For any continuously distributed payoff with a strictly increasing cumulative distribution function (CDF), the informed agent trades toward a target which shares the same distribution as the noise trades.[1] This trading target is constructed by composing the quantile function of the payoff distribution with the CDF of the noise trade. This composition generates the same normal distribution for any continuously distributed payoff. This is also the reason why the informed trader's optimal value and the distribution of her conditional expected payoff do not depend on the signal choice in our setting. This composition function is the unique monotone optimal transport map from the conditional expected payoff distribution to the noise trade distribution; see (Galichon, 2016, Section 4) for optimal transport between 1-dimensional distributions.

We show that both the optimality of continuous signals and the equivalence between

[1] Bagnoli, Viswanathan and Holden (2001) show that with one informed agent, the distribution of the informed agent's order must be identical to the noise trader distribution to to have a linear equilibrium.

them are robust to the choice of information acquisition cost. Both results still hold when we consider more flexible uniformly posterior-separable (UPS) cost functions, introduced by Caplin and Dean (2013) and Caplin, Dean and Leahy (2022).

With non-normal priors (either discrete or continuous), the interaction between the profit potential (variance of the conditional expected payoff), the information leakage cost measured by the Wasserstein distance, and the information acquisition cost determine the shape of the posterior expected payoff. Specifically, consider a prior distribution that exhibits heavier tails than a normal distribution. When the information acquisition cost parameter is small, the informed trader acquires a precise signal that generates a high variance of the conditional expected payoff with excess kurtosis compared to a normal distribution. However, as the information acquisition cost parameter increases, the variance of the conditional expected payoff decreases, and the information leakage cost becomes increasingly important, so that the conditional expected payoff increasingly resembles a normal distribution, with its tails becoming lighter and the kurtosis converging to 3 (the kurtosis for normal distributions). As the information acquisition cost parameter approaches infinity, the information content of the signal vanishes, and the conditional expected payoff converges to the prior expected payoff.

The impact of the noise trade volatility is exactly the opposite. As noise trades become more volatile, the cost of information leakage decreases, and the informed trader acquires a more precise signal that generates a heavy-tailed distribution of the conditional expected payoff. When the payoff has a two-state symmetric prior, the ratio between the information acquisition cost parameter and the standard deviation of the noise trade becomes a sufficient statistic for the conditional expected payoff distribution. A general implication of these results is that higher information acquisition costs or lower noise trade volatility lead to posterior distributions of the expected payoff that are more similar to normal distributions.

Our paper presents a new flexible approach to information acquisition in the continuous time trading model of Kyle (1985) and the extensions by Back (1992) and Back et al. (2021). In particular, we allow for general priors and signal structures, either of them can be discrete, continuous, and non-normal. We provide a complete characterization of the optimal information acquisition decision. We show that the posterior expected payoff has a continuous distribution, even for discretely distributed priors. Further, there is an indeterminacy in the signal choice, though the distributions of the conditional expected payoff and the informed trader's trading strategy are both determined. Therefore, without loss of generality, the informed trader can pick a normal signal. This rationalizes the normal

signal choice in Kyle (1984) and Back and Pedersen (1998). We show that the flexible entropy cost mediates between the two tensions — the desire for high trading profits from a precise signal, which generates a high variance of the conditional expected payoff, and the desire to reduce the information leakage cost induced by market makers' inference. Such a trade-off is robust for different types of prior payoff distributions.

This paper sits at the intersection of three different literatures. First, our paper contributes to the information acquisition decision in the Kyle (1985) model. Two notable recent contributions are Banerjee and Breon-Drish (2020, 2022) and Khorrami and Thereze (2025). Banerjee and Breon-Drish (2020) consider a dynamic information acquisition problem where the informed trader can either acquire a smooth flow of information or a lumpy information. They show that the informed trader increases the signal precision when the noise trading volatility is high. They assume that the payoff has a normal prior, the trader can acquire a normal signal about the payoff, and the information acquisition cost is a convex function of the signal precision. In our static information acquisition setting, general distributions of payoff and signal are considered. When the information acquisition incurs a fixed cost, Banerjee and Breon-Drish (2020) show the equilibrium may fail to exist. Banerjee and Breon-Drish (2022) model the costly research activity of the informed trader, who controls the research intensity and the likelihood of research success. Conditional on successful research, the informed trader observes a normal signal about payoff. Khorrami and Thereze (2025) consider a convex cost of acquiring normal signals in the generalized Kyle model studied by Lambert, Ostrovsky and Panov (2018). They tie the information efficiency in the large market limit with the marginal cost of signal precision at the prior. When the information structure is exogenous, Back and Baruch (2004) consider a Bernoulli-distributed payoff for the informed trader's signal. Caldentey and Stacchetti (2010) examine informed trading with a random announcement time. Collin-Dufresne and Fos (2016) study informed trading with stochastic volatility in noise trading.

In the context of noisy rational expectations model (Grossman and Stiglitz (1980) and Hellwig (1980)), signals on non-Gaussian payoffs have been considered by Breon-Drish (2015), Chabakauri, Yuan and Zachariadis (2022), and Detemple, Rindisbacher and Robertson (2020). However, this literature does not consider the optimal information acquisition for non-Gaussian payoffs. A notable exception is Han (2021), who study a dynamic information acquisition problem in the setting of Hellwig (1980) and Breon-Drish (2015). Han (2021) highlights the dynamic interaction of learning and price informativeness.

Second, our paper is related to the rational inattention literature; see Maćkowiak,

Matějka and Wiederholt (2023) for an extensive review. Matějka and McKay (2015) provide a foundation for the multinomial logit model in a discrete choice setting. Different from Matějka and McKay (2015), we show that the optimal signal has a continuous distribution even if the prior distribution on the state is discrete, due to the continuously distributed noise trades in the subsequent trading game. Our approach is similar to the posterior approach by Caplin and Dean (2013), Caplin, Dean and Leahy (2019, 2022). However, the informed trader's problem does not lie in the posterior separable class studied in this literature. In contrast to this literature, where the net utility of the decision maker depends only on the posterior distribution of states, our net utility depends on the entire marginal of the signal that is chosen.

Flexible information acquisition is also studied in security design problems. Hébert (2018) examines a moral hazard setting where the seller can modify the asset payoff distribution subject to a convex cost. When the cost is the Kullback-Leibler divergence, the optimal security is debt. Yang (2020) studies a setting where the buyer can acquire costly information about the asset. The optimality of debt is also illustrated there.

Finally, our paper connects to the entropy regularized optimal transport and the static Schrodinger bridge problem considered in physics and machine learning. Given the marginal distributions of payoff and signal, we solve the Lagrangian multipliers for Bayes plausibility via the Sinkhorn problem; see Sinkhorn (1964) and Rüschendorf and Thomsen (1993) for problems with discrete and continuous marginals. Peyré and Cuturi (2019) provide an extensive review on the computational aspect of the Sinkhorn problem and its connection with the entropy regularized optimal transport. Optimal transport has been applied to matching problems and other topics in economics (see Galichon (2016)). The entropy regularized optimal transport has also been used to study the a regularized version of the Bayesian persuasion problem recently by Justiniano et al. (2025). The entropy regularization helps to design an efficient numerical algorithm. Convergence of the algorithm is examined when the entropy regularization vanishes. Malamud and Schrimpf (2021) transform a multi-dimensional Bayesian persuasion problem into an optimal transport problem. The target of the transport map is the optimal information manifold in Malamud and Schrimpf (2021) and the optimal marginal distribution for signal in our model. Both these targets are determined endogenously. Çetin (2025) considers a continuous-time Kyle model with legal risk, where the insider pays a quadratic legal penalty for her trading rate. The expected legal penalty can be expressed as the relative entropy between the law of total demand the the law of noise trader demand, which represents how far the insider's trade pushes the

total demand away from the noise trader demand.

The rest of the paper is organized as follows. Section 2 introduces the model setup and an equivalent formulation of the problem. Section 3 considers the normally-distributed payoff. We start by assuming a normally-distributed signal and investigate the optimal signal precision. Then we prove that the normally-distributed signal is optimal for a normal prior. Section 4 presents the results for discretely-distributed payoffs. Section 4.1 presents the optimal signal structure among continuous signals. The optimality of continuous signals is presented in Section 4.2, followed by several examples in Section 4.3. Section 5 studies the continuously distributed payoff. Section 6 extends our results to the UPS cost functions. Section 7 concludes.

2 Model setup

2.1 Information acquisition and equilibrium

The asset payoff is represented by a random variable $\tilde{v}$, whose value is realized and made public at the time T. At time 0, the insider acquires a signal $\tilde{s}$. The state space for $\tilde{v}$ and $\tilde{s}$ is represented by the sets $\mathcal{V}$ and $\mathcal{S}$, respectively. We will specify these sets in different settings later. The marginal distribution of $\tilde{v}$ is denoted as $\{p(v)\}_{v\in\mathcal{V}}$, the marginal distribution of $\tilde{s}$ is $\{q(s)\}_{s\in\mathcal{S}}$, and the conditional distribution of $\tilde{v}$ given $\tilde{s}$ is $\{p(v|s)\}_{(v,s)\in\mathcal{V}\times\mathcal{S}}$.[2]

Information acquisition is costly, we follow the framework introduced by Sims (1998, 2003) to model the information acquisition cost by mutual information. Let $\mathbb{H}(\mu) := -\sum_s \mu(s)\log\mu(s)$ be the entropy for the distribution μ. The cost of acquiring the signal $\tilde{s}$ is given by

$$\theta I(\tilde{s},\tilde{v}) := \theta\big[\mathbb{H}(p) - \sum_s q(s)\mathbb{H}(p(\cdot|s))\big], \tag{2.1}$$

where $\theta > 0$ represents a cost parameter of information acquisition. The difference $\mathbb{H}(p) - \sum_s q(s)\mathbb{H}(p(\cdot|s))$ represents the reduction in entropy from the prior distribution of the payoff to its posterior. A more accurate signal leads to a lower posterior entropy, hence a larger reduction from the prior entropy.

The interaction among the informed trader, market makers, and noise traders is the same as in Kyle (1985) and Back (1992). The cumulative demand from the informed

[2]When $\tilde{v}$ and $\tilde{s}$ have discrete distributions, $p(\cdot), q(\cdot)$, and $p(\cdot|\cdot)$ are probability distribution functions. When $\tilde{v}$ and $\tilde{s}$ have (conditional) continuous distributions, p and q represent the densities.

trader is X_t and the cumulative demand from all noise traders is $Z_t := \sigma_Z B_t$, where B_t is a standard one-dimensional Brownian motion and independent with $\tilde{v}$. The cumulative aggregate demand observed by all market makers is $Y_t := X_t + Z_t$. Competition and risk neutrality force market makers set the price as

$$P_t = \mathbb{E}[\tilde{v} \,|\, \mathcal{F}_t^Y], \tag{2.2}$$

where $\{\mathcal{F}_t^Y\}_{t\geqslant 0}$ is the filtration generated by Y and it represents the information available to market makers. In the equilibrium considered, Y_t is a sufficient statistic for P_t, i.e., the equilibrium price is given by $P_t = h(t, Y_t)$ for some function h.

The informed trader observes the market price P_t and computes Y_t, then she is able to infer Z_t. Hence, we assume that the informed trader observes Z_t directly. The information available to the informed trader is the signal realization at time 0, drawn from the distribution $\big(p(\cdot|\cdot), q(\cdot)\big)$ that she chooses, and the history of Z. The informed trader's realized profit for a given trading strategy ϑ_t is given by[3]

$$(\tilde{v} - P_T)X_T + \int_0^T X_{t-} dP_t = \int_0^T (\tilde{v} - P_t)\vartheta_t dt.$$

The informed trader chooses the distribution $(p(\cdot|\cdot), q(\cdot))$ and a trading strategy ϑ_t, measurable with respect to $\sigma(\tilde{s}) \vee \mathcal{F}_t^Z$, to maximize her (unconditional) expected profit net the information acquisition cost:

$$\mathbb{E}\Big[\int_0^T (\tilde{v} - P_t)\vartheta_t dt\Big] - \theta I(\tilde{s}, \tilde{v}). \tag{2.3}$$

An equilibrium among market participants is defined as follows.

Definition 2.1. *A Nash equilibrium is a triplet $(p(\cdot|\cdot), q(\cdot)), \vartheta_t$, and P_t such that*

(i) Given the informed trader's strategy $(p(\cdot|\cdot), q(\cdot))$ and ϑ_t, the market makers set the price as in (2.2);

(ii) Given the market makers' price setting strategy P_t, the informed trader chooses a distribution $(p(\cdot|\cdot), q(\cdot))$ and a trading strategy ϑ_t to maximize the objective in (2.3).

[3] Back (1992) shows that it is suboptimal for the informed trader to submit orders so that X_t has jumps or a nonzero quadratic variation. Therefore, without loss of generality, we represent the trading rate of the informed trader by a trading rate ϑ_t.

Compared to the notion of equilibrium in Kyle (1985), the informed trader's information acquisition strategy $(p(\cdot|\cdot), q(\cdot))$ is incorporated into (2.1) and it is the main focus of our paper.

2.2 An equivalent formulation

In this section, we will introduce an equivalent formulation of an equilibrium. We call this equivalent formulation *equilibrium with the estimated payoff*. In such an equilibrium, the payoff is replaced by the informed trader's best estimate of the payoff.

To introduce this equivalent formulation, we start from the informed trader's objective function. Because the informed trader's strategy ϑ_t is measurable with respect to $\sigma(\tilde{s})$, the law of iterated conditional expectation yields

$$\mathbb{E}\Big[\int_0^T \big(\tilde{v} - \mathbb{E}[\tilde{v}|\tilde{s}]\big)\vartheta_t dt \,\Big|\, \tilde{s}\Big] = 0.$$

The previous identity and the risk neutrality of the informed trader imply that, conditioning on the signal observed, the expected payoff from trading is

$$\mathbb{E}\Big[\int_0^T \big(\tilde{v} - P_t\big)\vartheta_t dt \,\Big|\, \tilde{s} = s\Big] = \mathbb{E}\Big[\int_0^T \big(\mathbb{E}[\tilde{v}|\tilde{s}] - P_t\big)\vartheta_t dt \,\Big|\, \tilde{s}\Big].$$

After taking the unconditional expectations on both sides of the previous equality, the informed trader's objective in (2.3) can be equivalently stated as

$$\mathbb{E}\Big[\int_0^T \big(\mathbb{E}[\tilde{v}|\tilde{s}] - P_t\big)\vartheta_t dt\Big] - \theta I(\tilde{s}, \tilde{v}). \tag{2.4}$$

Equation (2.4) implies that the expected profit of the informed trader is the same if the asset payoff $\tilde{v}$ is replaced by its conditional expectation $\mathbb{E}[\tilde{v}|\tilde{s}]$, the best estimate (minimizing the mean squared error) for the payoff after the informed trader observes the signal.

Given the price p, consider the informed trader's problem of designing the optimal trading strategy to maximize the expected profit. Consider a pricing rule

$$P_t = h(t, Y_t), \tag{2.5}$$

where $h(t, y) := \mathbb{E}\big[h(T, Z_T) \,|\, Z_t = y\big]$ for a function $h(T, \cdot)$. Among all strategies satisfying 'no doubling strategies' condition, (Back, 1992, Lemma 2) shows that the optimal trading

strategy, which maximizes the conditional expected profit $\mathbb{E}\big[\int_0^T \big(\mathbb{E}[\tilde{v}|\tilde{s}]-P_t\big)\vartheta_t dt \,|\, \tilde{s}\big]$, satisfies

$$P_T = h(T, Y_T) = \mathbb{E}[\tilde{v}|\tilde{s}]. \tag{2.6}$$

At the end of the trading horizon, informed agent's trading strategy reveals all her private information so that the market price converges to the informed agent's best estimate of the asset payoff. Equation (2.6) implies that $\mathbb{E}[\tilde{v}|\tilde{s}]$ is measurable with respect to $\sigma(Y_T)$. Hence Y_T should be weakly more informative than $\tilde{s}$. Meanwhile, Y_T is measurable with respect to the information generated by the history of noise trades Z and $\tilde{s}$. Given that $\tilde{v}$ and Z are independent, predicting $\tilde{v}$ using Y_T should be less precise than predicting $\tilde{v}$ using $\tilde{s}$. Therefore, Y_T should be weakly less informative than $\tilde{s}$. Combining both directions of the argument above, we obtain the following result.

Lemma 2.2. *Suppose that $\mathbb{E}[\tilde{v}|\tilde{s}]$ is measurable with respect to $\sigma(Y_T)$. Then*

$$\mathbb{E}[\tilde{v}|\tilde{s}] = \mathbb{E}[\tilde{v}|\mathcal{F}_T^Y]. \tag{2.7}$$

Suppose that the total demand Y, in its own filtration, has the same distribution as Z, then (2.5), (2.6), and (2.7) combined implies

$$P_t = h(t, Y_t) = \mathbb{E}\big[h(T, Y_T) \,|\, \mathcal{F}_t^Y\big] = \mathbb{E}\Big[\mathbb{E}\big[\tilde{v}|\tilde{s}\big] \,\Big|\, \mathcal{F}_t^Y\Big] = \mathbb{E}\Big[\mathbb{E}\big[\tilde{v}|\mathcal{F}_T^Y\big] \,\Big|\, \mathcal{F}_t^Y\Big] = \mathbb{E}\big[\tilde{v} \,|\, \mathcal{F}_t^Y\big], \tag{2.8}$$

where second to the last equality holds thanks to (2.7), and the last equality follows from the law of iterated conditional expectation. Therefore, the pricing rule (2.5) also satisfies the equilibrium requirement (2.2). Equation (2.8) indicates that the pricing rule is the same when the payoff $\tilde{v}$ is replaced by its conditional expectation $\mathbb{E}[\tilde{v}|\tilde{s}]$. Therefore, market makers learn from the total order flow what the informed trader has learned from her private signal.

These observations on the informed trader's expected profit and the pricing rule motivate us to introduce the following notion of equilibrium.

Definition 2.3. *A Nash equilibrium with the estimated payoff is a triplet $(p(\cdot|\cdot), q(\cdot))$, ϑ_t, and P_t such that*

(i) Given the informed trader's strategy $(p(\cdot|\cdot), q(\cdot))$ and ϑ_t, the market makers set the price as

$$P_t = \mathbb{E}\big[\mathbb{E}[\tilde{v}|\tilde{s}] \,\big|\, \mathcal{F}_t^Y\big]. \tag{2.9}$$

(ii) Given the market makers' price setting strategy P_t, the informed trader chooses a distribution $(p(\cdot|\cdot), q(\cdot))$ and a trading strategy ϑ_t to maximize the objective in (2.4)

When the informed trader observes the payoff $\tilde{v}$ at time 0, Back (1992) establishes a Markovian equilibrium, where the total order Y is a sufficient static for market makers, and the informed agent's trade is inconspicuous, which implies the total order Y in its own filtration has the same distribution as the noise order Z. Considering among the Markovian inconspicuous equilibria, we obtain the following result.

Proposition 2.4. *The Markovian inconspicuous equilibria in Definitions 2.1 and 2.3 are 1-to-1 correspondence.*

This result allows us to focus on searching an equilibrium with the estimated payoff introduced in Definition 2.3. For a given signal structure $(p(\cdot|\cdot), q(\cdot))$, the distribution of $\mathbb{E}[\tilde{v}|\tilde{s}]$ is fixed. An equilibrium between the optimal trading strategy, which maximizes the expected profit

$$\mathbb{E}\Big[\int_0^T \big(\mathbb{E}[\tilde{v}|\tilde{s}] - P_t\big)\vartheta_t dt\Big],$$

and a pricing rule, which satisfies (2.9), is a standard equilibrium in Kyle's model, where the asset payoff is replaced by its conditional expectation. In the next section, we recall a characterization of such an equilibrium by Back et al. (2021).

2.3 Characterization of the Kyle's equilibrium via optimal transport

Back et al. (2021) provides a characterization of the Kyle's equilibrium from an optimal transport point of view. In particular, they identify a representation for the informed trader's ex ante expected profit, which is instrumental in our analysis of the optimal signal structure.

Given distribution of $\mathbb{E}[\tilde{v}|\tilde{s}]$, let us first recall the main results of Back et al. (2021) for a risk neutral informed trader. To this end, we denote F_s as the distribution function of $\mathbb{E}[\tilde{v}|\tilde{s}]$ and G as the distribution function of Z_T, which has a normal distribution $N(0, \sigma_Z^2 T)$. The following result restates (Back et al., 2021, Theorems 3.1 and 3.2) in a one-dimensional setting with the asset payoff therein replaced by $\mathbb{E}[\tilde{v}|\tilde{s}]$. We also assume that F_s is absolutely continuous and is strictly increasing. We will show in later sections that this without loss of generality when the signal $\tilde{s}$ is chosen optimally.

Proposition 2.5 (Back et al. (2021)). *Let $h(T,y) = F_s^{-1} \circ G(y)$ and consider the pricing rule in (2.5). The trading strategy $dX_t = \vartheta_t dt$ where*

$$\vartheta_t = \frac{\tilde{\zeta} - Y_t}{T-t}, \quad \tilde{\zeta} = G \circ F_s^{-1}\big(\mathbb{E}[\tilde{v}|\tilde{s}]\big), \tag{2.10}$$

maximizes the informed trader's expected profit $\mathbb{E}\big[\int_0^T \big(\mathbb{E}[\tilde{v}|\tilde{s}] - P_t\big)\vartheta_t dt \,|\, \tilde{s}\big]$ in the class of trading strategies satisfying $\mathbb{E}\big[\int_0^T h(t, X_t + Z_t) dZ_t\big] = 0$. Given the trading strategy in (2.10), Y, in its own filtration, has the same distribution as Z, and the pricing rule (2.5) satisfies the equilibrium condition (2.9). Moreover, the informed trader's optimal (unconditional) expected profit is

$$\max_\theta \mathbb{E}\Big[\int_0^T \big(\mathbb{E}[\tilde{v}|\tilde{s}] - P_t\big)\vartheta_t dt\Big] = \frac{1}{2}\Big\{\mathbb{E}\big[\mathbb{E}[\tilde{v}|\tilde{s}]^2\big] + \mathbb{E}[Z_T^2] - W_2^2(F_s, G)\Big\}, \tag{2.11}$$

where $W_2^2(F_s, G)$ is the squared Wasserstein-2 distance between F_s and G.

