The Design and Interpretation of Information

Shota Ichihashi[*] Delong Meng[†]

November 18, 2021

Abstract

A sender designs an experiment that generates a signal about an uncertain binary state, and then interprets the realized signal. The receiver adopts the sender's interpretation if the signal is more likely under the proposed interpretation than under the true experiment. Using a concavification approach, we characterize the sender's optimal strategy to generate and interpret signals. Strategic interpretation distorts the receiver's action compared to a Bayesian decision maker, but encourages the sender to choose a more informative experiment ex ante. When the latter effect dominates, the receiver's susceptibility to strategic interpretation benefits both the sender and the receiver.

[*] Bank of Canada, email: shotaichihashi@gmail.com. The opinions expressed in this article are the authors' own and do not reflect the views of the Bank of Canada.

[†] Shanghai Jiao Tong University, email: nealthcounts@gmail.com.

1 Introduction

Suppose that a public health agency conducts an experiment on wearing masks, or a clinician tests the effect of a medical intervention. The experiment shows a small or insignificant effect, but the experimenter still wants to promote the result because, e.g., they want more people to wear masks or adopt the medicine. To improve people's assessment of the treatment, the agency or the clinician could interpret the experimental result in a certain way. For example, they may emphasize the effect on a subgroup of patients, or the significance of the effect instead of the magnitude; if some results are insignificant, they may present a possible reason for which the results are false negatives. In scientific communication, such a reporting strategy is called a "spin" and is known to influence readers' interpretation of the experimental treatment (e.g., Boutron et al. (2014)).

Motivated by these examples, we study the strategic (mis)interpretation of experimental outcomes and its interaction with the design of an experiment. Specifically, we study a game in which the sender designs and interprets information to persuade the receiver. Ex ante, the sender chooses an experiment that generates a statistical signal about some binary state. After a signal is realized, the sender provides an "interpretation" of the signal. An interpretation is another Blackwell experiment, which maps each state to a distribution over signals. The receiver adopts the proposed interpretation instead of the true experiment and updates his belief accordingly, whenever the interpretation is a better fit, i.e., it makes the realized signal more likely than the true experiment. Finally the receiver takes a payoff-relevant action. The sender has a state-independent preference.

If strategic interpretation is absent—i.e., if the receiver always updates beliefs according to the true experiment—the model reduces to a version of Bayesian persuasion in Kamenica and Gentzkow (2011). If the true experiment and the realized signal are exogenously fixed, and the sender proposes an interpretation of the signal to affect the receiver's beliefs, the model reduces to "model persuasion" in Schwartzstein and Sunderam (2021). We combine these two approaches to study how the design of information interacts with strategic interpretation.

Our findings are two-fold. First, we characterize the sender's optimal strategy using a version of the concavification approach (Kamenica and Gentzkow, 2011). We cannot directly apply the concavification method, because the receiver's belief updating depends on the endogenous interpretation and on the entire structure of the true experiment. However, we show that the sender can focus on a simple class of experiments and interpretations to attain the optimal outcome. We then show that the sender's optimal outcome corresponds to the concavification of the modified value function, where the modification captures how the sender's strategic interpretation distorts the receiver's beliefs.

Second, strategic interpretation changes the welfare implication of persuasion in two ways. On one hand, interpretation distorts the receiver's posterior beliefs compared to the Bayesian decision maker, and makes the receiver more likely to take a suboptimal action. This *interpretation effect* captures the negative impact of interpretation on the receiver.

On the other hand, strategic interpretation affects the sender's ex ante incentive to design information, which we call the *information effect*. We show that the sender chooses a more informative experiment (as the true experiment) in our model than in Bayesian persuasion. To see the intuition, suppose that the sender wants to promote a new intervention and conducts an experiment. If the experiment reveals that the intervention may not be effective, the sender can come up with an interpretation of the result to convince the receiver that the false negative is ex ante likely and the negative signal is not meaningful. Because such an interpretation better fits the observation in terms of the ex ante likelihood of the negative signal, the receiver will adopt it and dismiss the signal as an uninformative one. Anticipating that strategic interpretation will mitigate the potential loss from unfavorable experimental results, the sender generates more information ex ante.

The information effect improves the ex ante quality of information compared to Bayesian persuasion, but the interpretation effect distorts the receiver's belief. Depending on the players' preferences, either effect can dominate. When the information effect dominates, the receiver's susceptibility to strategic interpretation increases the ex ante payoffs of both the sender and the receiver compared to Bayesian persuasion.

Our paper relates to three strands of literature. First, our work relates to the literature on Bayesian persuasion, in particular the one in which the receiver is not a Bayesian decision maker. de Clippel and Zhang (2020) study an information design problem in which the receiver's belief updating deviates from a rational Bayesian. In their model, the receiver updates beliefs according to an exogenous rule. In our model whether the receiver's inference deviates from Bayesian depends on the sender's interpretation strategy. Several papers