The construction of the equilibrium is obtained by Back (1992). The random variable $\tilde{\zeta}$ in (2.10) has the same distribution as Z_T. When the informed trader follows the strategy (2.10), the total order Y is a Brownian bridge satisfying $Y_T = \tilde{\zeta}$. Moreover, Y, in its own filtration, shares the same distribution as Z. Therefore, the equilibrium identified in Proposition 2.5 is a Markovian inconspicuous equilibrium.

To understand the informed trader's unconditional expected profit in (2.11), let us first recall the informed trader's conditional expected profit given the signal realization $\tilde{s} = s$. Replacing the distribution of $\tilde{v}$ in Back (1992) by the distribution of $\mathbb{E}[\tilde{v}|\tilde{s}]$, Lemma 1 and 2 therein state that the optimal conditional expected profit is

$$J(s, y, t) = \mathbb{E}\big[j(s, y + Z_T - Z_t)\big], \tag{2.12}$$

where

$$j(s,y) := \int_y^{b^{-1}(\mathbb{E}[\tilde{v}|s])} \big(\mathbb{E}[\tilde{v}|s] - b(x)\big)dx$$

and $b(y) = h(T,y) = F_s^{-1} \circ G(y)$. Back et al. (2021) show that the function b is the gradient of a convex function Γ, which is called a Brenier potential. This function is chosen to satisfy a normalization constraint $\mathbb{E}[\Gamma(Z_T)] = 0$. The function b is also the unique monotone map

that transports the distribution of $\tilde{\zeta}$ to the distribution of $\mathbb{E}[\tilde{v}|\tilde{s}]$.[4] Now, observe that

$$\partial_y j(s,y) = b(y) - \mathbb{E}[\tilde{v}|s] = \Gamma'(y) - \mathbb{E}[\tilde{v}|s].$$

Integrating both side with respect to y and requiring that the integral is nonnegative and only equals to zero when $y = b^{-1}(\mathbb{E}[\tilde{v}|s])$, we obtain that

$$j(s,y) = \Gamma^*\big(\mathbb{E}[\tilde{v}|s]\big) - \big(y\mathbb{E}[\tilde{v}|s] - \Gamma(y)\big), \tag{2.13}$$

where $\Gamma^*\big(\mathbb{E}[\tilde{v}|s]\big) = \sup_y \big\{y\mathbb{E}[\tilde{v}|s] - \Gamma(y)\big\}$ is the conjugate of Γ. Combining (2.12) and (2.13), we obtain

$$J(s,y,t) = \Gamma^*\big(\mathbb{E}[\tilde{v}|s]\big) + \mathbb{E}\big[\Gamma(y + Z_T - Z_t)\big] - y\mathbb{E}[\tilde{v}|s]. \tag{2.14}$$

Setting $(y,t) = (0,0)$ and taking expectation with respect to the law of $\tilde{s}$, we obtain from the previous equation and the normalization constraint $\mathbb{E}[\Gamma(Z_T)] = 0$ that

$$\max_\theta \mathbb{E}\Big[\int_0^T \big(\mathbb{E}[\tilde{v}|\tilde{s}] - P_t\big)\vartheta_t dt\Big] = \mathbb{E}\big[J(\tilde{s},0,0)\big] = \mathbb{E}\Big[\Gamma^*\big(\mathbb{E}[\tilde{v}|\tilde{s}]\big)\Big]. \tag{2.15}$$

Using Monge-Kantorovich duality, Back et al. (2021) further transform the right-hand side of (2.15) to the right-hand side of (2.11). Therefore, it is the optimal transport interpretation of the pricing function $F_s^{-1} \circ G$ chosen by Back (1992) that links the informed trader's unconditional expected profit with the Wasserstein-2 distance in (2.11). The optimal transport connects the expected payoff $\mathbb{E}[\tilde{v}|\tilde{s}]$ and the informed trader's trading target $\tilde{\zeta}$. When the expected payoff has a normal distribution, the trading target is a linear function of the expected payoff. Once the trading target $\tilde{\zeta}$ is obtained, Back et al. (2021) show that the trading strategy (2.10) is time consistent and the informed trader's optimal conditional expected profit is represented by (2.14).

Back et al. (2021) utilize this intriguing connection with optimal transport to generalize the solutions of Kyle model from Kyle (1985) and Back (1992) a multiple dimensional setting with potential discrete payoff distributions.[5]

Next, let us transform the representation (2.11) using the centered expected payoff

[4] See (Galichon, 2016, Section 4) for more discussions on the 1-dimensional optimal transport.

[5] Back et al. (2021) allow for multi-variate and discrete distributions for $\mathbb{E}[\tilde{v}|\tilde{s}]$. When $\mathbb{E}[\tilde{v}|\tilde{s}]$ has a discrete distribution, the optimal trading strategy still takes the form in (2.10), but $\tilde{\zeta}$ is constructed using additional randomization.

$\mathbb{E}[\tilde{v}|\tilde{s}] - \mathbb{E}[\tilde{v}]$. To this end, note that

$$\mathbb{E}\big[\mathbb{E}[\tilde{v}|\tilde{s}]^2\big] = \mathrm{Var}\big(\mathbb{E}[\tilde{v}|\tilde{s}]\big) + \mathbb{E}\big[\mathbb{E}[\tilde{v}|\tilde{s}]\big]^2 = \mathrm{Var}\big(\mathbb{E}[\tilde{v}|\tilde{s}]\big) + \mathbb{E}[\tilde{v}]^2.$$

Meanwhile, the Mallows distance decomposition property for Wasserstein-2 distance implies that

$$W_2^2(F_s, G) = \mathbb{E}[\tilde{v}]^2 + W_2^2(F_s^c, G)^2,$$

where F_s^c is the distribution for the centered expected payoff $\mathbb{E}[\tilde{v} \,|\, \tilde{s}] - \mathbb{E}[\tilde{v}]$. Plugging the previous two equations into the right-hand side of (2.11), the unconditional expected profit of the insider is transformed to

$$\frac{1}{2}\Big\{\mathrm{Var}\big(\mathbb{E}[\tilde{v}|\tilde{s}]\big) + \mathbb{E}\big[Z_T^2\big] - W_2^2(F_s^c, G)\Big\}. \tag{2.16}$$

The decomposition (2.16) highlights the impact of the conditional expected payoff $\mathbb{E}[\tilde{v}|\tilde{s}]$ and the noise trade distribution G on the informed trader's expected profit. The variance of the conditional expected payoff increases the expected profit of the informed trader. The larger the variance, the more informational advantage the informed trader has over the market makers. After observing her private signal, the informed trader updates her payoff estimate to $\mathbb{E}[\tilde{v}|\tilde{s}]$. Meanwhile, market markers only know the prior expectation $\mathbb{E}[\tilde{v}]$. A higher variance of $\mathbb{E}[\tilde{v}|\tilde{s}]$ indicates a higher information asymmetry, hence a better trading opportunity for the informed trader. We call this the *profit potential* of information. However, a higher variance of $\mathbb{E}[\tilde{v}|\tilde{s}]$ also introduces a higher cost to the informed trader. If the informed trader submits order aggressively, it is harder for her to hide her informed orders among noise trades, particularly if the distribution of trades in the tails differs from the normal distribution; the market makers learn informed trader's private information more efficiently, reducing the informational advantage of the informed trader. We call this the information *leakage cost*. This effect introduces the third term in (2.16). When the variance of $\mathbb{E}[\tilde{v}|\tilde{s}]$ increases, the centered distribution F_s^c is further away from the noise orders' normal distribution G. This leads to a higher Wasserstein-2 distance, hence reduces the informed trader's expected profit.

When the information acquisition cost is present, we will show in the next section that the information acquisition cost interacts with both the profit potential of information and the information leakage cost. The optimal design of the signal will balance the profit potential, the leakage cost, and the information acquisition cost.

To conclude this section, we plug (2.11) into (2.4) to obtain the following information structure design problem for the informed trader:

$$\max_{p(\cdot|\cdot),q(\cdot)} \frac{1}{2}\Big\{\mathbb{E}\big[\mathbb{E}[\tilde{v}|\tilde{s}]^2\big] + \mathbb{E}[Z_T^2] - W_2^2(F_s, G)\Big\} - \theta I(\tilde{s}, \tilde{v}), \tag{2.17}$$

subject to

$$\int q(s)ds = 1, \tag{2.18}$$

$$\int p(v|s)dv = 1, \quad \text{for any } s, \tag{2.19}$$

$$\int p(v|s)q(s)ds = p(v), \quad \text{for any } v. \tag{2.20}$$

As we discussed before, the first three terms in (2.17) represents the informed trader's optimal (unconditional) expected profit from trading when the informed trader samples from a given signal distribution $(p(\cdot|\cdot), q(\cdot))$. The last term represents the information acquisition cost. The first two constraints ensures $q(\cdot)$ and $p(\cdot|\cdot)$ are probability densities. The last constraint is the Bayes formula between the posterior, marginal, and the given prior $p(\cdot)$. This is the Bayes plausibility condition.

In the next three sections, we will examine the trade-offs among the profit potential, the information leakage, and the information acquisitions cost. We will present the optimal signal structure for different payoff distributions.

3 Normal-distributed payoff

In this section, we consider the case where the payoff $\tilde{v}$ follows a normal distribution. We first illustrate the optimal signal precision when the signal has a normal distribution. Then we show that the normal signal distribution is optimal.

Consider a normal-distributed payoff $\tilde{v} \sim N(\mu_v, \sigma_v^2)$. Assume that the informed trader observes a signal

$$\tilde{s} = \tilde{v} + \tilde{\epsilon}, \tag{3.1}$$

where $\tilde{\epsilon}$ is independent of $\tilde{v}$ and follows a normal distribution $N(0, \sigma_\epsilon^2)$. Let us first derive the optimal signal precision $1/\sigma_\epsilon^2$, then prove that the signal structure in (3.1) is optimal.

Given the signal structure in (3.1), the posterior distribution of $\tilde{v}$, conditioning on $\tilde{s}$, is

normal with the mean and variance

$$\mathbb{E}[\tilde{v}|\tilde{s}] = \mathbb{E}[\tilde{v}] + \xi\big(\tilde{s} - \mathbb{E}[\tilde{v}]\big) \quad \text{and} \quad \sigma^2_{v|s} = \sigma^2_v(1-\xi),$$

respectively, where $\xi := \frac{\sigma_v^2}{\sigma_s^2} \in [0,1]$ and $\sigma_s^2 := \sigma_v^2 + \sigma_\epsilon^2$. Therefore, the conditional expectation of payoff $\mathbb{E}[\tilde{v}|\tilde{s}]$ follows a normal distribution with mean $\mathbb{E}[\tilde{v}]$ and variance $\xi^2\sigma_s^2 = \xi\sigma_v^2$.

The informed trader's objective function in (2.17) can be simplified when both the payoff and the signal are normal distributed. To this end, given two normal distributions $F_s \sim N(\mathbb{E}[\tilde{v}], \xi\sigma_v^2)$ and $G \sim N(0, \sigma_Z^2 T)$, their Wasserstein-2 distance is

$$W_2^2(F_s, G) = \mathbb{E}[\tilde{v}]^2 + \big(\sigma_Z\sqrt{T} - \sqrt{\xi}\sigma_v\big)^2.$$

Therefore, the informed trader's optimal (unconditional) expected profit in (2.11) is

$$\begin{aligned}\frac{1}{2}\Big\{\mathbb{E}\big[\mathbb{E}[\tilde{v}|\tilde{s}]^2\big] + \mathbb{E}[Z_T^2] - W_2^2(F_s, G)\Big\} =& \frac{1}{2}\Big\{\xi\sigma_v^2 + \mathbb{E}[\tilde{v}]^2 + \sigma_Z^2 T - \mathbb{E}[\tilde{v}]^2 - \big(\sigma_Z\sqrt{T} - \sqrt{\xi}\sigma_v\big)^2\Big\} \\ =& \sigma_v\sigma_Z\sqrt{T}\sqrt{\xi}.\end{aligned}$$

In the previous equation, the net between the profit potential $\mathbb{E}\big[\mathbb{E}[\tilde{v}|\tilde{s}]^2\big]$ and the information leakage cost $W_2^2(F_s, G)$ increases with $\sqrt{\xi}$. Hence, a more precise signal increases the expected profit from trading.

For the information acquisition cost, recall that the entropy for a normal distribution $N(\mu, \sigma^2)$ is $\frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2}$. Therefore, the information acquisition cost is

$$\theta I(\tilde{s}; \tilde{v}) = \frac{\theta}{2}\log\big(\frac{\sigma_v^2}{\sigma_{v|s}^2}\big) = \frac{\theta}{2}\log\big(\frac{1}{1-\xi}\big).$$

Plugging the previous two equations into (2.17), we obtain the following optimization problem for the informed trader:

$$\max_{\xi\in[0,1]} \Big\{\sigma_v\sigma_Z\sqrt{T}\sqrt{\xi} - \frac{\theta}{2}\log\big(\frac{1}{1-\xi}\big)\Big\}. \tag{3.2}$$

The first term in the objective function of (3.2) represents the expected profit. More precise signal yields a higher ξ, which improves the expected profit from trading. The second term in (3.2) represents the information acquisition cost, whereas a higher ξ means a more precise signal hence a higher cost. Therefore, the optimal choice of ξ balances the marginal benefit

and the marginal cost of information. The optimal ξ and the optimal signal precision are presented in the following result.

Proposition 3.1. *Suppose that the payoff $\tilde{v}$ follows a normal distribution $N(\mu_v, \sigma_v^2)$ and the informed trader observes a signal of the form (3.1). Then the optimal conditional precision of the signal is*

$$\frac{1}{\sigma_\epsilon^2} = \frac{1}{\sigma_v^2} \frac{\xi^*}{1-\xi^*}, \tag{3.3}$$

where ξ^* is the unique maximizer of the problem (3.2). The optimal precision and ξ^* both decrease with $\frac{\theta}{\sigma_v \sigma_Z \sqrt{T}}$. Moreover, the Kyle's λ decreases, as the information acquisition parameter θ increases, the prior standard deviation σ_v decreases, or the noise trade volatility σ_Z increases.

The previous result shows that there is a negative relation between the optimal signal precision and $\frac{\theta}{\sigma_v \sigma_Z \sqrt{T}}$. When θ increases, information acquisition is more costly, the informed trader acquires less precise signal as a result. When σ_v increases, ξ^* increases as well, leading to a larger variance of the conditional expected payoff $\mathbb{E}[\tilde{v}|\tilde{s}]$. Even though a higher variance of $\mathbb{E}[\tilde{v}|\tilde{s}]$ increases the information leakage cost, the improvement in profit potential $\mathbb{E}\big[\mathbb{E}[\tilde{v}|\tilde{s}]^2\big]$ is more, leading to a higher expected profit for the informed trader. When $\sigma_Z\sqrt{T}$ increases, noise trade is more volatile, the informed trader can better hide her orders, hence the concern on information leakage is reduced. Therefore, the informed trader increases the signal precision to take advantage of this market environment.

When the expected payoff $\mathbb{E}[\tilde{v}|\tilde{s}]$ has a normal distribution, Kyle's λ is given by $\frac{\sqrt{\xi^*}\sigma_v}{\sigma_Z\sqrt{T}}$, where the numerator is the standard deviation of $\mathbb{E}[\tilde{v}|\tilde{s}]$. When either θ increases or σ_v decreases, ξ^* decreases, leading to a lower Kyle's λ. When σ_Z increases, even though ξ^* increases as well, the proof of Proposition 3.1 shows that the effect in the denominator of Kyle's λ dominates, hence λ decreases.

As a comparison, let us consider a case where the information acquisition cost is an increasing and convex function of the noise precision $1/\sigma_\epsilon^2$ in (3.1). This cost formulation is widely used when the signal is assumed to have a normal distribution. Definition of ξ implies $\frac{\sigma_v^2}{\sigma_\epsilon^2} = \frac{\xi}{1-\xi}$. Therefore, we consider the following optimization problem for the informed trader:

$$\max_{\xi\in[0,1]} \Big\{\sigma_v\sigma_Z\sqrt{T}\sqrt{\xi} - \theta\, c\Big(\frac{\xi}{1-\xi}\Big)\Big\},$$

where c is an increasing and convex function. It follows from the first-order condition of ξ that the optimal $\hat{\xi}$ and also the optimal precision both decrease with $\frac{\theta}{\sigma_v \sigma_Z \sqrt{T}}$. Therefore, the

precision-based cost delivers the same qualitative result as in Proposition 3.1. However, when the payoff has a normal-distributed prior, the entropy-based cost ensures that the normal signal distribution $(p(\cdot|\cdot), q(\cdot))$ is optimal among a wide class of distributions satisfying some non-degeneracy condition. Moreover, this optimal signal structure design can be implemented by the standard signal-plus-noise form in (3.1).

Proposition 3.2. *Suppose that the payoff is normal-distributed $\tilde{v} \sim N(\mu_v, \sigma_v^2)$. Then an optimal $(p(\cdot|\cdot), q(\cdot))$ for the problem (2.17), together with the constraints (2.18) - (2.20), is normal, among all distributions so that $\frac{d}{ds}\mathbb{E}[\tilde{v}|\tilde{s} = s] \neq 0$ for almost all s. Moreover, this optimal signal structure can be implemented by (3.1).*

The restriction $\frac{d}{ds}\mathbb{E}[\tilde{v}|\tilde{s} = s] \neq 0$ for almost all s is a non-degeneracy condition. Without this restriction, there can be two signals s_1 and s_2 such that $\mathbb{E}[\tilde{v}|\tilde{s} = s_1] = \mathbb{E}[\tilde{v}|\tilde{s} = s_2]$. Information acquisition is costly, then it is wasteful for the informed trader to design two distinct signals which lead to the same estimates on the payoff.

Intuition for Proposition 3.2 is the following. The Wasserstein-2 distance between the distribution F_s for $\mathbb{E}[\tilde{v}|\tilde{s}]$ and the normal distribution G for noise trades admits the Gelbrich (1990) lower bound:

$$W_2^2(F_s, G) \geqslant (\mu_G - \mathbb{E}[\tilde{v}])^2 + (\sigma_G - \sqrt{K})^2, \tag{3.4}$$

where $\mu_G = 0$ is the mean of G, $\sigma_G = \sigma_Z\sqrt{T}$ and $\sqrt{K}$ are the standard deviations for G and F_s, respectively. Moreover, the lower bound is attained when F_s is a normal distribution. We replace $W_2^2(F_s, G)$ in (2.17) by the lower bound in (3.4), hence obtain a new objective function which is an upper bound of the objective in (2.17). The entropy information acquisition cost yields an normal distribution $(p^*(\cdot|\cdot), q^*(\cdot))$ as a maximizer for this upper bound objective function. Because $(p^*(\cdot|\cdot), q^*(\cdot))$ is normal, the Gelbrich lower bound is attained. Hence, $(p^*(\cdot|\cdot), q^*(\cdot))$ is also a maximizer for the original objective function in (2.17). This optimal distribution can be implemented by the signal (3.1). Hence, this signal structure, which we assumed for Proposition 3.1, is optimal.

4 Discretely distributed payoff

In this section, we consider a discretely distributed payoff $\tilde{v}$ with a finite state space $\mathcal{V} = \{v_1, \ldots, v_N\}$. Even though the payoff state space is discrete and finite, our analytical

result shows that optimal signals have continuous distributions. Moreover, any continuous signal leads to the same ex ante optimal value for the informed trader and the same distribution of the conditional expected payoff.

4.1 Continuous signal

We consider a signal $\tilde{s}$ with a continuous distribution which admits a density with respect to the Lebesgue measure on $\mathbb{R}$. We denote it cumulative distribution function (CDF) as Q and its probability distribution function (density) as q. The support of q can be a subset of $\mathbb{R}$ or $\mathbb{R}$ itself. Because the signal states can be reordered, it is without loss of generality to consider the class of distributions such that $\text{support}(q) \ni s \mapsto \mathbb{E}[\tilde{v}|s]$ is strictly increasing.

Let us first formulate the optimization problem for the informed trader. To this end, for a given signal structure and the associated distribution of $\mathbb{E}[\tilde{v}|\tilde{s}]$, recall that the unconditional expected profit of the insider is given by (2.11). Using the quantile functions F_s^{-1} and G^{-1}, the Wasserstein 2 distance between F_s and G can be expressed as

$$W_2^2(F_s, G) = \int_0^1 \left[F_s^{-1}(u) - G^{-1}(u)\right]^2 du.$$

Therefore, the expected profit of the informed trader is

$$\begin{aligned}\frac{1}{2}\Big\{\mathbb{E}\big[\mathbb{E}[\tilde{v}|\tilde{s}]^2\big] + \mathbb{E}[Z_T^2] - W_2^2(F_s, G)\Big\} &= \frac{1}{2}\int_0^1 \Big\{\big(F_s^{-1}(u)\big)^2 + \big(G^{-1}(u)\big)^2 - \big[F_s^{-1}(u) - G^{-1}(u)\big]^2\Big\} du \\ &= \int_0^1 F_s^{-1}(u) G^{-1}(u)\, du. \qquad (4.1)\end{aligned}$$

The first equality above follows from the fact that $\mathbb{E}\big[\mathbb{E}[\tilde{v}|\tilde{s}]^2\big] = \int_0^1 (F_s^{-1}(u))^2 du$ and $\mathbb{E}[Z_T^2] = \int_0^1 (G^{-1}(u))^2 du$. Using the quantile formulation in (4.1), the following result presents the informed trader's optimization problem.

Lemma 4.1. *The informed trader's optimization problem is*

$$\max_{p(\cdot|\cdot),q(\cdot)} \underbrace{\int \mathbb{E}[\tilde{v}|s]G^{-1}\big(Q(s)\big)q(s)\,ds}_{\textit{Unconditional Expected Profit}} - \underbrace{\theta\Big\{\sum_{n=1}^{N}\int \log\big(p(v_n|s)\big)p(v_n|s)q(s)\,ds - \sum_{n=1}^{N}\log\big(p(v_n)\big)p(v_n)\Big\}}_{\textit{Information Acquisition Cost}}, \tag{4.2}$$

subject to

$$\sum_{n=1}^{N} p(v_n|s) = 1, \quad \textit{for any } s, \tag{4.3}$$

$$\int p(v_n|s)q(s)\,ds = p(v_n), \quad \textit{for any } 1 \leqslant n \leqslant N. \tag{4.4}$$

The first term in (4.2) represents the unconditional expected profit from trading and the second term is the information acquisition cost. The constraint (4.3) ensures that $p(\cdot|\cdot)$ is a posterior probability and the constraint (4.4) represents the Bayes plausibility. The following result presents the informed trader's optimal choice of signal structure among continuous-distributed signals and the associated optimal value.

Proposition 4.2. *The following statements hold.*

(i) The optimal posterior probabilities are given by

$$p(v_n|s) = \frac{e^{\frac{1}{\theta}\left[v_n G^{-1}(Q(s))+\mu_n\right]}}{\sum_{n'=1}^{N} e^{\frac{1}{\theta}\left[v_{n'} G^{-1}(Q(s))+\mu_{n'}\right]}}, \quad s \in \textit{support}(q), n \in \{1,\dots,N\}. \tag{4.5}$$

In the previous equation, the vector $(\mu_1,\dots,\mu_N)$ exists and is unique (up to additive constants, i.e., $(\mu_1,\dots,\mu_N)$ is identical to $(\mu_1+c,\dots,\mu_N+c)$ for any constant c) such that the Bayes plausibility (4.4) is satisfied.

(ii) Any continuous signal $\tilde{s}$, such that $s \mapsto \mathbb{E}[\tilde{v}|s]$ is strictly increasing, achieves the same optimal value

$$\theta\,\mathbb{E}\Big[\log\Big(\sum_{n=1}^{N} e^{(v_n\tilde{z}+\mu_n)/\theta}\Big)\Big] + \theta\sum_{n=1}^{N}\log\big(p(v_n)\big)p(v_n) - \sum_{n=1}^{N}\mu_n p(v_n), \tag{4.6}$$

where $\tilde{z}$ is a random variable with normal distribution $N(0,\sigma_Z^2 T)$.

(iii) The conditional expected payoff $\mathbb{E}[\tilde{v}|\tilde{s}]$ *has the same distribution as*

$$\frac{\sum_{n=1}^{N} v_n e^{(v_n \tilde{z}+\mu_n)/\theta}}{\sum_{n=1}^{N} e^{(v_n \tilde{z}+\mu_n)/\theta}}. \tag{4.7}$$

In particular, any continuous signal $\tilde{s}$ *leads to the same expected payoff distribution.*

It is surprising that any continuous signal achieves the same optimal value. To understand this result, we start from the optimization problem (4.2). Introducing Lagrangian multipliers $\nu(s)$ for the constraint (4.3) and $(\mu_1, \ldots, \mu_N)$ for the constraint (4.4), we obtain the optimal posteriors in (4.5) from the first-order condition of $p(v_n|s)$. We also obtain the following objective function for the informed trader:

$$\underbrace{\theta \int \log\Big(\sum_{n=1}^{N} e^{\frac{1}{\theta}\left(v_n G^{-1}(Q(s))+\mu_n\right)}\Big) q(s)\, ds}_{\text{Expected Profit } +\theta\, \mathbb{E}[\mathbb{H}(\tilde{v}|\tilde{s})]} + \underbrace{\theta \sum_{n=1}^{N} \log\big(p(v_n)\big) p(v_n)}_{-\theta\, \mathbb{H}(\tilde{v})} - \underbrace{\sum_{n=1}^{N} \mu_n p(v_n)}_{\text{Prior Adjustment}}. \tag{4.8}$$

The first term represents the expected profit from trading and the expected entropy of the posterior payoff, when the informed trader chooses the optimal posterior as in (4.5). The second term in (4.8) is proportional to the entropy of the prior payoff distribution. This term does not depend on choice of the informed trader. The last term in (4.8) represents an adjustment term associated with the Bayes plausibility (4.4). The Lagrangian multipliers $(\mu_1, \ldots, \mu_N)$ are chosen so that (4.4) is satisfied. These Lagrangian multipliers introduce this adjustment term. In a two-state symmetric case (i.e., $N = 2$ and $p(v_1) = p(v_2) = 0.5$), due to the symmetric of noise trade distribution, Proposition A.2 in the appendix shows that $\mu_1 = \mu_2 = 0$, hence this adjustment term vanishes.

Let us focus on the first term in (4.8). Recall that $Q(\tilde{s})$ has a uniform distribution on $[0, 1]$, when $\tilde{s}$ has a continuous distribution. Hence $G^{-1}(Q(\tilde{s}))$ has a normal distribution $N(0, \sigma_Z^2 T)$. Introducing a random variable $\tilde{z}$ with such a normal distribution, we obtain the following equivalent form for the first term in (4.8):

$$\begin{aligned}\theta \int \log\Big(\sum_{n=1}^{N} e^{\frac{1}{\theta}\left(v_n G^{-1}(Q(s))+\mu_n\right)}\Big) q(s)\, ds =& \theta\, \mathbb{E}\Big[\log\Big(\sum_{n=1}^{N} e^{\frac{1}{\theta}\left(v_n G^{-1}(Q(\tilde{s}))+\mu_n\right)}\Big)\Big] \\ =& \theta\, \mathbb{E}\Big[\log\Big(\sum_{n=1}^{N} e^{\frac{1}{\theta}\left(v_n \tilde{z}+\mu_n\right)}\Big)\Big].\end{aligned} \tag{4.9}$$

Combining (4.8) and (4.9) yields the value in (4.6). Meanwhile, we show in the proof of

Proposition (4.2) that $(\mu_1, \ldots, \mu_N)$ exists, unique (up to an additive constant), and is the same for any choice of continuously distributed signal $\tilde{s}$. Therefore, the value (4.6) is the same for any continuous signal, it is the optimal value for the informed trader's optimization problem in Lemma 4.1.

For any continuous signal $\tilde{s}$ and the associated optimal posterior in (4.5), the conditional expected payoff is given by (4.7). The distribution of the conditional expected payoff is the same for any continuous signal choice. This follows again from the fact that $G^{-1}(Q(\tilde{s}))$ has the same distribution as $\tilde{z}$ for any continuously distributed $\tilde{s}$. Given that the distribution of the conditional expected payoff is independent with the signal choice, both the unconditional expected profit and the information acquisition cost in (4.2) are independent of the signal choice as well, when the informed trader chooses the optimal posterior (4.5).

Similar to the rational inattention literature, the first two terms in the objective (4.8) can be rewritten as *posterior separable* form:

$$\int V^{\text{net}}(s) q(s)\, ds - \theta \mathbb{H}(\tilde{v}),$$

where $V^{\text{net}}(s) := \theta \log \left(\sum_{n=1}^{N} e^{\frac{1}{\theta}(v_n G^{-1}(Q(s)) + \mu_n)} \right)$ is the net utility associate with signal state s; see Caplin and Dean (2013); Caplin, Dean and Leahy (2019, 2022). Different from these papers, where the net utility only depends on the posterior distribution of states, our net utility depends on the entire marginal for the signal via the signal CDF Q and the Lagrangian multipliers $(\mu_1, \cdots, \mu_N)$.

Even though any continuous signal leads to the same optimal value for the informed trader, the optimal posterior in (4.5) depends on the choice of signal. Any pair of $\big(p(\cdot|\cdot), q(\cdot)\big)$ leads to the same optimal value. Hence the informed trader can choose the normal distributed $\tilde{z}$ as the signal. The associated optimal posterior is then

$$p(v_n|z) = \frac{e^{(v_n z + \mu_n)/\theta}}{\sum_{n'=1}^{N} e^{(v_{n'} z + \mu_{n'})/\theta}}, \quad z \in \mathbb{R}, n \in \{1, \ldots, N\}. \tag{4.10}$$

This optimal posterior has a logit form. The optimal posteriors are determined by three forces. First, the posterior probability $p(v_n|z)$ increases with $v_n z$. When the informed trader receives a positive/negative signal, higher/lower payoff states receive larger posterior probabilities. When z increases, high payoff states are associated with higher weights. In the limit $z \to \infty$, the highest payoff state receives the posterior probability one and other payoff states have zero posterior probability, i.e., $\lim_{z\to\infty} p(v_{\max}|z) = 1$, for $v_{\max} = \max\{v_1, \ldots, v_N\}$,

and $\lim_{z\to\infty} p(v_n|z) = 0$ for $v_n \neq v_{\max}$.

Second, the distinction in the posterior probabilities $p(v_n|z)$ among different v_n and z reduces as θ increases. A higher information acquisition cost increases the cost of distinguishing among different signal and payoff states. Therefore, the informed trader decreases the differences between different $p(v_n|z)$, i.e., the signal is less informative about the payoff. In the limit as $\theta \to \infty$, $(\mu_1, \ldots, \mu_N)$ adjusts so that the posterior $p(v_n|z)$ converges to the prior $p(v_n)$ for any n and z. Hence the conditional expectation $\mathbb{E}[\tilde{v}|z]$ converges to the prior expectation $\mathbb{E}[\tilde{v}]$.

Third, the vector $(\mu_1, \ldots, \mu_N)$ represents the Lagrangian multipliers for the Bayes plausibility (4.4). Plugging (4.10) into (4.4), we obtain

$$\int \xi_n k_n(z)\zeta(z)\, dz = p(v_n), \quad \text{where } \xi_n := e^{\mu_n/\theta},\ k_n(z) := e^{v_n z/\theta}, \text{ and } \zeta(z) := \frac{q(z)}{\sum_{n'=1}^{N} \xi_{n'} k_{n'}(z)}, \tag{4.11}$$

for $1 \leqslant n \leqslant N$. These N identities are equivalent to the following matrix identity

$$\int \operatorname{diag}(\xi_1, \ldots, \xi_N)\, k(z)\, \zeta(z)\, dz = p, \tag{4.12}$$

where $\operatorname{diag}(\xi_1, \ldots, \xi_N)$ is the diagonal matrix with diagonal entries $(\xi_1, \ldots, \xi_N)$, $k(z)$ is a N-dimensional vector $(k_1(z), \ldots, k_N(z))$, and p is the column vector $(p(v_1), \ldots, p(v_N))$. The identity (4.12) states that the row integral of $\operatorname{diag}(\xi_1, \ldots, \xi_N)\, k(z)\, \zeta(z)$ is p. Meanwhile, the definition of $\zeta(z)$ implies that the column sums of $\operatorname{diag}(\xi_1, \ldots, \xi_N)\, k(z)\, \zeta(z)$ is $q(z)$. Problem (4.12) is the (continuous) *Sinkhorn* problem, which matches a discrete marginal p and a continuous marginal q. For given k, marginals p and q, there exists a positive vector $(\xi_1, \ldots, \xi_N)$ satisfying (4.12). Moreover, this vector is unique up to a multiplicative scaling factor, i.e., if $(\xi_1, \ldots, \xi_N)$ solves (4.12), then so is $(c\,\xi_1, \ldots, c\,\xi_N)$ for any nonzero c; see Rüschendorf and Thomsen (1993).[6] From the definition of ξ's, we see that there exists a vector $(\mu_1, \ldots, \mu_N)$ so that p, defined in (4.10), satisfies (4.4). Moreover, $(\mu_1, \ldots, \mu_N)$ is

[6]Suppose that the support of the marginal q is on finite number of points $\{s_1, \ldots, s_M\}$. The Sinkhorn problem is also called the matrix scaling problem. When $M = N$, the existence and uniqueness of the solution for the Sinkhorn problem were obtained by Sinkhorn (1964) and Sinkhorn and Knopp (1967). Convergence of an iterative algorithm to identify the solution was also obtained in these papers. The existence and uniqueness of the solution are presented by (Peyré and Cuturi, 2019, Proposition 4.3) when M and N can be different. The result was generalized by Rüschendorf and Thomsen (1993) for general marginals. The convergence of the iterative algorithm for continuous marginals was proved by Rüschendorf (1995). We present the iterative algorithm to solve (4.11) in Appendix A.13. The Sinkhorn problem also relates to the entropic regularization of optimal transport; see (Peyré and Cuturi, 2019, Section 4) for the connection and further references.

unique up to an additive constant.

Given k and q, there is a positive relationship between μ_n and $p(v_n)$.[7] Therefore, when the marginal $p(v_n)$ increases, μ_n increases as well, leading to a higher posterior $p(v_n|z)$ for all signal realizations.

In an rational inattention discrete choice problem, Matějka and McKay (2015) obtain a logit form of optimal state-dependent choice probabilities, which provides a foundation for the multinomial logit model. Due to the recommendation lemma (Matějka and McKay, 2015, Lemma 1), signals one-to-one correspond to actions in Matějka and McKay (2015), hence their optimal choice probabilities are the conditional probabilities of signals or actions given states; this is not the case in our problem. Different from Matějka and McKay (2015), our posterior probabilities in (4.10) represent the conditional probabilities of states given signals.

Choosing the normal-distributed signal $\tilde{z}$, we show in the proof of Proposition 4.2 that $\mathbb{R} \ni z \mapsto \mathbb{E}[\tilde{v}|z]$ is strictly increasing. Hence the signal has a monotone relation with the informed trader's best estimate of the payoff. In particular, when the payoff has two states, i.e., $N = 2$, the conditional expected payoff $\mathbb{E}[\tilde{v}|\tilde{z}]$ admits the following explicit density.

Corollary 4.3. *When $N = 2$, the density of the conditional expected payoff $\mathbb{E}[\tilde{v}|\tilde{z}]$ is*

$$f_s(v) = \frac{\theta}{\sigma_Z\sqrt{T}(v_2 - v)(v - v_1)}\phi\Big(\frac{\theta}{\sigma_Z\sqrt{T}(v_1 - v_2)}\log\Big(\frac{v_2 - v}{v - v_1}\Big) - \frac{\mu_1 - \mu_2}{\sigma_Z\sqrt{T}(v_1 - v_2)}\Big), \quad (4.13)$$

for $v \in (v_1, v_2)$, where $\phi(\cdot)$ is the density of the standard normal distribution.

We will illustrate several comparative statics for the conditional expected payoff after discussing the optimality of the continuous signals.

4.2 Continuous signal is optimal

In our previous section, signal $\tilde{s}$ is assumed to have a continuous distribution. However,

Is continuous signal optimal for a discretely-distributed payoff?

To answer this question, we first consider discretely distributed signals: The signal $\tilde{s}$ could take M different values $\{s_1, \ldots, s_M\}$, where the number of signal states M is not necessarily equal to the number of payoff states N. The probability distribution function of

[7] A proof is provided in Appendix A.13.

the discretely distributed signal is denoted as $q(s_m)$ and the CDF is $Q_m := \sum_{i=1}^{m} q(s_i)$ for $1 \leqslant m \leqslant M$. We still denote the CDF for the conditional expected payoff $\mathbb{E}[\tilde{v}|\tilde{s}]$ as F_s.

For discretely distributed signals, let us first formulate the informed trader's optimization problem. To this end, note that the quantile function of $\mathbb{E}[\tilde{v}|\tilde{s}]$ takes the following form:

$$F_s^{-1}(u) = \mathbb{E}[\tilde{v}|s_m], \qquad \text{when } Q_{m-1} < u \leqslant Q_m. \tag{4.14}$$

Therefore, the insider's unconditional expected profit in (4.1) is

$$\int_0^1 F_s^{-1}(u)G^{-1}(u)\,du = \sum_{m=1}^{M} \mathbb{E}[\tilde{v}|s_m] \int_{Q_{m-1}}^{Q_m} G^{-1}(u)du = \sum_{m=1}^{M} \mathbb{E}[\tilde{v}|s_m]I_m q(s_m), \tag{4.15}$$

where $I_m := \frac{1}{q(s_m)} \int_{Q_{m-1}}^{Q_m} G^{-1}(u)du$. Incorporate the information acquisition cost, the insider's optimization problem is

$$\max_{p(\cdot|\cdot),q(\cdot)} \sum_{m=1}^{M} \mathbb{E}[\tilde{v}|s_m]I_m q(s_m) - \theta\Big\{ \sum_{m=1}^{M}\sum_{n=1}^{N} \log\big(p(v_n|s_m)\big)p(v_n|s_m)q(s_m) - \sum_{n=1}^{N} \log\big(p(v_n)\big)p(v_n)\Big\}, \tag{4.16}$$

subject to

$$\sum_{n=1}^{N} p(v_n|s_m) = 1, \quad \text{for } 1 \leqslant m \leqslant M, \tag{4.17}$$

$$\sum_{m=1}^{M} p(v_n|s_m)q(s_m) = p(v_n), \quad \text{for } 1 \leqslant n \leqslant N. \tag{4.18}$$

The problem (4.16) is the discrete analogue of (4.2). To see the connection between these two problems, let us fix a signal state s_m and let the number of signals go to infinite. In the limit, the probability of each signal, in particular $q(s_m)$, approaches zero. We then expect that

$$\lim_{M\to\infty} I_m = \lim_{M\to\infty} \frac{1}{q(s_m)} \int_{Q_{m-1}}^{Q_m} G^{-1}(u)du = G^{-1}(Q(s_m)). \tag{4.19}$$

Changing the summations over M in (4.16) to integrals, we obtain the objective (4.2) for the continuous signal case.

The following result summarizes the optimal signal structure when only M signal states are allowed.

Proposition 4.4. *The optimal posteriors for the signal state s_m with $q(s_m) > 0$ are given*

by

$$p(v_n|s_m) = \frac{e^{(v_n I_m+\mu_n)/\theta}}{\sum_{n'=1}^{N} e^{(v_{n'} I_m+\mu_{n'})/\theta}}, \quad 1 \leqslant n \leqslant N, 1 \leqslant m \leqslant M. \tag{4.20}$$

In the previous equation, the vector $(\mu_1, \ldots, \mu_N)$ *exists and is unique (up to an additive constant) such that the Bayes plausibility* (4.18) *is satisfied for given* $q(s_1), \ldots, q(s_N)$*. The optimal value is*

$$\max_{q(\cdot)} \Big\{ \theta \sum_{m=1}^{M} \log \Big(\sum_{n=1}^{N} e^{(v_n I_m+\mu_n)/\theta} \Big) q(s_m) + \theta \sum_{n=1}^{N} \log \big(p(v_n)\big) p(v_n) - \sum_{n=1}^{N} \mu_n p(v_n) \Big\}, \tag{4.21}$$

and there exists an optimal q*.*

The optimal posterior in (4.20) and the objective in (4.21) are the discrete analogue of (4.5) and (4.8) respectively.

Come back to our question whether continous signals are optimal for discretely distributed payoffs. The following result provides the answer.

Proposition 4.5. *When the payoff has a discrete distribution, continuous signals are optimal. The optimal posterior, optimal value, and the conditional expected payoff are presented in Proposition 4.2*

To understand the logic behind this result, recall the optimal value (4.21). the informed trader picks the signal marginal q to maximize the first term. The signal marginal is not only used to sum over different log terms, it also impacts $I_m = \frac{1}{q(s_m)} \int_{Q_{m-1}}^{Q_m} G^{-1}(du)$. For a signal state s_m, the informed trader can replace it by two new signal states $\underline{s_m}$ and $\overline{s_m}$ such that their associated probabilities satisfy $\underline{q_m} + \overline{q_m} = q(s_m)$, meanwhile all other signal states remain the same. Importantly, for given $(\mu_1, \ldots, \mu_N)$, the function $x \mapsto \log \big(\sum_{n=1}^{N} e^{(v_n x+\mu_n)/\theta} \big)$ is strictly convex. Therefore, it is strictly better off by splitting one signal states into two, i.e.,

$$\log \Big(\sum_{n=1}^{N} e^{(v_n I_m+\mu_n)/\theta} \Big) q(s_m) < \log \Big(\sum_{n=1}^{N} e^{(v_n \underline{I_m}+\mu_n)/\theta} \Big) \underline{q_m} + \log \Big(\sum_{n=1}^{N} e^{(v_n \overline{I_m}+\mu_n)/\theta} \Big) \overline{q_m},$$

where $\underline{I_m} := \frac{1}{\underline{q_m}} \int_{Q_{m-1}}^{Q_{m-1}+\underline{q_m}} G^{-1}(u)du$ and $\overline{I_m} := \frac{1}{\overline{q_m}} \int_{Q_m-\overline{q_m}}^{Q_m} G^{-1}(u)du$ satisfy

$$I_m = \underline{I_m} \frac{\underline{q_m}}{q_m} + \overline{I_m} \frac{\overline{q_m}}{q_m}.$$

Therefore, for given $(\mu_1, \ldots, \mu_N)$, splitting signals is welfare improving.

Consider Lagrangian multipliers $(\mu_1, \ldots, \mu_N)$ as dual variables, we obtain from the argument above that the dual value function improves as the number of signals increases. Finally, we use a duality argument to show that the primal value function of the prior p also weakly improves when the number of signals increases, leading to the optimality of continuously distributed signals.

When there is a finite number of possible actions, a recommendation lemma (Matějka and McKay, 2015, Lemma 1) shows that it is optimal to design a signal one-to-one mapping to each action. This is because information acquisition is costly, it is suboptimal to distinguish two signals which lead to the same action and payoff. Therefore, the number of signals is equal to the number of actions, and discrete signals are optimal. In our setting, the conditional expected payoff $\mathbb{E}[\tilde{v}|\tilde{s}]$ corresponds to actions. Each realization of $\mathbb{E}[\tilde{v}|\tilde{s}]$ corresponds to a different trading strategy in (2.10) and a different conditional profit. This matches the distribution of noise trades, which is also continuous. Therefore, the number of possible actions is infinite, and number of signals is infinite as well.

4.3 Examples

In this section, we present comparative statics for several examples of the conditional expected payoff distribution. We will illustrate how the trade-off between the profit potential and the information leakage cost interact with the information acquisition cost in determining the conditional expected payoff.

Let us start from the case with two payoff states and a symmetric prior, i.e., $N = 2$ and $p(v_1) = p(v_2) = 0.5$ with $v_1 < v_2$. Proposition 4.2 implies that it is without loss of generality to consider a normal signal $\tilde{z} \sim N(0, \sigma_Z^2 T)$. Meanwhile, Proposition A.2 in the appendix shows that $\mu_1 = \mu_2 = 0$. Therefore, (4.7) implies that as θ converges to zero, the posterior expected payoff $\mathbb{E}[\tilde{v}|\tilde{z}]$ converges to the prior. Specifically, $\lim_{\theta\to 0} \mathbb{P}\big(\mathbb{E}[\tilde{v}|\tilde{z} < 0] = v_1\big) = \lim_{\theta\to 0} \mathbb{P}\big(\mathbb{E}[\tilde{v}|\tilde{z} > 0] = v_2\big) = 0.5$. When the information acquisition cost vanishes, the posterior expected payoff is the low payoff state v_1 when the informed trader receives a negative signal and the high payoff state v_2 when the informed trader observes a positive signal. When θ converges to infinity, (4.7) implies that the posterior expected payoff converges to the prior expectation $(v_1 + v_2)/2$. When the information acquisition cost is infinite, the signal is not informative, hence the posterior expectation does not provide additional information beyond the prior mean.

For an intermediate value of θ, the density of the posterior expected payoff $\mathbb{E}[\tilde{v}|\tilde{z}]$ is given by (4.13) with $\mu_1 = \mu_2 = 0$. From this expression, we can see that $\frac{\theta}{\sigma_Z\sqrt{T}}$ is a sufficient statistic for the distribution of the expected payoff. Therefore, the information acquisition cost parameter has exactly the opposite effect as the noise trade volatility. When the information acquisition becomes less costly (θ decreases), it has the same effect on the posterior expected payoff as an increasing noise trade volatility. When the noise trade volatility increases, inference by market makers becomes less precise, the informed trader can better hide her informed trades among noise trades and the information leakage cost is reduced. Therefore, the informed trader can acquire a more precise signal to exploit the profit potential. The limit $\theta \to 0$ is equivalent to the limit $\sigma_Z\sqrt{T} \to \infty$. In such a limit, the noise trade is infinitely volatile, the informed trader acquires the perfect signal to maximize the variance of the posterior expected payoff, which indicates the highest profit potential.

Figure 1 Here

Figure 1 presents one example with two payoff states and a symmetric prior, but different information acquisition cost parameter θ. Panel (A) presents the densities of $\mathbb{E}[\tilde{v}|\tilde{z}]$ for different θ. They are all distributed symmetrically around zero, which is the prior expected payoff. Therefore, the skewness of $\mathbb{E}[\tilde{v}|\tilde{z}]$ is identically zero for different θ, as the black solid line in Panel (D) shows. However, when θ is small, information acquisition is cheap. The informed trader acquires a precise signal, the resulting distribution of $\mathbb{E}[\tilde{v}|\tilde{z}]$ has a large variance and is close to the prior distribution of $\tilde{v}$. When $\theta = 1$, Panel (A) shows that the density of $\mathbb{E}[\tilde{v}|\tilde{s}]$ peaks at -2 and 2, which are the value of v_1 and v_2, respectively. However, the distribution of $\mathbb{E}[\tilde{v}|\tilde{z}]$ is distinct from the noise trade distribution $N(0, \sigma_Z^2 T)$, generating a large Wasserstein-2 distance between the centered conditional payoff distribution F_s^c and the normal distribution of noise trades. Therefore, even though the informed trader benefits from a large profit potential from a precise signal and trades aggressively, she also faces high information leakage cost, because the market markers learn more from her aggressive trades. In the limit $\theta \to 0$, the profit potential dominates the information leakage cost. The informed trader acquires a prefect signal. The variance of $\mathbb{E}[\tilde{v}|\tilde{z}]$ is the variance of the prior payoff distribution, which is the largest possible variance of the conditional expected payoff.

As θ increases, information acquisition becomes increasingly costly, hence signal is less informative about payoffs. As Panel (B) of Figure 1 shows, the posterior probability of payoffs becomes less sensitive in signal, as θ increases. As a result, Panel (A) shows that the distribution of $\mathbb{E}[\tilde{v}|\tilde{z}]$ becomes more concentrated around the prior mean and

its variance decreases. As θ increases, Panel (C) shows that the Wasserstein-2 distance between F_s^c and G has a U-shape. The reason is the following. When θ is close to zero, the distribution of $\mathbb{E}[\tilde{v}|\tilde{s}]$ is close to the prior distribution of $\tilde{v}$ and far away from the distribution of noise trade Z_T. As θ increases, the distribution of $\mathbb{E}[\tilde{v}|\tilde{s}]$ increasingly resembles the distribution of Z_T, reducing their Wasserstein-2 distance. However, as θ continues to increase, difference between the variance of $\mathbb{E}[\tilde{v}|\tilde{s}]$ and Z_T increases, which becomes the dominating component in $W_2(F_s^c, G)$. When we exclude the difference of mean and standard deviation, i.e., the Gelbrich lower bound, the orange dash line in Panel (C) shows that $W_2^2(F_s^c, G) - \mathbb{E}[\tilde{v}]^2 - \left(\text{std}(\mathbb{E}[\tilde{v}|\tilde{s}]) - \sigma_Z\sqrt{T}\right)^2$ vanishes as θ increases. This implies that $\mathbb{E}[\tilde{v}|\tilde{s}]$ increasingly resembles a norm distribution with mean $\mathbb{E}[\tilde{v}]$ and standard deviation $\text{std}(\mathbb{E}[\tilde{v}|\tilde{s}])$. In particular, the orange dashed line in Panel (D) shows that the Kurtosis of $\mathbb{E}[\tilde{v}|\tilde{z}]$ converges to 3 (the Kurtosis of normal distributions) as θ increase. Therefore, the informed trader extracts less profit potential from less precise signal, meanwhile the information leakage also reduces due to less aggressive trades. In the limit $\theta \to \infty$, the informed trader no longer acquires signal, the conditional expected payoff $\mathbb{E}[\tilde{v}|\tilde{z}]$ converges to the prior expectation $\mathbb{E}[\tilde{v}]$, the information advantage disappears and the informed trader's value converges to zero.

Figure 2 Here

In the two-state symmetric prior case, $\frac{\theta}{\sigma_Z\sqrt{T}}$ is a sufficient statics for the distribution of the conditional expected payoff. Therefore, decreasing σ_Z has an equivalent impact as increasing θ. Figure 2 presents an example with different σ_Z. As σ_Z decreases, it becomes harder for the informed trader to hide her order among noise trades. Panel (A) shows that the conditional expected payoff becomes more concentrated around the prior mean. Both the Wasserstein-2 distance $W_2(F_s^c, G)$ and the version excluding Gelbrich lower bound vanish, as σ_Z approaches zero. The orange dashed line in Panel (D) shows that the Kurtosis of the conditional expected payoff also converges to 3. Meanwhile, as σ_Z decreases, Panel (B) shows that the posterior probability of states become less sensitive in signal, i.e., signal becomes less informative. Even if the informed trader has precise information on the payoff, she cannot trade on it aggressively due to severe information leakage in a thin market with low volume of noise trades. Hence, the informed trader optimally acquires low precision signal to save on the information acquisition cost.

Figure 3 Here

Figure 3 compares discrete and continuous signals. As we have seen from Proposition 4.5, as the number of signal state M increases, the optimal value improves. This is confirmed by Panel (A). Such improvement in the optimal value is an joint effect of changes in the variance of the conditional expected payoff, the squared Wasserstein-2 distance between F_s^c and G, and the information acquisition cost. As M increases, Panel (B) shows that the variance of $\mathbb{E}[\tilde{v}|\tilde{s}]$ reduces. This is because when M increases, there are more signal states with the posterior mean $\mathbb{E}[\tilde{v}|\tilde{s}]$ close to the prior mean $\mathbb{E}[\tilde{v}]$. In contrast, if there were only two signal states s_1 and s_2, the informed trader would spread out $\mathbb{E}[\tilde{v}|s_1]$ and $\mathbb{E}[\tilde{v}|s_2]$, in order to benefit from their associated profit potential. Hence, the informed trader would not choose signals such that either $\mathbb{E}[\tilde{v}|s_1]$ or $\mathbb{E}[\tilde{v}|s_2]$ is close to the prior mean $\mathbb{E}[\tilde{v}]$. As a result, the variance of $\mathbb{E}[\tilde{v}|\tilde{s}]$ is larger when $\tilde{s}$ has only two states. As M increases, the distribution of $\mathbb{E}[\tilde{v}|\tilde{s}]$ increasingly resemble a continuous distribution, hence $W_2^2(F_s^c, G)$ reduced, as Panel (C) shows. Meanwhile, when the informed trader chooses the optimal posterior and the signal marginal, Panel (D) shows that the information acquisition cost decreases with M when $M \geqslant 3$. When M increases, the informed trader redesign the posterior and the signal marginal, achieving a marginally lower information acquisition cost, even though the number of signals states is more.

The reduction in both costs, mainly from the reduction of the squared Wasserstein-2 distance, dominates the decrease in the variance of conditional expected payoff, leading to a higher insider's value as the number of signal states increases.

We present several comparative statics when the prior distribution is asymmetric in Appendix B. All comparative statics discussed above hold for the asymmetric prior case. Two new features emerge: i) due to the asymmetric prior, the distribution of $\mathbb{E}[\tilde{v}|\tilde{s}]$ is asymmetric as well. Therefore, μ_1 and μ_2 are nonzero, and (4.13) shows that $\frac{\theta}{\sigma_Z\sqrt{T}}$ is no longer a sufficient statics for the conditional expected payoff distribution; ii) the skewness of $\mathbb{E}[\tilde{v}|\tilde{s}]$ converges to zero (the skewness for normal distribution) as $\theta \to \infty$ or $\sigma_Z \to 0$.

5 Continuously distributed payoff

Having understand the optimal signal structure for discretely distributed payoffs, we consider continuously distributed payoffs in this section. We show that the main result for discretely distributed payoffs (Proposition 4.2) has a natural extension in the case with continuously distributed payoffs.

Consider a continuously distributed payoff $\tilde{v}$ with a density p. We assume that $\mathbb{E}[\tilde{v}^2]$

is finite so that the informed trader's value in (2.17) is finite.[8] We focus on continuously distributed signals such that support$(q) \ni s \mapsto \mathbb{E}[\tilde{v}|s]$ is strictly increasing.[9] Extending Lemma 4.1 to the current case, the informed trader's optimization problem is

$$\max_{p(\cdot|\cdot),q(\cdot)} \int \mathbb{E}[\tilde{v}|s] G^{-1}\big(Q(s)\big) q(s)\, ds - \theta \Big\{ \iint \log\big(p(v|s)\big) p(v|s) q(s)\, ds dv - \int \log\big(p(v)\big) p(v)\, dv \Big\}, \tag{5.1}$$

subject to

$$\int p(v|s)\, dv = 1, \quad \text{for any } s, \tag{5.2}$$

$$\int p(v|s) q(s)\, ds = p(v), \quad \text{for any } v. \tag{5.3}$$

Compared with (4.2) and its constraints, only the summations over different payoff states are replaced by integrals above.

The following result is a natural extension of Proposition 4.2.

Proposition 5.1. *Suppose that* $\mathbb{E}[\tilde{v}^2] < \infty$. *The following statements hold.*

(i) The optimal posterior probabilities are

$$p(v|s) = \frac{e^{\frac{1}{\theta}\left[vG^{-1}(Q(s))+\mu(v)\right]}}{\int e^{\frac{1}{\theta}\left[v'G^{-1}(Q(s))+\mu(v')\right]} dv'}, \quad s \in \textit{support}(q), v \in \textit{support}(p). \tag{5.4}$$

In the previous equation, the function $v \mapsto \mu(v)$ *exists and is unique up to additive constants.*

(ii) Any continuous signals $\tilde{s}$, *such that* $s \mapsto \mathbb{E}[\tilde{v}|s]$ *is strictly increasing, achieves the same optimal value*

$$\theta\, \mathbb{E}\Big[\log\Big(\int e^{(v\tilde{z}+\mu)/\theta} dv \Big) \Big] + \theta \int \log\big(p(v)\big) p(v)\, dv - \int \mu(v) p(v)\, dv, \tag{5.5}$$

where $\tilde{z}$ *is a random variable with normal distribution* $N(0, \sigma_Z^2 T)$.

[8] Jensen's inequality implies that $\mathbb{E}\big[\mathbb{E}[\tilde{v}|\tilde{s}]^2\big] \leqslant \mathbb{E}[\tilde{v}^2]$. Both the Wasserstein-2 distance and the information acquisition cost are nonnegative. Hence the value of (2.17) is bounded from above.

[9] Proposition 5.1 below show that the convexity structure of the informed trader's objective function with respect to signals remains the same when the payoff is continuously distributed. Therefore, the same statements of Proposition 4.5 hold for continuously distributed payoffs, indicating that continuous signals are optimal.

(iii) The conditional expected payoff $\mathbb{E}[\tilde{v}|\tilde{s}]$ *has the same distribution as*

$$\frac{\int v e^{(v\tilde{z}+\mu(v))/\theta} dv}{\int e^{(v\tilde{z}+\mu(v))/\theta} dv}. \tag{5.6}$$

In particular, any continuous signal $\tilde{s}$ *leads to the same expected payoff distribution.*

When there are infinite payoff states, the Lagrangian multiplier μ becomes a function of states. For a given signal marginal q and payoff a prior density p, the Sinkhorn problem becomes to find $v \mapsto \mu(v)$ so that

$$\int \xi(v)k(v,z)\zeta(z)\, dz = p(v), \qquad \text{for any v,}$$

where $\xi(v) = e^{\mu(v)/\theta}$, $k(v,z) = e^{vz/\theta}$, and $\zeta(z) = \frac{q(z)}{\int \xi(v)k(v,z)dv}$. The existence and uniqueness (up to additive constants) are obtained by Rüschendorf and Thomsen (1993).

Similar to the discretely distributed payoff case, any continuous signal yields the same optimal value. Therefore, without loss of generality, the informed trader can choose a normal signal $\tilde{z} \sim N(0, \sigma_Z^2 T)$, regardless of the payoff distribution. Therefore, restricting to a normally-distributed signal, such as Kyle (1984) and Back and Pedersen (1998), may not restrict the generality of their results.

Figures 4 and 5 present an example where the payoff follows a double exponential distribution with density $\rho(v) = e^{-v}$ for $v > 0$ and $\rho(v) = e^{v}$ for $v < 0$. The informed trader chooses the normal signal $\tilde{z} \sim N(0, \sigma_Z^2 T)$. The comparative statics of the expected payoff distribution with respect to θ and σ_Z are consistent with those for discretely distributed payoffs in the previous section.

Figure 4 Here

Figure 4 shows that as θ increases, the expected payoff distribution increasingly resembles a normal distribution, with the kurtosis decreases from 6 (the kurtosis for a double-exponential distribution) to 3 (the kurtosis for a normal distribution). As the information acquisition becomes more costly, optimal signal becomes less informative, profit potential decreases, and the informed trader focuses more on preventing information leakage, so that the expected payoff becomes more similar to a normal distribution, in order to protect the small informational advantage of the informed trader.

Figure 5 Here

Figure 5 illustrates the opposite impact of σ_Z. As σ_Z increase, the expected payoff increasingly resembles the prior distribution, with the kurtosis increases from 3 to 6. As the noise trades become more volatile, it becomes easier for informed trades to hide her trades among noise trades. The information leakage cost decreases, the informed trader acquires a more informative signal to increase the profit potential.

6 General information acquisition cost

In the previous sections, we work with the entropy cost of Sims (1998, 2003). Although much of the existing literature adopts the entropy cost to represent information acquisition costs, some of the behavioral implications of this specification are inconsistent with experimental evidence (see, e.g., Woodford (2012), Caplin and Dean (2013), and Dean and Neligh (2020, 2023)). In response to these findings, several studies propose more general formulations of information acquisition costs, including Caplin and Dean (2013), Caplin, Dean and Leahy (2022), Hébert and Woodford (2021), and Pomatto, Strack and Tamuz (2023).

In this section, we consider the uniformly posterior-separable (UPS) cost functions introduced by Caplin and Dean (2013); Caplin, Dean and Leahy (2022). Consider a concave function $H : [0,1] \to \mathbb{R} \cup -\infty$. We assume that $H \in C^1(0,1)$, H is strictly concave on $(0,1)$, and $H'(0+) = \infty$.[10] Given a prior distribution p and a signal structure, which includes the signal marginal $q(\cdot)$ and the posterior distribution $p(\cdot|\cdot)$, we consider the information acquisition cost

$$\theta I^H(\tilde{s}, \tilde{v}) := \theta\Big[\sum_v H(p(v)) - \sum_s q(s) \sum_v H(p(v|s))\Big]. \tag{6.1}$$

In this section, we will show that the qualitative results that we obtained in the previous section are robust for the general specification in (6.1): the continuous signal is optimal and any continuous signal, together with its associated posterior, leads to the same ex ante value for the informed trader.

We illustrate our results using the setting of Section 4. We start with a signal $\tilde{s}$ with a continuous distribution such that $s \mapsto \mathbb{E}[\tilde{v}|s]$ is strictly increasing. Reformulation of the informed trader's problem in Lemma 4.1 holds with the information acquisition cost replaced by (6.1):

[10] When the domain of H includes $[0,1]$, we consider its restriction on $[0,1]$. The condition $H'(0+) = \infty$ ensures that the optimal posterior is strictly positive.

$$\max_{p(\cdot|\cdot),q(\cdot)} \int \mathbb{E}[\tilde{v}|s] G^{-1}\big(Q(s)\big) q(s)\, ds - \theta\Big[\sum_{n=1}^{N} H(p(v_n)) - \sum_{n=1}^{N} \int H(p(v_n|s)) q(s)\, ds\Big], \tag{6.2}$$

subject to the constraints (4.3) and (4.4).

The transformation $G^{-1}\big(Q(s)\big)$ still presents in the objective function (6.2). Therefore, for any continuously distributed signal $\tilde{s}$ with strictly increasing CDF Q, the informed trader can transform the problem (6.2) to an equivalent formulation with the normal distributed signal $\tilde{z} = G^{-1}\big(Q(\tilde{s})\big)$. Therefore, all such continuously distributed signals lead to the same ex ante value for the informed trader, and the informed trader can choose the normal signal $\tilde{z}$ without loss of generality. Moreover, in the proof of the following result, we show that the informed trader's value increases with the number of signal states. Therefore, continuously distributed signals remain optimal.

Proposition 6.1. *Consider the information acquisition cost in* (6.1)*. The informed trader's optimization problem* (6.2) *with the constraints* (4.3) *and* (4.4) *admit a solution. Any continuous signal $\tilde{s}$, such that $s \mapsto \mathbb{E}[\tilde{v}|s]$ is strictly increasing, achieves the same optimal value. Moreover, continuous signals are optimal.*

7 Conclusion

We study a flexible information acquisition problem in Kyle (1985) model. The prior distribution of the asset payoff can be general and the signal choice is flexible; both can have discrete or continuous, including non-normal, distributions. The information acquisition cost is modelled by mutual information, following Sims (1998, 2003). We provide a complete characterization of the optimal information acquisition decision. We show that continuous signals are optimal regardless of the prior distribution of the asset payoff. Moreover, any continuous signals, which lead to a higher posterior expected payoff after a higher signal realization, lead to the same ex-ante value for the informed trader and the same distribution for the posterior expected payoff. Therefore, restricting to a normal signal is without loss of generality. The optimality of the continuous signals and the equivalence among continuous signals are robust to the choice of information acquisition cost. Both results hold when we consider more flexible UPS cost of Caplin and Dean (2013) and Caplin, Dean and Leahy (2022). We show that the entropy information acquisition cost mediates between two

tensions — the desire for high trading profits from a precise signal, which generates a high variance of the posterior expected payoff, and the desire to reduce the information leakage cost induced by market makers' inference.

We consider a static information acquisition problem in this paper. When the informed trader can acquire multiple pieces of information during the trading game, such as the setting of Banerjee and Breon-Drish (2020, 2022) or Han (2021), the incentive to acquire information interacts with future information acquisition and trading opportunities. For uniformly posterior-separable cost functions, a dynamic discrete choice problem is studied by Miao and Xing (2024). Integrating the forward-backward algorithm introduced therein with the dynamic flexible information acquisition problem in the Kyle model is another interesting future research topic.

References

Admati, Anat R., and Paul Pfleiderer. 1988. "A Theory of Intraday Patterns: Volume and Price Variability." *The Review of Financial Studies*, 1(1): 3–40.

Back, Kerry. 1992. "Insider trading incontinuous time." *Review of Financial Studies*, 5(3): 387–409.

Back, Kerry, and Hal Pedersen. 1998. "Long-lived information and intraday patterns." *Journal of Financial Markets*, 1(3-4): 385–402.

Back, Kerry, and Shmuel Baruch. 2004. "Information in Securities Markets: Kyle Meets Glosten and Milgrom." *Econometrica*, 72(2): 433–465.

Back, Kerry, Francois Cocquemas, Ibrahim Ekren, and Abraham Lioui. 2021. "Optimal transport and risk aversion in Kyle's model of infomred trading." Working paper, Rice University, Florida State University, EDHEC Business School.

Bagnoli, Mark, S. Viswanathan, and Craig Holden. 2001. "On the Existence of Linear Equilibria in Models of Market Making." *Mathematical Finance*, 11(1): 1–31.

Banerjee, Snehal, and Bradyn Breon-Drish. 2020. "Strategic trading and unobservable information acquisition." *Journal of Financial Economics*, 138(2): 458–482.

Banerjee, Snehal, and Bradyn Breon-Drish. 2022. "Dynamics of Research and Strategic Trading." *Review of Financial Studies*, 35(2): 908–961.

Breon-Drish, Bradyn. 2015. "On Existence and Uniqueness of Equilibrium in a Class of Noisy Rational Expectations Models." *The Review of Economic Studies*, 82(3): 868–921.

Caldentey, René, and Ennio Stacchetti. 2010. "Insider Trading With a Random Deadline." *Econometrica*, 72(1): 245–283.

Caplin, Andrew, and Mark Dean. 2013. "Behavioral Implications of Rational Inattention with Shannon Entropy." National Bureau of Economic Research Working Paper 19318.

Caplin, Andrew, Mark Dean, and John Leahy. 2019. "Rational Inattention, Optimal Consideration Sets, and Stochastic Choice." *The Review of Economic Studies*, 86(3): 1061–1094.

Caplin, Andrew, Mark Dean, and John Leahy. 2022. "Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy." *Journal of Political Economy*, 130(6): 1676–1715.

Çetin, Umut. 2025. "Insider trading with penalties in continuous time." *Journal of Economic Theory*.

Chabakauri, Georgy, Kathy Yuan, and Konstantinos E. Zachariadis. 2022. "Multi-asset Noisy Rational Expectations Equilibrium with Contingent Claims." *The Review of Economic Studies*, 89(5): 2445–2490.

Collin-Dufresne, Pierre, and Vyacheslav Fos. 2016. "Insider Trading, Stochastic Liquidity, and Equilibrium Prices." *Econometrica*, 84(4): 1441–1475.

Dean, Mark, and Nathaniel Neligh. 2020. "Estimating information cost functions in models of rational inattention." *Journal of Economic Theory*, 187: 1–32.

Dean, Mark, and Nathaniel Neligh. 2023. "Experimental tests of rational inattention." *Journal of Political Economy*, 131(12): 3415–3461.

Detemple, Jerome, Marcel Rindisbacher, and Scott Robertson. 2020. "Dynamic Noisy Rational Expectations Equilibrium With Insider Information." *Econometrica*, 88(6): 2697–2737.

Galichon, Alfred. 2016. *Optimal Transport Methods in Economics.* Princeton, NJ:Princeton University Press.

Gelbrich, Matthias. 1990. "On a Formula for the L2 Wasserstein Metric between Measures on Euclidean and Hilbert Spaces." *Mathematische Nachrichten*, 147: 185–203.

Grossman, Sanford J., and Joseph E. Stiglitz. 1980. "On the Impossibility of Informationally Efficient Markets." *The American Economic Review*, 70(3): 393–408.

Han, Brandon. 2021. "Dynamic information acquisition and asset prices." Working paper, Smith School, University of Maryland.

Hébert, Benjamin. 2018. "Moral Hazard and the Optimality of Debt." *The Review of Economic Studies*, 85(4): 2214–2252.

Hébert, Benjamin, and Michael Woodford. 2021. "Neighborhood-based information costs." *American Economic Review*, 111(10): 3225–3255.

Hellwig, Martin F. 1980. "On the Aggregation of Information in Competitive Markets." *Journal of Economic Theory*, 22(3): 477–498.

Justiniano, Jorge, Andreas Kleiner, Benny Moldovanu, Martin Rumpf, and Philipp Strack. 2025. "Entropy-Regularized Optimal Transport in Information Design." ACM.

Khorrami, Paymon, and João Thereze. 2025. "The Magic of Markets: Information Acquisition and Aggregation with Strategic Traders." Duke University Working Paper.

Kyle, Albert S. 1984. "Market Structure, Information, Futures Markets, and Price Formation." In *International Agricultural Trade: Advanced Readings in Price Formation, Market Structure, and Price Instability.* , ed. Gary G. Storey, Andrew Schmitz and Alexander H. Sarris, 45–64. Boulder, CO:Westview Press.

Kyle, ALbert S. 1985. "Continuous auctions and insider trading." *Econometrica*, 53(6): 1315–1335.

Lambert, Nicolas S., Michael Ostrovsky, and Mikhail Panov. 2018. "Strategic Trading in Informationally Complex Environments." *Econometrica*, 86(4): 1119–1157.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt. 2023. "Rational Inattention: A Review." *Journal of Economic Literature*, 61(1): 226–273.

Malamud, Semyon, and Andreas Schrimpf. 2021. "Persuasion by Dimension Reduction." arXiv Working Paper arXiv:2110.08884.

Matějka, Filip, and Alisdair McKay. 2015. "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model." *American Economic Review*, 105(1): 272–298.

Miao, Jianjun, and Hao Xing. 2024. "Dynamic discrete choice under rational inattention." *Economic Theory*, 77: 597–652.

Peyré, Gabriel, and Marco Cuturi. 2019. "Computational Optimal Transport." *Foundations and Trends in Machine Learning*, 11(5-6): 355–607.

Pomatto, Luciano, Philipp Strack, and Omer Tamuz. 2023. "The cost of information: the case of constant marginal costs." *American Economic Review*, 113(5): 1360–1393.

Rüschendorf, Ludger. 1995. "Convergence of the Iterative Proportional Fitting Procedure." *The Annals of Statistics*, 23(4): 1160–1174.

Rüschendorf, Ludger, and Wolfgang Thomsen. 1993. "Note on the Schrödinger equation and I-projections." *Statistics & Probability Letters*, 17(5): 369–375.

Sims, Christopher A. 1998. "Stickiness." *Carnegie-Rochester Conference Series on Public Policy*, 49(1): 317–356.

Sims, Christopher A. 2003. "Implications of rational inattention." *Journal of Monetary Economics*, 50: 665–690.

Sinkhorn, Richard. 1964. "A Relationship Between Arbitrary Positive Matrices and Doubly Stochastic Matrices." *The Annals of Mathematical Statistics*, 35(2): 876–879.

Sinkhorn, Richard, and Paul Knopp. 1967. "Concerning nonnegative matrices and doubly stochastic matrices." *Pacific Journal of Mathematics*, 21(2): 343–348.

Vives, Xavier. 2014. "On the Possibility of Informationally Efficient Markets." *Journal of the European Economic Association*, 12(5): 1200–1239.

Woodford, Michael. 2012. "Inattentive valuation and reference-dependent choice." Working Paper.

Yang, Ming. 2020. "Optimality of Debt under Flexible Information Acquisition." *The Review of Economic Studies*, 87(1): 487–536.

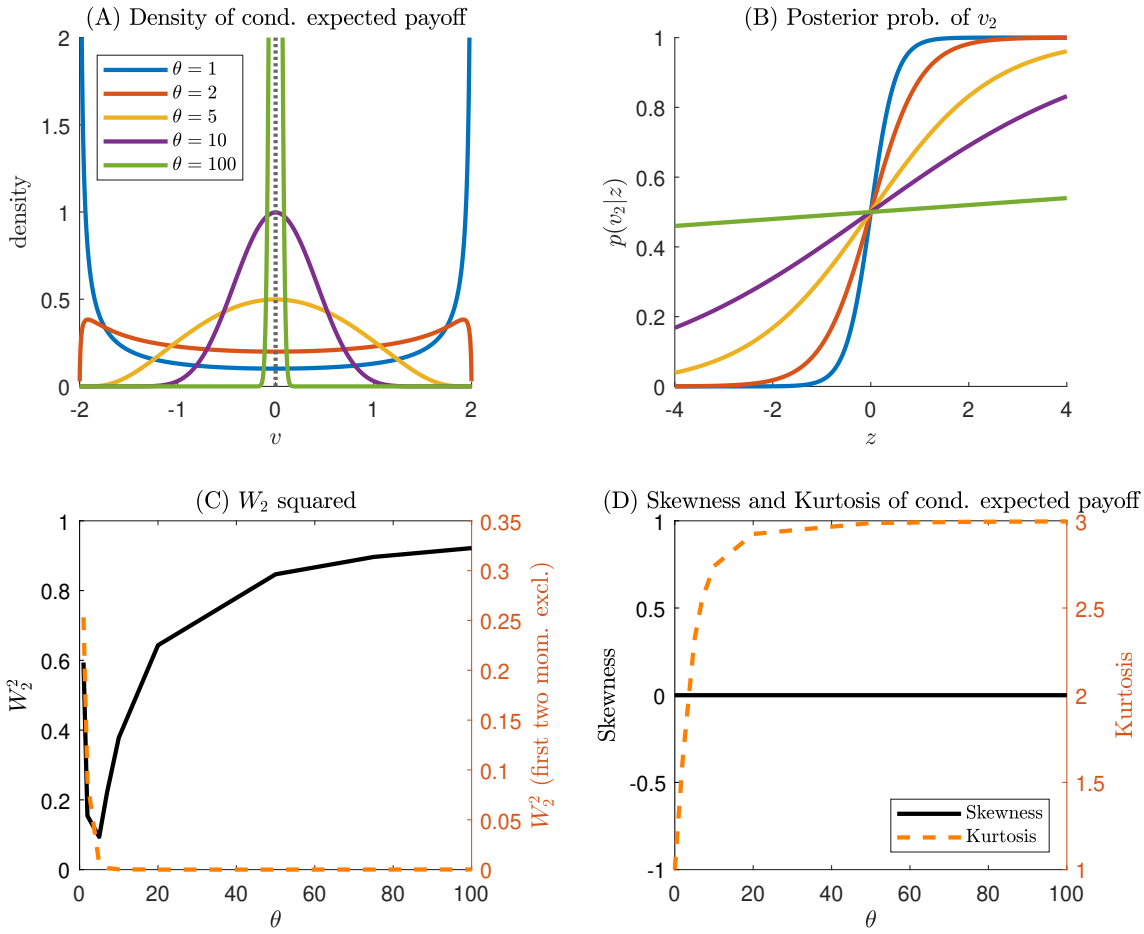

Figure 1: Two-state symmetric prior (different θ)

This figure presents the distribution of the conditional expected payoff, its skewness, Kurtosis, its W_2-distance with noise trade distribution, and the optimal posterior $p(v_2|z)$ for the two-state symmetric prior case. The signal $\tilde{z}$ has a normal distribution $N(0, \sigma_Z^2 T)$. Parameters: $N = 2$, $v_1 = -2, v_2 = 2$, $p(v_1) = p(v_2) = 0.5$, $\sigma_Z = 1$, and $T = 1$.

A Proofs

A.1 Proof of Lemma 2.2

Given that $\mathbb{E}[\tilde{v}|\tilde{s}]$ is measurable with respect to $\sigma(Y_T)$, the definition of conditional expectation, that it minimizes the mean squared error, implies

$$\mathbb{E}\big[\big(\tilde{v} - \mathbb{E}[\tilde{v}|\tilde{s}]\big)^2\big] \geqslant \mathbb{E}\big[\big(\tilde{v} - \mathbb{E}[\tilde{v}|\mathcal{F}_T^Y]\big)^2\big]. \tag{A.1}$$

Meanwhile, because $\tilde{v}$ and Z are independent, $\mathbb{E}\big[\tilde{v}|\mathcal{F}_T^Z \vee \sigma(\tilde{s})\big] = \mathbb{E}[\tilde{v}|\tilde{s}]$. Then $\mathcal{F}_T^Y \subset \mathcal{F}_T^Z \vee \sigma(\tilde{s})$ yields

$$\mathbb{E}\big[\big(\tilde{v} - \mathbb{E}[\tilde{v}|\mathcal{F}_T^Y]\big)^2\big] \geqslant \mathbb{E}\big[\big(\tilde{v} - \mathbb{E}[\tilde{v}|\mathcal{F}_T^Z \vee \sigma(\tilde{s})]\big)^2\big] = \mathbb{E}\big[\big(\tilde{v} - \mathbb{E}[\tilde{v}|\tilde{s}]\big)^2\big]. \tag{A.2}$$

Combining (A.1) and (A.2) yields

$$\mathbb{E}\big[\big(\tilde{v} - \mathbb{E}[\tilde{v}|\tilde{s}]\big)^2\big] = \mathbb{E}\big[\big(\tilde{v} - \mathbb{E}[\tilde{v}|\mathcal{F}_T^Y]\big)^2\big].$$

Due to the uniqueness of conditional expectation, we conclude $\mathbb{E}[\tilde{v}|\tilde{s}] = \mathbb{E}\big[\tilde{v}|\mathcal{F}_T^Y\big]$.

A.2 Proof of Proposition 2.4

Consider an equilibrium in Definition 2.1. Because the informed agent's trading strategy is measurable with respect to $\sigma(\tilde{s})$, we have seen from the discussion leading to (2.4) that

$$\mathbb{E}\Big[\int_0^T \big(\mathbb{E}[\tilde{v}] - P_t\big)\vartheta_t dt\Big] = \mathbb{E}\Big[\int_0^T \big(\mathbb{E}[\tilde{v}|\tilde{s}] - P_t\big)\vartheta_t dt\Big]. \tag{A.3}$$

Therefore, the informed trader's strategy $(p(\cdot|\cdot), q(\cdot))$ and ϑ_t also maximizes the objective in (2.4). Meanwhile, for a Markovian pricing rule, (Back, 1992, Lemma 2) shows that the optimal trading strategy satisfies

$$P_T = h(T, Y_T) = \mathbb{E}[\tilde{v}|\tilde{s}]. \tag{A.4}$$

Therefore,

$$P_t = \mathbb{E}\big[\tilde{v} \,|\, \mathcal{F}_t^Y\big] = \mathbb{E}\Big[\mathbb{E}\big[\tilde{v}|\mathcal{F}_T^Y\big]\,\Big|\,\mathcal{F}_t^Y\Big] = \mathbb{E}\Big[\mathbb{E}\big[\tilde{v}|\tilde{s}\big]\,\Big|\,\mathcal{F}_t^Y\Big],$$

where the second equality follows from the law of iterated conditional expectation, the third equality holds due to Lemma 2.2. Therefore, the pricing rule satisfies (2.9), and the equilibrium satisfies the requirement in Definition 2.3.

Conversely, consider an equilibrium in Definition 2.3. Because the informed agent's trading strategy is measurable with respect to $\sigma(\tilde{s})$, Equation (A.3) still holds. Hence the informed trader's strategy $(p(\cdot|\cdot), q(\cdot))$ and ϑ_t also maximizes the objective in (2.3). Meanwhile, (Back, 1992, Lemma 2) implies (A.4) for a Markovian pricing rule. Therefore, Lemma 2.2 implies that

$$P_t = \mathbb{E}\Big[\mathbb{E}\big[\tilde{v}|\tilde{s}\big]\,\Big|\,\mathcal{F}_t^Y\Big] = \mathbb{E}\Big[\mathbb{E}\big[\tilde{v}|\mathcal{F}_T^Y\big]\,\Big|\,\mathcal{F}_t^Y\Big] = \mathbb{E}\big[\tilde{v}\,|\,\mathcal{F}_t^Y\big].$$

Therefore, this equilibrium also satisfies all requirements in Definition 2.1.

A.3 Proof of Proposition 3.1

The first order condition of (3.2) in ξ is

$$\xi^{-\frac{1}{2}} = \frac{\theta}{\sigma_v\sigma_Z\sqrt{T}}\frac{1}{1-\xi}. \tag{A.5}$$

The left-hand side of the equation above decreases from infinity to zero and the right-hand side increases from θ to infinity, as ξ increases from 0 to 1. Therefore, there exists a unique solution $\xi^* \in (0,1)$ to this first order condition. The objective function in (3.2) is concave in ξ, because the second order derivative of the objective function is negative. Therefore, ξ^* is the unique maximizer.

When $\frac{\theta}{\sigma_v\sigma_Z\sqrt{T}}$ increases, the right-hand side of (A.5) increases for a fixed ξ. Therefore, ξ^* is smaller in order to satisfy the first-order condition (A.5).

Given ξ^*, σ_ϵ^2 satisfies $\xi^* = \frac{\sigma_v^2}{\sigma_v^2+\sigma_\epsilon^2}$. Hence, the optimal conditional precision of the signal satisfies (3.3). Because the right-hand side of (3.3) increases with ξ^*, the observation from the previous paragraph implies that the optimal signal precision $\frac{1}{\sigma_\epsilon^2}$ decreases with $\frac{\theta}{\sigma_v\sigma_Z\sqrt{T}}$.

When either θ increases or σ_v decreases, ξ^* decreases, hence $\lambda = \frac{\sqrt{\xi^*}\sigma_v}{\sigma_Z\sqrt{T}}$ decreases. For the monotonicity in σ_Z, we obtain from (A.5) that

$$\sigma_Z\sqrt{T}(\xi^*)^{-\frac{1}{2}} = \frac{\theta}{\sigma_v} + \sigma_Z\sqrt{T}(\xi^*)^{\frac{1}{2}}.$$

When σ_Z increases, ξ^* increases as well, hence the right-hand side of the previous equation

increases, indicating that $\lambda = \frac{\sqrt{\xi^*}\sigma_v}{\sigma_Z\sqrt{T}}$ decreases.

A.4 Proof of Proposition 3.2

We divide the proof into several steps. First, we add another constraint to the problem (2.17) that the variance of $\mathbb{E}[\tilde{v}|\tilde{s}]$ is a given constant K. We show that the optimal value of the problem (2.17) with the additional constraint is bounded from above by the optimal value of an alternative problem. Second, we show that a normal distribution $(p(\cdot|\cdot), q(\cdot))$ achieves the optimal value for this alternative problem, and it also achieves the optimal value for the original problem. Third, we allow the informed trader to vary K and finally confirm the statement in Proposition 3.2.

Step 1: Consider the problem (2.17) with the constraints (2.18), (2.19), and (2.20). We consider an additional constraint that the variance of $\mathbb{E}[\tilde{v}|\tilde{s}]$ is a given constant K. Equivalently, the second moment of $\mathbb{E}[\tilde{v}|\tilde{s}]$ is given by

$$\mathbb{E}\big[\mathbb{E}[\tilde{v}\,|\,\tilde{s}]^2\big] = \int \Big(\int v\, p(v|s) dv \Big)^2 q(s) ds = K + \mathbb{E}[\tilde{v}]^2. \tag{A.6}$$

We introduce an alternative problem whose optimal value provides an upper bound for the problem (2.17) with the constraints (2.18), (2.19), (2.20), and (A.6). To this end, recall that $\mathbb{E}[\tilde{v}|\tilde{s}]$ and Z_T are independent. For an arbitrary CDF F_s for $\mathbb{E}[\tilde{v}|\tilde{s}]$, its Wasserstein-2 distance admits the following Gelbrich (1990) lower bound :

$$W_2^2(F_s, G) \geqslant (\mu_G - \mathbb{E}[\tilde{v}])^2 + (\sigma_G - \sqrt{K})^2, \tag{A.7}$$

where $\mu_G = 0$ is the mean of Z_T and $\sigma_G = \sigma_Z\sqrt{T}$ is the standard deviation. Moreover, the lower bound on the right-hand side of (A.7) is attained when F_s is a normal CDF with mean $\mathbb{E}[\tilde{v}]$ and variance K. Let us now replace $W_2^2(F_s, G)$ in (2.17) by its lower bound and consider the following problem:

$$\max_{p(\cdot|\cdot), q(\cdot)} \frac{1}{2}\mathbb{E}\big[\mathbb{E}[\tilde{v}|\tilde{s}]^2\big] - \theta I(\tilde{s};\tilde{v}) + \frac{1}{2}\sigma_Z^2 T - \frac{1}{2}\mathbb{E}[\tilde{v}]^2 - \frac{1}{2}\big(\sigma_G - \sqrt{K}\big)^2, \tag{A.8}$$

subject to (2.18), (2.19), (2.20), and (A.6).

The value of problem (A.8) provides an upper bound to the value of problem (2.17). Moreover, if we show the value of (A.8) is attained by a normal distribution $(p^*(\cdot|\cdot), q^*(\cdot))$,

then F_s is a normal CDF and the Gelbrich lower bound in (A.7) is attained. Hence $(p^*(\cdot|\cdot), q^*(\cdot))$ also attains the optimal value of problem (2.17).

Step 2: We consider the problem (A.8) with the constraints (2.18), (2.19), (2.20), and (A.6). To this end, recall from (2.1) that

$$\begin{aligned}\theta I(\tilde{s};\tilde{v}) =&\theta\big[\mathbb{H}(p) - \int \mathbb{H}(p(\cdot|s))q(s)ds\big]\\ =&\theta \iint \log\big(p(v|s)\big)p(v|s)q(s)\,dvds - \theta\int \log(p(v))p(v)\,dv,\end{aligned}$$

Plugging the previous expression of $\theta I(\tilde{s};\tilde{v})$ into (A.8) and omitting quantities independent of $p(\cdot|\cdot)$ and $q(\cdot)$, we obtain the following problem:

$$\max_{p(\cdot|\cdot),q(\cdot)} \frac{1}{2}\int \Big(\int vp(v|s)dv\Big)^2 q(s)\,ds - \theta\iint \log\big(p(v|s)\big)p(v|s)q(s)\,dvds \tag{A.9}$$

subject to (2.18), (2.19), (2.20), and (A.6).

Introducing Lagrangian multiplier for the four constraints above, we obtain the following unconstrained problem

$$\begin{aligned}&\max_{p(\cdot|\cdot),q(\cdot)} \frac{1}{2}\int \Big(\int vp(v|s)dv\Big)^2 q(s)\,ds - \theta\iint \log\big(p(v|s)\big)p(v|s)q(s)\,dvds\\ &+\int \mu(v)\Big(\int p(v|s)q(s)ds - p(v)\Big)dv + \tau\Big(\int\Big(\int vp(v|s)dv\Big)^2 q(s)ds - K - \mathbb{E}[\tilde{v}]^2\Big)\\ &+\int \nu(s)\Big(\int p(v|s)dv - 1\Big)ds + \eta\Big(\int q(s)ds - 1\Big).\end{aligned} \tag{A.10}$$

The objective function above is linear in $q(s)$. Hence, for s with $q^*(s) \neq 0$, the following first-order condition must be satisfied at optimal

$$0 = \frac{1}{2}\Big(\int vp^*(v|s)dv\Big)^2 - \theta\int \log(p^*(v|s))p^*(v|s)dv + \int \mu(v)p^*(v|s)dv + \tau\Big(\int vp^*(v|s)dv\Big)^2 + \eta. \tag{A.11}$$

For s with $q^*(s) \neq 0$, the first-order condition in $p(v|s)$ yields

$$0 = (1+2\tau)v\int v'p^*(v'|s)dv' - \theta - \theta\log p^*(v|s) + \mu(v) + \frac{\nu(s)}{q(s)}. \tag{A.12}$$

Because the marginal value of $-\log\big(p(v|s)\big)p(v|s)$ is infinite at $p(v|s) = 0$, the optimizer

$p^*(v|s)$ must be nonzero except for some v's with measure zero. Therefore (A.12) must be satisfied for almost all v.

The equation (A.12) implies

$$p^*(v|s) = M(s)e^{(1+2\tau)v\,\mathbb{E}[\tilde{v}|\tilde{s}=s]/\theta+\mu(v)/\theta}, \tag{A.13}$$

for some function $M(s)$ which ensures the constraint $\int p^*(v|s)dv = 1$ to be satisfied for any s. Taking log on both sides of (A.13), multiplying both sides by $p^*(v|s)$, and integrating with respect to v, we obtain

$$\int \log(p^*(v|s))p^*(v|s)dv = \log M(s)\int p^*(v|s)dv + \frac{1+2\tau}{\theta}\big(\mathbb{E}[\tilde{v}|\tilde{s}=s]\big)^2 + \frac{1}{\theta}\int \mu(v)p^*(v|s)dv.$$

For the function $M(s)$ which ensures the constraint $\int p^*(v|s)dv = 1$ for any s, we obtain from the previous equation that

$$\theta\int \log(p^*(v|s))p^*(v|s)dv - \int \mu(v)p^*(v|s)dv = \theta\log M(s) + (1+2\tau)\mathbb{E}[\tilde{v}|\tilde{s}=s]^2.$$

Plugging the previous equation back into (A.11), we obtain

$$M(s) = e^{-(1/2+\tau)\mathbb{E}[\tilde{v}|\tilde{s}=s]^2/\theta+\eta/\theta}.$$

Combining the previous equation with (A.13), we obtain

$$p^*(v|s) = e^{\frac{1/2+\tau}{\theta}2v\mathbb{E}[\tilde{v}|\tilde{s}=s]-\frac{1/2+\tau}{\theta}\mathbb{E}[\tilde{v}|\tilde{s}=s]^2+\frac{\mu(v)+\eta}{\theta}} = e^{\frac{\mu(v)+\eta}{\theta}+\frac{1/2+\tau}{\theta}v^2}\,e^{-\frac{1/2+\tau}{\theta}\left(v-\mathbb{E}[\tilde{v}|\tilde{s}=s]\right)^2}. \tag{A.14}$$

Denote $\sigma_{v|s} := \sqrt{\frac{\theta}{1+2\tau}}$. Lemma A.1 below implies that $p^*(v|s)$ must be a normal density with mean $\mathbb{E}[\tilde{v}|\tilde{s}=s]$ and standard deviation $\sigma_{v|s}$. Hence $\mu(v)$ must satisfy

$$e^{\frac{\mu(v)-\eta}{\theta}+\frac{1/2+\tau}{\theta}v^2} = \frac{1}{\sqrt{2\pi}\sqrt{\frac{\theta}{1+2\tau}}}, \qquad \text{for any } v. \tag{A.15}$$

Given that both the posterior density $p^*(v|s)$ and the prior density $p(v)$ are normal, the marginal density $q^*(s)$ must be chosen to satisfy the Bayes constraint $\int p^*(v|s)q^*(s)ds = p(v)$ and optimize the objective function in (A.10). Plugging the densities for $p^*(v|s)$ and $p(v)$,

the Bayes constraint is rewritten as

$$\int \frac{1}{\sqrt{2\pi}\sigma_{v|s}} e^{-\frac{(v-\mathbb{E}[\tilde{v}|\tilde{s}=s])^2}{2\sigma^2_{v|s}}} q^*(s)ds = \frac{1}{\sqrt{2\pi}\sigma_v} e^{-\frac{(v-\mathbb{E}[\tilde{v}])^2}{2\sigma^2_v}}, \quad \text{for any } v. \tag{A.16}$$

This constraint can be satisfied by choosing a normal density $\widetilde{q}^*$ with mean $\mathbb{E}[\tilde{v}]$ and variance σ_v^2/ξ, with $\xi = 1 - \frac{\sigma^2_{v|s}}{\sigma^2_v}$. Meanwhile, note that the objective function in (A.10) is linear in $q(s)$, hence (A.11) implies that $\widetilde{q}^*$ does not change the value of the objective function. This choice of $\tilde{q}^*$ also ensures that $\tilde{q}^*(s) \neq 0$ for any s, verifying a condition used in (A.12). Therefore, $(p^*(\cdot|\cdot), \widetilde{q}^*(\cdot))$ is an optimizer for (A.10), and both $p^*(\cdot|\cdot)$ and $\widetilde{q}^*(\cdot)$ are normal.

To complete this step, we note that both $\sigma_{v|s}$ and ξ are determined by τ, which is the Langrangian multiplier associated to the constraint (A.6). For a given τ, the distribution $(p^*(\cdot|\cdot), \tilde{q}^*(\cdot))$ can be implemented by introducing a single

$$\tilde{s} = \tilde{v} + \epsilon, \tag{A.17}$$

where $\epsilon \sim N(0, \sigma_\epsilon^2)$ with $\sigma_\epsilon^2 = \sigma_v^2(\frac{1}{\xi} - 1)$. It follows from (A.17) that $\mathbb{E}[\tilde{v}|\tilde{s} = s] = \mathbb{E}[\tilde{v}] + \xi(s - \mathbb{E}[\tilde{v}])$. Therefore,

$$K = \mathrm{Var}\big(\mathbb{E}[\tilde{v}|\tilde{s}]\big) = \xi^2 \mathrm{Var}(\tilde{s}) = \xi\sigma_v^2.$$

When τ increases, $\sigma_{v|s}$ decreases and ξ increases. Hence, there is an 1-to-1 monotone increasing relation between τ and K.

Step 3: As the informed trader varies K to maximize the objective in (A.8), there is a unique associated τ and the optimal $(p^*(\cdot|\cdot), \tilde{q}^*(\cdot))$ are both normal and their variances are determined by τ, or equivalently, by ξ. As we have seen in the proof of Proposition 3.1, there exists a unique ξ^* which maximizes the objective in (3.2). Meanwhile, the associated optimal $(p^*(\cdot|\cdot), \tilde{q}^*(\cdot))$ is normal, hence the Gelbrich lower bound is attained, hence $(p^*(\cdot|\cdot), \tilde{q}^*(\cdot))$ is also an optimizer for the problem (2.17) with the constraints (2.18), (2.19), and (2.20).

A.5 An auxiliary lemma for the proof of Proposition 3.2

Lemma A.1. *Suppose that* $\frac{d}{ds}\mathbb{E}[\tilde{v}|\tilde{s} = s] \neq 0$ *for any* s *with* $q(s) \neq 0$. *Then the density* $p(v|s)$, *which satisfies*

$$p(v|s) = e^{\frac{\mu(v)+\eta}{\theta} + \frac{1/2+\tau}{\theta}v^2} e^{-\frac{1/2+\tau}{\theta}\left(v - \mathbb{E}[\tilde{v}|\tilde{s}=s]\right)^2}, \quad \textit{for any } v \textit{ and } s, \tag{A.18}$$

must be a normal density with mean $\mathbb{E}[\tilde{v}|\tilde{s}=s]$ *and standard deviation* $\sigma_{v|s}=\sqrt{\frac{\theta}{1+2\tau}}$.

Proof. It follows from the definition of conditional expectation that $\int vp(v|s)dv=\mathbb{E}[\tilde{v}|\tilde{s}=s]$. Plugging the expression of $p(v|s)$ from (A.18) into the previous equation and taking derivative with respect to s on both sides, we obtain

$$\int vp(v|s)\frac{v-\mathbb{E}[\tilde{v}|s]}{\sigma^2_{v|s}}\frac{d}{ds}\mathbb{E}[\tilde{v}|\tilde{s}=s]\,dv=\frac{d}{ds}\mathbb{E}[\tilde{v}|\tilde{s}=s].$$

Because $\frac{d}{ds}\mathbb{E}[\tilde{v}|\tilde{s}=s]\neq 0$, the previous equation and $\int vp(v|s)dv=\mathbb{E}[\tilde{v}|\tilde{s}=s]$ imply that

$$\int\big(v-\mathbb{E}[\tilde{v}|\tilde{s}=s]\big)^2p(v|s)dv=\sigma^2_{v|s}. \tag{A.19}$$

Taking derivatives with respect to s on both sides of (A.19), because $\sigma^2_{v|s}=\frac{\theta}{2+2\tau}$ is independent of s, we obtain

$$-2\frac{d}{ds}\mathbb{E}[\tilde{v}|\tilde{s}=s]\int\big(v-\mathbb{E}[\tilde{v}\,|\,s]\big)p(v|s)dv+\frac{d}{ds}\mathbb{E}[\tilde{v}|\tilde{s}=s]\int\big(v-\mathbb{E}[\tilde{v}|\tilde{s}=s]\big)^3\frac{1}{\sigma^2_{v|s}}p(v|s)dv=0.$$

This previous equation, together with $\int vp(v|s)dv=\mathbb{E}[\tilde{v}|\tilde{s}=s]$ and $\frac{d}{ds}\mathbb{E}[\tilde{v}|\tilde{s}=s]\neq 0$, implies that

$$\int\big(v-\mathbb{E}[\tilde{v}|\tilde{s}=s]\big)^3p(v|s)dv=0. \tag{A.20}$$

Taking the derivative with respect to s to both sides of the equation above and using the similar argument as before, we obtain

$$\int\big(v-\mathbb{E}[\tilde{v}|\tilde{s}=s]\big)^4p(v|s)dv=3\sigma^2_{v|s}. \tag{A.21}$$

Using induction and similar argument as before, we prove that

$$\int\big(v-\mathbb{E}[\tilde{v}|\tilde{s}=s]\big)^{2n-1}p(v|s)dv=0\quad\text{and}\quad\int\big(v-\mathbb{E}[\tilde{v}|\tilde{s}=s]\big)^{2n}p(v|s)dv=\sigma^{2n}_{v|s}\prod_{k=1}^{n}(2k-1),$$

for any positive integer n. These are exactly the central moments of a normal distribution with mean $\mathbb{E}[\tilde{v}|\tilde{s}=s]$ and variance $\sigma^2_{v|s}$. The statement follows, because moments uniquely identify a distribution. □

A.6 Proof of Lemma 4.1

We claim that

$$F_s^{-1}\big(Q(s)\big) = \mathbb{E}[\tilde{v}|s], \quad \text{for any } s. \tag{A.22}$$

To see this, suppose that $Q(s) = F_s(v)$ for some v. Then

$$Q(s) = F_s(v) = \mathbb{P}\big(\mathbb{E}[\tilde{v}|\tilde{s}] \leqslant v\big). \tag{A.23}$$

Meanwhile,

$$Q(s) = \mathbb{P}(\tilde{s} \leqslant s) = \mathbb{P}\big(\mathbb{E}[\tilde{v}|\tilde{s}] \leqslant \mathbb{E}[\tilde{v}|s]\big), \tag{A.24}$$

where the second equality holds because $s \mapsto \mathbb{E}[\tilde{v}|s]$ is strictly increasing. Comparing (A.23) and (A.24), we obtain $v = \mathbb{E}[\tilde{v}|s]$, and hence (A.22) is confirmed.

Plug (A.22) into (4.1) and use change-of-variable $u = Q(s)$,

$$\int_0^1 F_s^{-1}(u)G^{-1}(u)\,du = \int F_s^{-1}\big(Q(s)\big)G^{-1}\big(Q(s)\big)q(s)\,ds = \int \mathbb{E}[\tilde{v}|s]G^{-1}\big(Q(s)\big)q(s)\,ds.$$

The information acquisition cost is

$$\theta I(\tilde{s};\tilde{v}) = \theta\Big\{\sum_{n=1}^{N}\int \log\big(p(v_n|s)\big)p(v_n|s)q(s)\,ds - \sum_{n=1}^{N}\log\big(p(v_n)\big)p(v_n)\Big\}.$$

Combining the previous two equations, we obtain the objective function in (4.2). The constraint (4.3) ensures that the posterior $p(\cdot|s)$ is a probability for any s and the constraint (4.4) is the Bayes plausibility.

A.7 Proof of Proposition 4.2

Statements in (i): Introduce the Lagrangian multiplier μ_n and $\nu(s)$, $1 \leqslant n \leqslant N, s \in \text{support}(q)$, for the constraints (4.3) and (4.4). We obtain the following unconstrained

problem for (4.2):

$$\min_{\mu,\nu}\max_{p(\cdot|\cdot),q(\cdot)}\int \mathbb{E}[\tilde{v}|s]G^{-1}\big(Q(s)\big)q(s)ds$$
$$-\theta\Big\{\int\sum_{n=1}^{N}\log\big(p(v_n|s)\big)p(v_n|s)q(s)\,ds-\sum_{n=1}^{N}\log\big(p(v_n)\big)p(v_n)\Big\}$$
$$+\int\nu(s)\big[\sum_{n=1}^{N}p(v_n|s)-1\big]\,ds+\sum_{n=1}^{N}\mu_n\big[\int p(v_n|s)q(s)ds-p(v_n)\big]. \tag{A.25}$$

Recall that $\mathbb{E}[\tilde{v}|s]=\sum_{n=1}^{N}v_np(v_n|s)$. When $q(s)\neq 0$, the first-order condition of $p(v_n|s)$ in (A.25) yields

$$v_nG^{-1}\big(Q(s)\big)q(s)-\theta q(s)-\theta\log\big(p(v_n|s)\big)q(s)+\nu(s)+\mu_nq(s)=0.$$

This implies that

$$p(v_n|s)=C(s)\,e^{\frac{1}{\theta}(v_nG^{-1}(Q(s))+\mu_n)},$$

for some factor $C(s)$. The constraint $\sum_{n=1}^{N}p(v_n|s)=1$ pins down the factor $C(s)$ and leading to the expression

$$p(v_n|s)=\frac{e^{\frac{1}{\theta}(v_nG^{-1}(Q(s))+\mu_n)}}{\sum_{n'=1}^{N}e^{\frac{1}{\theta}(v_{n'}G^{-1}(Q(s))+\mu_{n'})}}. \tag{A.26}$$

Given that the objective function in (A.25) is concave in $p(v_n|s)$ due to the entropy cost, $p(v_n|s)$ is also the maximizer for given $q(s)$ and $(\mu_1,\ldots,\mu_N)$.

Given $p(v_n|s)$ and $q(s)$ for $1\leqslant n\leqslant N$ and $s\in\text{support}(q)$, the existence and uniqueness (up to an additive constant) of $(\mu_1,\ldots,\mu_N)$ has been discussed in the text. Such choice of μ's ensures the Bayes plausibility $\int p(v_n|s)q(s)\,ds=p(v_n)$ for any n.

Plugging (A.26) into (A.25), we obtain the following objective function for q:

$$\theta\int\log\Big(\sum_{n=1}^{N}e^{\frac{1}{\theta}\left(v_nG^{-1}(Q(s))+\mu_n\right)}\Big)q(s)\,ds+\theta\sum_{n=1}^{N}\log\big(p(v_n)\big)p(v_n)-\sum_{n=1}^{N}\mu_np(v_n). \tag{A.27}$$

Statements in (ii): For any continuously distributed $\tilde{s}$ with the cumulative distribution function Q, $Q(\tilde{s})$ is uniformly distributed on $[0,1]$. Hence $G^{-1}(Q(\tilde{s}))$ has a normal distribution

$N(0, \sigma_Z\sqrt{T})$. Therefore,

$$\begin{aligned}\int \log\Big(\sum_{n=1}^{N} e^{\frac{1}{\theta}\left(v_n G^{-1}(Q(s))+\mu_n\right)}\Big) q(s)\, ds =& \mathbb{E}\Big[\log\Big(\sum_{n=1}^{N} e^{\frac{1}{\theta}\left(v_n G^{-1}(Q(\tilde{s}))+\mu_n\right)}\Big)\Big] \\ =& \mathbb{E}\Big[\log\Big(\sum_{n=1}^{N} e^{\frac{1}{\theta}\left(v_n \tilde{z}+\mu_n\right)}\Big)\Big],\end{aligned} \tag{A.28}$$

where $\tilde{z}$ is a normal random variable with distribution $N(0, \sigma_Z^2 T)$. Plugging (A.28) into (A.27), we obtain (4.6). For the constraint (4.4), it can be stated equivalently as

$$\begin{aligned}p(v_n) =& \int p(v_n|s)q(s)ds = \int \frac{e^{\frac{1}{\theta}(v_n G^{-1}(Q(s))+\mu_n)}}{\sum_{n'=1}^{N} e^{\frac{1}{\theta}(v_{n'} G^{-1}(Q(s))+\mu_{n'})}} q(s)\, ds \\ =& \mathbb{E}\Big[\frac{e^{\frac{1}{\theta}(v_n G^{-1}(Q(\tilde{s}))+\mu_n)}}{\sum_{n'=1}^{N} e^{\frac{1}{\theta}(v_{n'} G^{-1}(Q(\tilde{s}))+\mu_{n'})}}\Big] = \mathbb{E}\Big[\frac{e^{\frac{1}{\theta}(v_n \tilde{z}+\mu_n)}}{\sum_{n'=1}^{N} e^{\frac{1}{\theta}(v_{n'} \tilde{z}+\mu_{n'})}}\Big].\end{aligned} \tag{A.29}$$

The previous identities show that the constraint (4.4) is the same for any continuously distributed $\tilde{s}$. Therefore, the constraint (4.4) is satisfied by choosing $(\mu_1, \ldots, \mu_N)$ as the solution of the Sinkhorn problem associated with (A.29), which exists and is unique (up to additive constants); see Rüschendorf and Thomsen (1993). Finally, note that the value in (A.29) does not change if $(\mu_1, \ldots, \mu_N)$ is replaced by $(\mu_1 + c, \ldots, \mu_N + c)$ for any c. The statement in (ii) is confirmed.

Statements in (iii): It follows from (A.26) that

$$\mathbb{E}[\tilde{v}|\tilde{s}] = \sum_{n=1}^{N} v_n p(v_n|\tilde{s}) = \frac{\sum_{n=1}^{N} v_n e^{\frac{1}{\theta}(v_n G^{-1}(Q(\tilde{s}))+\mu_n)}}{\sum_{n=1}^{N} e^{\frac{1}{\theta}(v_n G^{-1}(Q(\tilde{s}))+\mu_n)}} \sim \frac{\sum_{n=1}^{N} v_n e^{\frac{1}{\theta}(v_n \tilde{z}+\mu_n)}}{\sum_{n=1}^{N} e^{\frac{1}{\theta}(v_n \tilde{z}+\mu_n)}}.$$

In the equation above $\sim$ means that two random variables share the same distribution. The equal distribution above holds due to the fact that $G^{-1}(Q(\tilde{s}))$ and $\tilde{z}$ share the same distribution. Hence, any continuous signal leads to the same expected payoff distribution. In particular, the informed trader can choose $\tilde{z}$ as the signal.

Finally, we verify that $z \mapsto \mathbb{E}[\tilde{v}|z]$ is strictly increasing. To this end, denote $w_n := \frac{e^{(v_n z+\mu_n)/\theta}}{\sum_{n=1}^{N} e^{(v_n z+\mu_n)/\theta}}$. Note that $\sum_{n=1}^{N} w_n = 1$ and $\mathbb{E}[\tilde{v}|z] = \sum_{n=1}^{N} v_n w_n$. Then

$$\frac{d}{dz}\mathbb{E}[\tilde{v}|z] = \frac{1}{\theta}\Big\{\sum_{n=1}^{n} v_n^2 w_n - \Big(\sum_{n=1}^{N} v_n w_n\Big)^2\Big\} > 0,$$

where the inequality follows from Cauchy-Schwarz inequality.

A.8 Proof of Corollary 4.3

In this case, $N = 2$. For any $v \in (v_1, v_2)$, we compute the CDF for $\mathbb{E}[\tilde{v}|\tilde{z}]$:

$$\begin{aligned} F_s(v) =& \mathbb{P}\big(\mathbb{E}[\tilde{v}|\tilde{z}] \leqslant v\big) = \mathbb{P}\Big(\sum_{n=1}^{2} v_n e^{(v_n\tilde{z}+\mu_n)/\theta} \leqslant v \sum_{n=1}^{2} e^{(v_n\tilde{z}+\mu_n)/\theta}\Big) \\ =& \mathbb{P}\Big((v_1 - v)e^{(v_1\tilde{z}+\mu_1)/\theta} \leqslant (v - v_2)e^{(v_2\tilde{z}+\mu_2)/\theta}\Big) = \mathbb{P}\left(\tilde{z} \leqslant \frac{\theta}{v_1 - v_2}\log\left(\frac{v_2 - v}{v - v_1}\right) - \frac{\mu_1 - \mu_2}{v_1 - v_2}\right) \\ =& \mathbb{P}\left(\frac{\tilde{z}}{\sigma_Z\sqrt{T}} \leqslant \frac{\theta}{\sigma_Z\sqrt{T}(v_1 - v_2)}\log\left(\frac{v_2 - v}{v - v_1}\right) - \frac{\mu_1 - \mu_2}{\sigma_Z\sqrt{T}(v_1 - v_2)}\right) \end{aligned}$$

where the second equality follows from the form of $\mathbb{E}[\tilde{v}|\tilde{z}]$ in (4.7) and $\frac{\tilde{z}}{\sigma_Z\sqrt{T}}$ has a standard normal distribution. Taking the derivative with respect to v in $F_s(v)$, we obtain the density in (4.13).

A.9 Proof of Proposition 4.4

Introduce the Lagrangian multiplier ν_m and μ_n, $1 \leqslant n \leqslant N, 1 \leqslant m \leqslant M$, for the constraints (4.17) and (4.18). We obtain the following unconstrained problem for (4.16):

$$\begin{aligned} \min_{\mu,\nu} \max_{p(\cdot|\cdot),q(\cdot)} & \sum_{m=1}^{M} \mathbb{E}[\tilde{v} \,|\, s_m] I_m q(s_m) \\ & - \theta\Big\{ \sum_{m=1}^{M}\sum_{n=1}^{N} \log\big(p(v_n \,|\, s_m)\big) p(v_n \,|\, s_m) q(s_m) - \sum_{m=1}^{M} \log\big(p(v_n)\big) p(v_n)\Big\} \\ & + \sum_{m=1}^{M} \nu_m \Big[\sum_{n=1}^{N} p(v_n \,|\, s_m) - 1\Big] + \sum_{n=1}^{N} \mu_n \Big[\sum_{m=1}^{M} p(v_n \,|\, s_m) q(s_m) - p(v_n)\Big]. \end{aligned} \tag{A.30}$$

Recall that $\mathbb{E}[\tilde{v} \,|\, s_m] = \sum_{n=1}^{N} v_n p(v_n \,|\, s_m)$. When $q(s_m) \neq 0$, the first-order condition of $p(v_n \,|\, s_m)$ in (A.30) yields

$$v_n I_m q(s_m) - \theta q(s_m) - \theta \log\big(p(v_n \,|\, s_m)\big) q(s_m) + \nu_m + \mu_n q(s_m) = 0.$$

This implies that

$$p(v_n \,|\, s_m) = C_m e^{(v_n I_m + \mu_n)/\theta},$$

for some factor C_m. The constraint $\sum_{n=1}^{N} p(v_n \mid s_m) = 1$ pins down the factor C_m and leading to the expression

$$p(v_n \mid s_m) = \frac{e^{(v_n I_m + \mu_n)/\theta}}{\sum_{n'=1}^{N} e^{(v_{n'} I_m + \mu_{n'})/\theta}}. \tag{A.31}$$

Given that the objective function in (A.30) is concave in $p(v_n \mid s_m)$ due to the entropy cost, $p(v_n \mid s_m)$ is also the maximizer for given q's and μ's.

Given $p(v_n \mid s_m)$ and $q(s_m)$ for $1 \leqslant n \leqslant N, 1 \leqslant m \leqslant M$, the existence and uniqueness (up to an additive constant) of $(\mu_1, \ldots, \mu_N)$ is the consequence of the (rectangular) Sinkhorn-Knopp problem. Here rectangular means that $M \neq N$. This is the discrete analogue of the (continuous) Sinkhorn-Knopp problem in Proposition 4.2. Such choice of μ's ensures the Bayes plausibility $\sum_{m=1}^{M} p(v_n \mid s_m) q(s_m) = p(v_n)$ for any n.

Plugging (A.31) into (A.30), we obtain the objective function for q's in (4.21). Note that for signal state s_m with $q(s_m) = 0$, it does not contribute to the objective function in (4.21). Finally, we show that there exists a maximizer of the objective function (A.30) on the simplex of $\mathbb{R}^M$. To this end, note that $\mathbb{E}[\tilde{v}|s_m]$ is bounded from above by the maximum value of $\{v_1, \ldots, v_N\}$. Meanwhile,

$$\left|I_m q_m\right| = \left| \int_{Q_{m-1}}^{Q_m} G^{-1}(d) du \right| \leqslant \int_0^{\infty} G^{-1}(u) du = \frac{\sigma_Z \sqrt{T}}{\sqrt{2\pi}},$$

where the last equality holds because G is the CDF of $N(0, \sigma_Z^2 T)$. Meanwhile, the information acquisition cost in (4.16) is nonnegative. Therefore, the objective function in (4.16) is uniformly bounded from above. Consequently, the objective function (4.21) is bounded from above on the simplex of $\mathbb{R}^M$. The objective function is also continuous. Hence there exists an optimizer.

A.10 Proof of Proposition 4.5

If the signal has M different signal states, we denote the value in (A.30) as $V^M(p)$, where $p = (p(v_1), \ldots, p(v_N))$ is the prior payoff distribution. Equation (A.30) implies that the

value function $V^M(p)$ can be rewritten as

$$V^M(p) = \min_{\mu} \Big\{ \widehat{V}^M(\mu) - \sum_{n=1}^{N} \mu_n p(v_n) \Big\}, \quad \text{where } \mu = (\mu_1, \dots, \mu_N) \text{ and} \tag{A.32}$$

$$\begin{aligned} \widehat{V}^M(\mu) := \min_{\nu} \max_{p(\cdot|\cdot), q(\cdot)} & \sum_{m=1}^{M} \mathbb{E}[\tilde{v} \,|\, s_m] I_m q(s_m) \\ & - \theta \Big\{ \sum_{m=1}^{M} \sum_{n=1}^{N} \log\big(p(v_n \,|\, s_m)\big) p(v_n \,|\, s_m) q(s_m) - \sum_{n=1}^{N} \log\big(p(v_n)\big) p(v_n) \Big\} \\ & + \sum_{m=1}^{M} \nu_m \Big[\sum_{n=1}^{N} p(v_n \,|\, s_m) - 1 \Big] + \sum_{n=1}^{N} \mu_n \sum_{m=1}^{M} p(v_n \,|\, s_m) q(s_m). \end{aligned} \tag{A.33}$$

For each fixed μ, we claim that

$$\widehat{V}^M(\mu) < \widehat{V}^{M+1}(\mu). \tag{A.34}$$

Then, subtract $\sum_{n=1}^{N} \mu_n p(v_n)$ on both sides,

$$\widehat{V}^M(\mu) - \sum_{n=1}^{N} \mu_n p(v_n) < \widehat{V}^{M+1}(\mu) - \sum_{n=1}^{N} \mu_n p(v_n).$$

Taking infimum over μ first on the left-hand side, then on the right-hand side, we obtain that

$$V^M(p) \leqslant V^{M+1}(p). \tag{A.35}$$

It then remains to prove the claim (A.34). To this end, it follows from Proposition 4.4 that

$$\widehat{V}^M(\mu) = \max_{q(\cdot)} \Big\{ \theta \sum_{m=1}^{M} \log\Big(\sum_{n=1}^{N} e^{(v_n I_m + \mu_n)/\theta} \Big) q(s_m) \Big\} + \theta \sum_{n=1}^{N} \log\big(p(v_n)\big) p(v_n).$$

Similar argument as in the proof of Proposition 4.4 shows that the objective function in q is uniformly bounded from above on the simplex of $\mathbb{R}^M$. Meanwhile, this objective function is continuous. Therefore, an optimizer exists and we call it q^M.

Next, we are going to construct a new signal with $M+1$ signal states which achieves strictly higher value than q^M. To achieve this goal, let us consider a signal state s_m with the associated probability $q^M(s_m)$. Split the probability $q^M(s_m)$ into $\overline{q_m}$ and $\underline{q_m}$ such that $\overline{q_m} + \underline{q_m} = q^M(s_m)$. Replace the signal state s_m by two new signal states $\overline{s_m}$ and $\underline{s_m}$,

meanwhile, keep all other signal states and associated probabilities the same. Given that all $q^M(s_{m'})$ remain the same for $m' \neq m$, let us compare the following terms affected by the splitted signal:

$$\log\Big(\sum_{n=1}^{N} e^{(v_n I_m+\mu_n)/\theta}\Big)q(s_m) \quad \text{and} \quad \log\Big(\sum_{n=1}^{N} e^{(v_n \underline{I_m}+\mu_n)/\theta}\Big)\underline{q_m} + \log\Big(\sum_{n=1}^{N} e^{(v_n \overline{I_m}+\mu_n)/\theta}\Big)\overline{q_m},$$

where $\underline{I_m} := \frac{1}{\underline{q_m}}\int_{Q_{m-1}}^{Q_{m-1}+\underline{q_m}} G^{-1}(u)du$ and $\overline{I_m} := \frac{1}{\overline{q_m}}\int_{Q_m-\overline{q_m}}^{Q_m} G^{-1}(u)du$. Denote $\alpha := \frac{\underline{q_m}}{q(s_m)}$, then $1-\alpha = \frac{\overline{q_m}}{q(s_m)}$. Observe that

$$\alpha \underline{I_m} + (1-\alpha)\overline{I_m} = I_m.$$

Consider the function

$$f(x) := \log\Big(\sum_{n=1}^{N} e^{(v_n x+\mu_n)/\theta}\Big). \tag{A.36}$$

We have that

$$f''(x) = \frac{1}{\theta^2}\Big\{\sum_{n=1}^{N}(v_n)^2 w_n - \Big(\sum_{n=1}^{N} v_n w_n\Big)^2\Big\}, \quad w_n = \frac{e^{(v_n x+\mu_n)/\theta}}{\sum_{n'=1}^{N} e^{(v_{n'} x+\mu_{n'})/\theta}}.$$

The Cauchy-Schwarz inequality implies that $f''(x) > 0$, hence f is strictly convex. It then follows from the Jensen's inequality that

$$\log\Big(\sum_{n=1}^{N} e^{(v_n I_m+\mu_n)/\theta}\Big)q(s_m) = q(s_m)f(I_m) = q(s_m)f\big(\alpha\underline{I_m} + (1-\alpha)\overline{I_m}\big)$$

$$< q(s_m)\alpha f(\underline{I_m}) + q(s_m)(1-\alpha)f(\overline{I_m}) = \log\Big(\sum_{n=1}^{N} e^{(v_n \underline{I_m}+\mu_n)/\theta}\Big)\underline{q_m} + \log\Big(\sum_{n=1}^{N} e^{(v_n \overline{I_m}+\mu_n)/\theta}\Big)\overline{q_m}.$$

Therefore, splitting the signal strictly improves the value. Given that q^M is the optimal marginal when M signal states are allowed, splitting $q(s_m)$ strictly improves the value. Optimizing over signals with $M+1$ states weakly improve the value further, hence (A.34) is confirmed.

A.11 Two-state symmetric case

Proposition A.2. *Suppose $N = 2$ and $p(v_1) = p(v_2) = 0.5$. The Lagrangian multiplier μ are zero for symmetric signals.*

Proof. We consider discrete signal with the number of signal states M. The proof for continuous signal is similar.

Given that $q(s_m) = q(s_{M+1-m})$ for any m and the distribution of $N(0, \sigma_Z^2 T)$ is symmetric around zero, we obtain from the definition of I_m that $I_m = -I_{M+1-m}$ for any $1 \leqslant m \leqslant M$. Consider the posterior

$$p(v_n|s_m) = \frac{e^{v_n I_m/\theta}}{e^{v_1 I_m/\theta} + e^{v_2 I_m/\theta}}, \quad n = 1, 2.$$

We claim that this posterior satisfies the Bayesian identity

$$\sum_{m=1}^{M} p(v_n|s_m) q(s_m) = p(v_n) = 0.5, \quad n = 1, 2. \tag{A.37}$$

It then follows from the uniqueness of (μ_1, μ_2) (up to additive constants) that $\mu_1 = \mu_2 = 0$.

To see why (A.37) holds, we observe that

$$\begin{aligned}\sum_{m=1}^{M} p(v_1|s_m) q(s_m) &= \sum_{m=1}^{M} \frac{q(s_m)}{1 + e^{(v_2 - v_1) I_m/\theta}} = \sum_{m=1}^{M} \frac{q(s_{M+1-m})}{1 + e^{(v_1 - v_2) I_{M+1-m}/\theta}} \\ &= \sum_{m'=1}^{M} \frac{q(s_{m'})}{1 + e^{(v_1 - v_2) I_{m'}/\theta}} = \sum_{m'=1}^{M} p(v_2|s_{m'}) q(s_{m'}),\end{aligned}$$

where the second equality follows from $q(s_m) = q(s_{M+1-m})$ and $I_m = -I_{M+1-m}$, the third equality holds due to $m' = M + 1 - m$. Because $p(v_1|s_m) + p(v_2|s_m) = 1$ for any m. The previous displayed equation implies (A.37). □

A.12 Proof of Proposition 5.1

The proof follows that for Proposition 4.2 verbatim, except by replacing all summations over v_n with integrals with respect to v. Because $\mathbb{E}[\tilde{v}^2] < \infty$, the objective function in (2.17) is bounded from above. Therefore, all objective functions in the proof are bounded from above as well.

A.13 Sinkhorn algorithm

The Sinkhorn problem (4.11) can be solved by the Sinkhorn algorithm:

Step 0: Initalize $\xi_n^{(0)} = 1$ for any $n \in \{1, \ldots, N\}$;

Step i: Update $\zeta(z)$ via

$$\zeta^{(i)}(z) = \frac{q(z)}{\sum_{n'=1}^{N} \xi_{n'}^{(i-1)} k_{n'}(z)}, \tag{A.38}$$

Update ξ_n via

$$\xi_n^{(i)} = \frac{p_n}{\int k_n(z)\zeta^{(i)}(z)dz}, \quad 1 \leqslant n \leqslant N. \tag{A.39}$$

Iterate until convergence.

Rüschendorf (1995) prove that this algorithm converges to the unique solution of the Sinkhorn problem (4.11). For the finite state cases, the convergence of the algorithm was proved by Sinkhorn (1964). See also Peyré and Cuturi (2019) for a recent survey of the Sinkhorn algorithm. When the payoff is continuous or the signal state is finite, (4.11) and (A.39) by using summations in finite state space cases and integrals in continuous state space cases.

When p_n is larger, $\xi_n^{(i)}$ in (A.39) increases. Then $\zeta^{(i+1)}(z)$ increases, because the denominator of (A.38) decreases. It then follows from (A.39) for the step $i+1$ that $\xi_n^{(i+1)}$ increases, because the denominator in (A.39) decreases. Following the previous iteration, the limit ξ_n, which is a part of Sinkhorn problem solution, increases. Because $\mu_n = \theta \log(\xi_n)$, we confirm that ξ_n increases with p_n.

A.14 Proof of Proposition 6.1

Step 1: Equivalent formulation using a normal signal. Let us first transform the objective (6.2) using a normal signal. To this end, define $z := G^{-1}\big(Q(s)\big)$, for any $s \in \text{supp}(q)$. Because both Q and G are strictly increasing, this introduces an one-to-one correspondence between z and s. Define $\hat{p}(v_n|z) := p\big(v_n|Q^{-1}(G(z))\big)$ and note that

$$\int \mathbb{E}[\tilde{v}|s]G^{-1}\big(Q(s)\big)q(s)\,ds = \int \sum_{n=1}^{N} v_n p(v_n|s)G^{-1}\big(Q(s)\big)q(s)\,ds = \int \sum_{n=1}^{N} v_n z \hat{p}(v_n|z)g(z)\,dz,$$

where the second equality holds due to $q(s)ds = dQ(s) = dG(z) = g(z)dz$. We transform the problem (6.2) and its constraints equivalently to

$$\max_{\hat{p}(\cdot|\cdot)} \int \sum_{n=1}^{N} v_n z \hat{p}(v_n|z) g(z)\, dz + \theta \int \sum_{n=1}^{N} H(\hat{p}(v_n|z)) g(z)\, dz - \theta \sum_{n=1}^{N} H(p(v_n)), \tag{A.40}$$

subject to

$$\sum_{n=1}^{N} \hat{p}(v_n|z) = 1, \quad \text{for any } z, \tag{A.41}$$

$$\int \hat{p}(v_n|z) g(z)\, dz = p(v_n), \quad \text{for any } 1 \leqslant n \leqslant N. \tag{A.42}$$

This equivalent transformation shows that any continuously distributed signal, which yields an increasing map $s \mapsto \mathbb{E}[\tilde{v}|s]$, leads to the same value.

Step 2: Existence of an optimizer. The feasible set specified by the constraints (A.41) and (A.42) is non-empty (because the prior belongs to thsi feasible set), bounded, and closed (because both constraints are linear). It then follows from the Banach-Alaoglu Theorem that the feasible set is compact in both the weak* topology and the Euclidean topology. The objective (A.40) has a finite value for any $\hat{p}$ in the feasible set. This is because $\hat{p}(v_n|z) \in [0,1]$, $\int |z| g(z) dz < \infty$, and H is finite-valued. Moreover, the objective function is continuous. It then follows from the Weierstrass' Theorem that a maximizer exists.

Step 3: Characterization of an optimizer. Introduce the Lagrangian multiplier $\nu(z)$ for (A.41) and μ_n for (A.42). Consider the unconstrained problem

$$\begin{aligned} \min_{\nu(\cdot),\{\mu_n\}} \max_{\hat{p}(\cdot|\cdot)} & \int \sum_{n=1}^{N} v_n z \hat{p}(v_n|z) g(z)\, dz + \theta \int \sum_{n=1}^{N} H(\hat{p}(v_n|z)) g(z)\, dz - \theta \sum_{n=1}^{N} H(p(v_n)) \\ & + \int \nu(z) \Big[\sum_{n=1}^{N} \hat{p}(v_n|z) - 1 \Big] dz + \sum_{n=1}^{N} \mu_n \Big[\int \hat{p}(v_n|z) g(z) dz - p(v_n) \Big]. \end{aligned} \tag{A.43}$$

Given $\nu(\cdot)$ and $\{\mu_n\}$, the first-order condition of $\hat{p}(v_n|z)$ in (A.43) yields

$$\theta H'\big(\hat{p}^*(v_n|z)\big) = -v_n z - \mu_n - \frac{\nu(z)}{g(z)}, \tag{A.44}$$

for the optimal posterior $\hat{p}^*(v_n|z)$. Because $H'(0+) = \infty$, the optimal posterior $\hat{p}^*(v_n|z)$ is interior, i.e., $\hat{p}^*(v_n|z) > 0$ for any v_n and z.

For given $\{\mu_n\}$, let $\nu^*(\cdot)$ be the optimal Lagrangian multipliers for (A.43). Introduce

the net utility function

$$V^{\text{net}}(z;\mu) := \sum_{n=1}^{N} \Big(v_n z + \mu_n + \frac{\nu^*(z)}{g(z)}\Big)\hat{p}^*(v_n|z) + \theta H\big(\hat{p}^*(v_n|z)\big) - \frac{v^*(z)}{g(z)}. \tag{A.45}$$

Then the optimal value in (A.43) can be rewritten as

$$\min_{\{\mu_n\}} \Big\{ \int V^{\text{net}}(z;\mu) g(z)\, dz - \sum_{n=1}^{N} \mu_n p(v_n) \Big\} - \theta \sum_{n=1}^{N} H(p(v_n)).$$

Step 4: Continuous signal is optimal. For given $\{\mu_n\}$, we are going to show that $V^{\text{net}}(z;\mu)$ is convex in z. Then the proof of the optimality of continuous signals is the same as the proof for Proposition 4.5, where the convexity of f in (A.36) plays a key role.

To this end, it follows from the envelope theorem that

$$\partial_z V^{\text{net}}(z;\mu) = \sum_{n=1}^{N} v_n \hat{p}^*(v_n|z) = \mathbb{E}[\tilde{v}|z]. \tag{A.46}$$

Therefore,

$$\partial_{zz}^2 V^{\text{net}}(z;\mu) = \sum_{n=1}^{N} v_n \partial_z \hat{p}^*(v_n|z). \tag{A.47}$$

To derive $\partial_z p^*(v_n|z)$, we take derivative with respect to z on both sides of (A.44) to obtain

$$\partial_z \hat{p}^*(v_n|z) = \frac{-\hat{\nu}'(z) - v_n}{\theta H''(\hat{p}^*(v_n|z))}, \tag{A.48}$$

where $\hat{\nu}(z) = \nu(z)/g(z)$. Summing over n on both sides of the previous equation and using the fact that $\sum_{n=1}^{N} \hat{p}^*(v_n|z) = 1$ for any z, we obtain that

$$0 = \sum_{n=1}^{N} \partial_z \hat{p}^*(v_n|z) = -\hat{\nu}'(z) \sum_{n=1}^{N} \frac{1}{\theta H''(\hat{p}^*(v_n|z))} - \sum_{n=1}^{N} \frac{v_n}{\theta H''(\hat{p}^*(v_n|z)}.$$

Therefore

$$-\hat{\nu}'(z) = \frac{\sum_{n=1}^{N} \frac{v_n}{\theta H''(\hat{p}^*(v_n|z)}}{\sum_{n=1}^{N} \frac{1}{\theta H''(\hat{p}^*(v_n|z))}}. \tag{A.49}$$

Plugging (A.48) and (A.49) into (A.47), we obtain

$$\partial_{zz}^2 V^{\text{net}}(z;\mu) = \frac{\sum_{n=1}^{N} \frac{1}{-\theta H''(\hat{p}^*(v_n|z))} \sum_{n=1}^{N} \frac{v_n^2}{-\theta H''(\hat{p}^*(v_n|z))} - \left(\sum_{n=1}^{N} \frac{v_n}{-\theta H''(\hat{p}^*(v_n|z)}\right)^2}{\sum_{n=1}^{N} \frac{1}{-\theta H''(\hat{p}^*(v_n|z))}}. \tag{A.50}$$

Given that H is strictly concave, the denominator is positive. Due to the Cauchy Schwartz inequality, the numerator is positive. Therefore, the convexity of $V^{\text{net}}(z;\mu)$ in z is confirmed.

B Two payoff states with asymmetric prior

Figures B.1, B.2, and B.3 present examples for two-state asymmetric prior, i.e., $N = 2$ and $p(v_1) \neq p(v_2)$. The results are similar to those from the symmetric prior case (see Figures 1, 2, and 3). Because the prior distribution is asymmetric, the distributions for $\mathbb{E}[\tilde{v}|\tilde{s}]$ are skewed. However, Panel (D) of Figures B.1 and B.2 shows that the skewness of $\mathbb{E}[\tilde{v}|\tilde{s}]$ converges to 0 (the skewness of a normal distribution) as either the information acquisition cost parameter θ approaches infinity or the noise trade volatility σ_Z converges to zero.

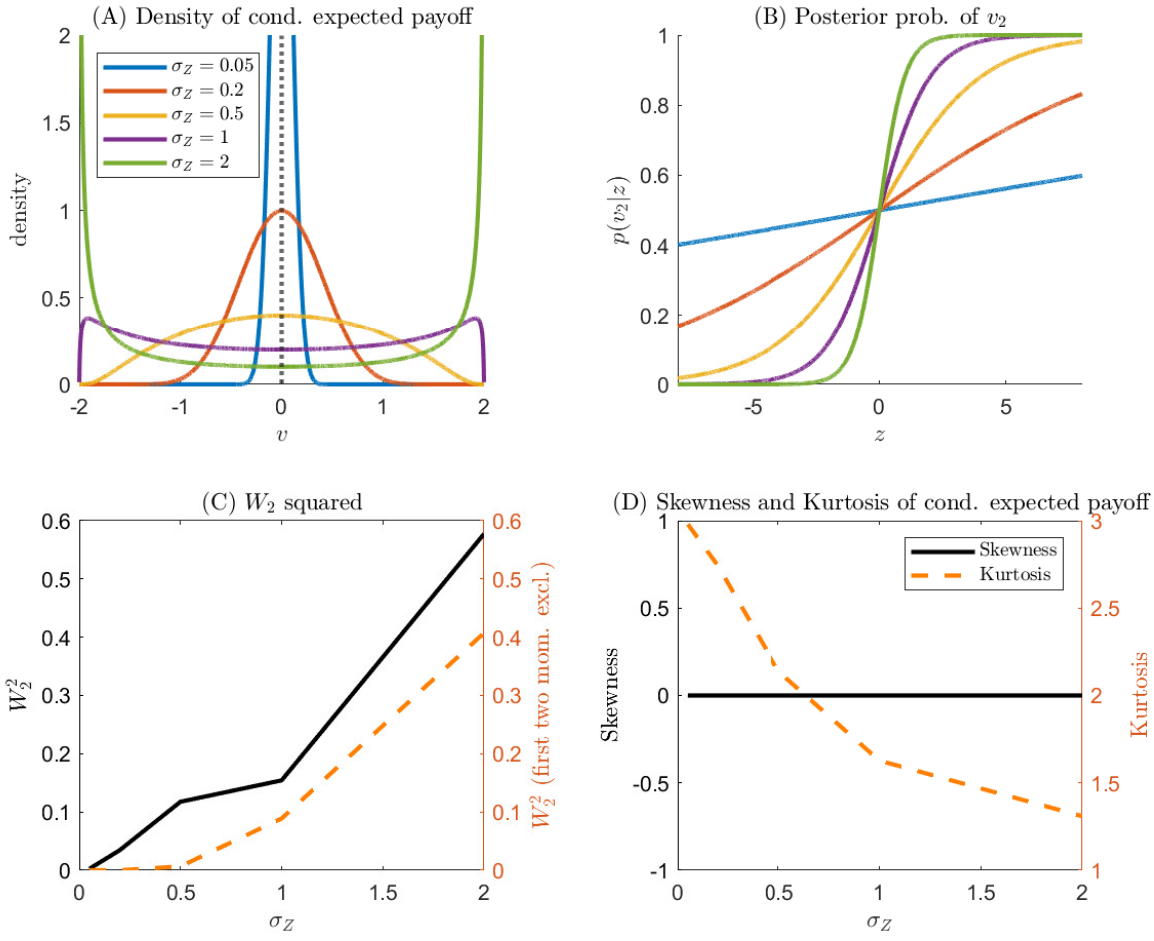

Figure 2: Two-state symmetric prior (different σ_Z)

This figure presents the distribution of the conditional expected payoff, its skewness, Kurtosis, its W_2-distance with noise trade distribution, and the optimal posterior $p(v_2|z)$ for the two-state symmetric prior case. The signal $\tilde{z}$ has a normal distribution $N(0, \sigma_Z^2 T)$. Parameters: $N = 2$, $v_1 = -2, v_2 = 2$, $p(v_1) = p(v_2) = 0.5$, $\theta = 2$, and $T = 1$.

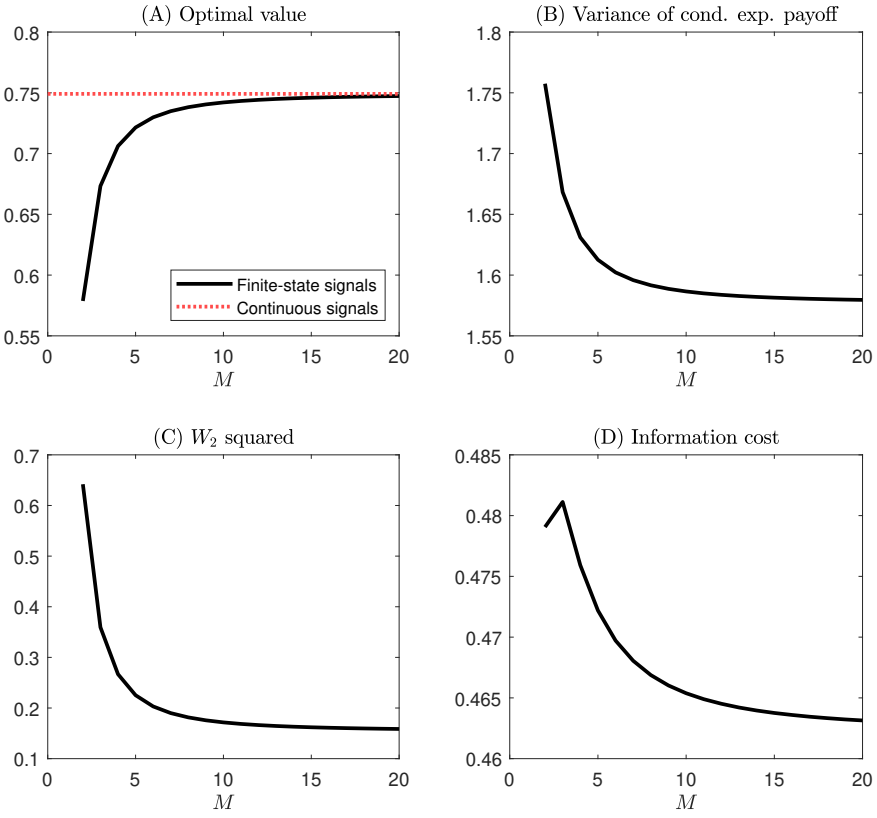

Figure 3: Convergence of values (symmetric prior $p(v_1) = 0.5$)

This figure presents the convergence of optimal value and its different components for the informed trader, as the number os signal states M increases. Parameters: $N = 2$, $v_1 = -2, v_2 = 2$, $p(v_1) = p(v_2) = 0.5$, $\theta = 2$, $\sigma_Z = 1$, and $T = 1$.

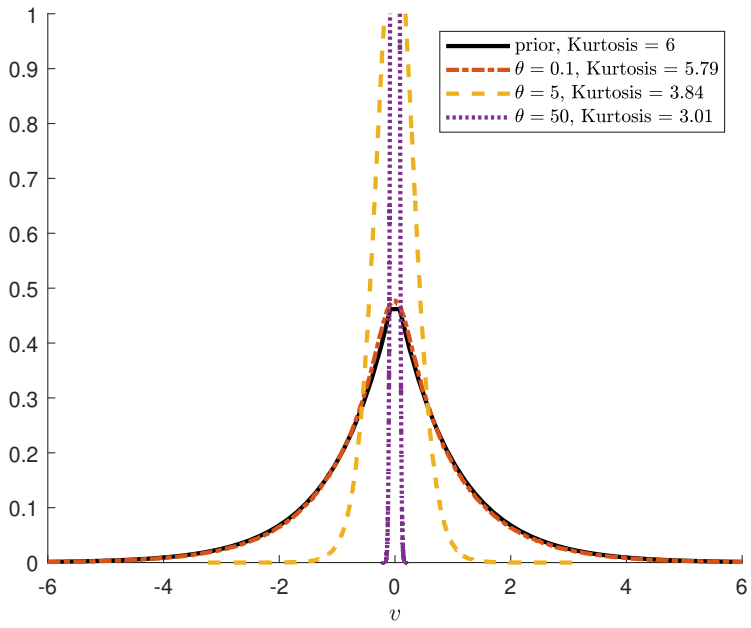

Figure 4: Double exponential distributed payoff (different θ)

This figure presents the density of the conditional expected payoff and its Kurtosis, when the payoff $\tilde{v}$ has a double exponential distribution. The signal $\tilde{z}$ has a normal distribution $N(0, \sigma_Z^2 T)$. The density of the prior distribution of $\tilde{v}$ is $\rho(v) = 0.5e^{-v}$ for $v > 0$ and $\rho(v) = 0.5e^{v}$ for $v < 0$. Other parameters are $\sigma_Z = 1$ and $T = 1$.

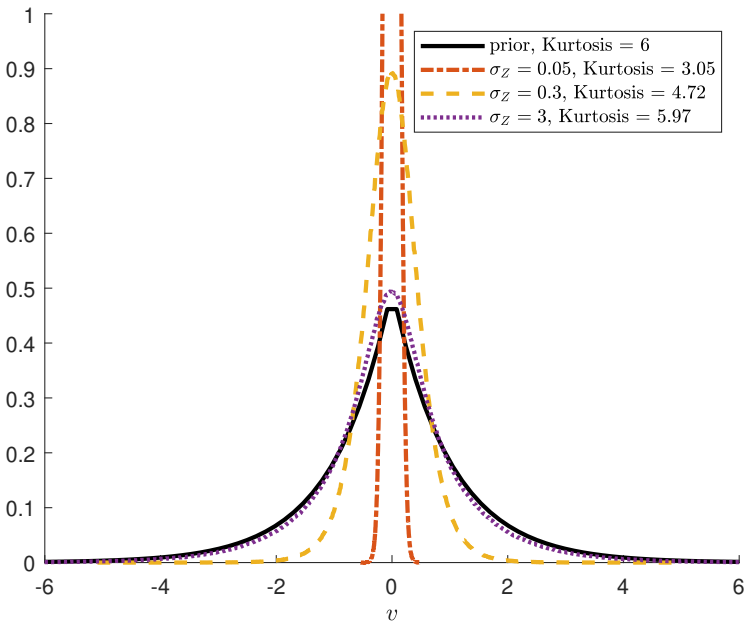

Figure 5: Double exponential distributed payoff (different σ)

This figure presents the density of the conditional expected payoff and its Kurtosis, when the payoff $\tilde{v}$ has a double exponential distribution. The signal $\tilde{z}$ has a normal distribution $N(0, \sigma_Z^2 T)$. The density of the prior distribution of $\tilde{v}$ is $\rho(v) = 0.5e^{-v}$ for $v > 0$ and $\rho(v) = 0.5e^{v}$ for $v < 0$. Other parameters are $\theta = 1$ and $T = 1$.

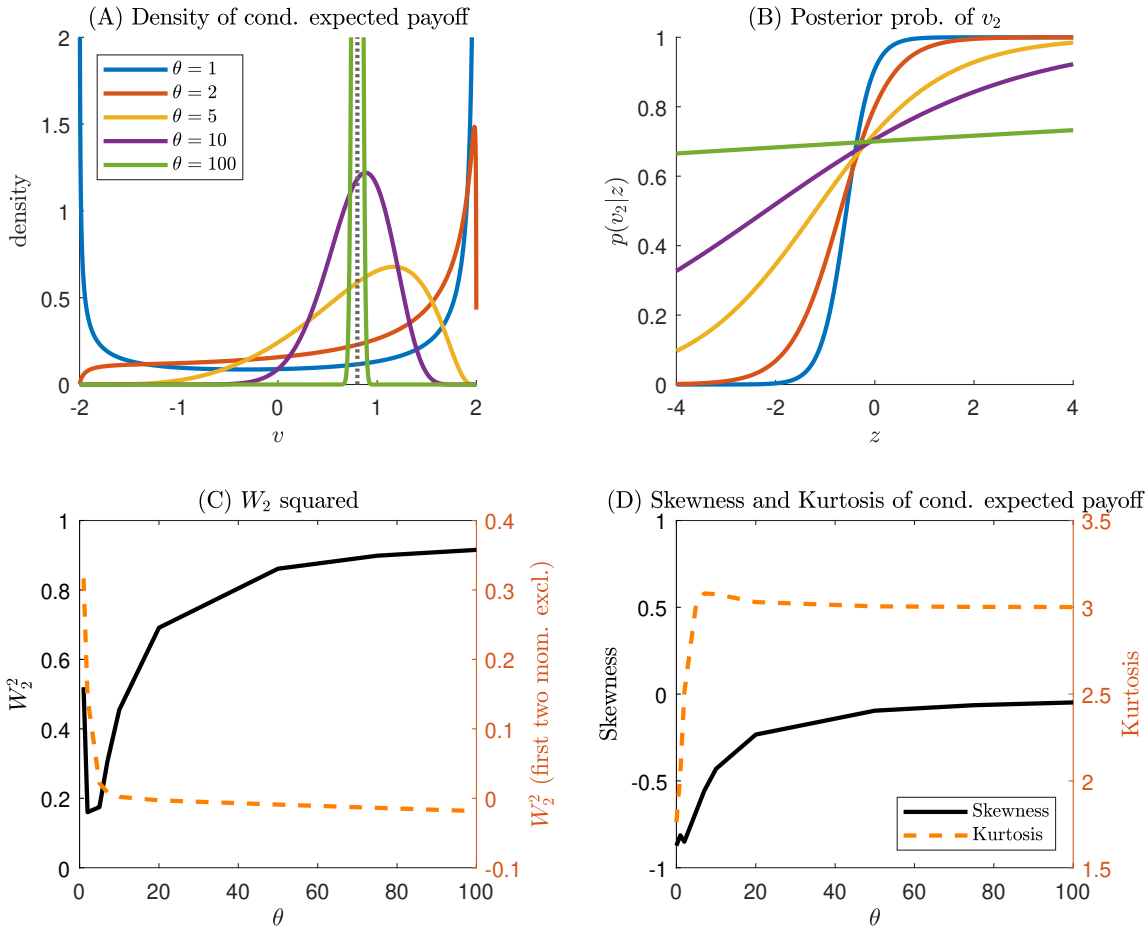

Figure B.1: Two-state asymmetric prior (different θ)

This figure presents the distribution of the conditional expected payoff, its skewness, Kurtosis, its W_2-distance with noise trade distribution, and the optimal posterior $p(v_2|z)$ for the two-state asymmetric prior case. The signal $\tilde{z}$ has a normal distribution $N(0, \sigma_Z^2 T)$. Parameters: $N = 2$, $v_1 = -2, v_2 = 2$, $p(v_1) = 0.3, p(v_2) = 0.7$, $\sigma_Z = 1$, and $T = 1$.

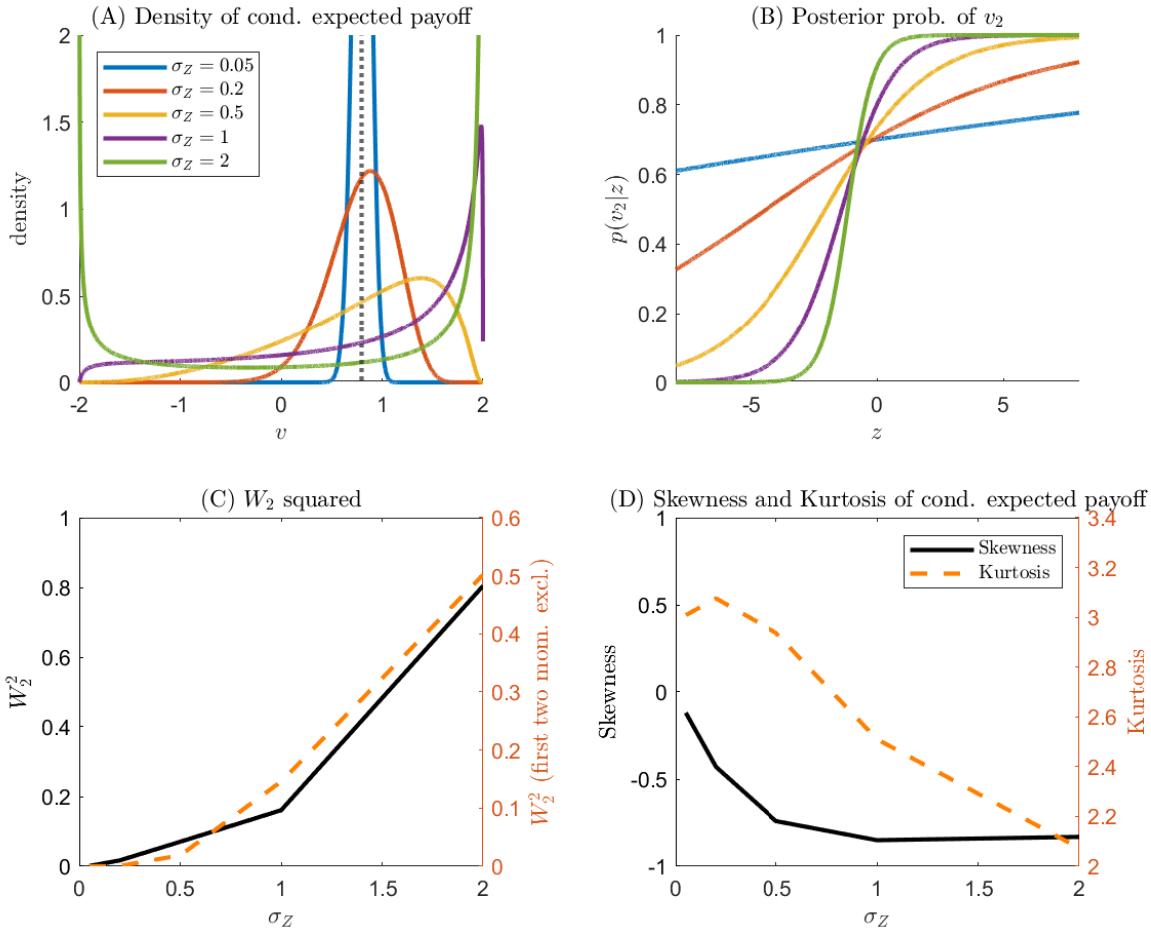

Figure B.2: Two-state asymmetric prior (different σ_Z)

This figure presents the distribution of the conditional expected payoff, its skewness, Kurtosis, its W_2-distance with noise trade distribution, and the optimal posterior $p(v_2|z)$ for the two-state asymmetric prior case. The signal $\tilde{z}$ has a normal distribution $N(0, \sigma_Z^2 T)$. Parameters: $N = 2$, $v_1 = -2, v_2 = 2$, $p(v_1) = 0.3, p(v_2) = 0.7$, $\theta = 2$, and $T = 1$.

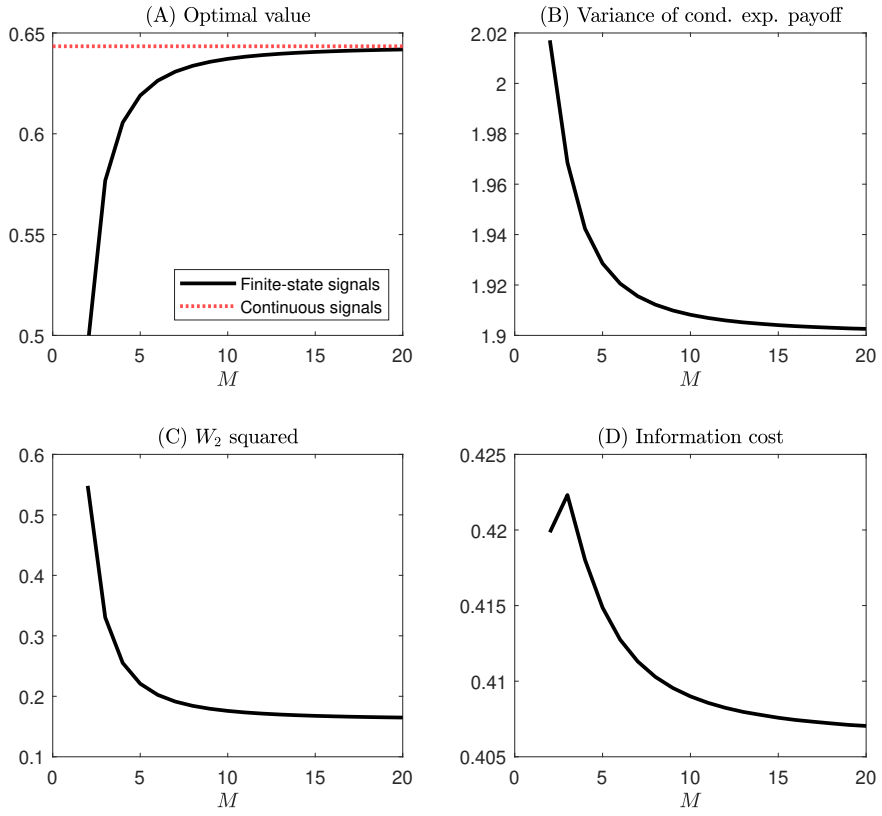

Figure B.3: Convergence of values (asymmetric prior $p(v_1) = 0.3$)

This figure presents the convergence of optimal value and its different components for the informed trader, as the number os signal states M increases. Parameters: $N = 2$, $v_1 = -2, v_2 = 2$, $p(v_1) = 0.3, p(v_2) = 0.7$, $\theta = 2$, $\sigma_Z = 1$, and $T = 1$.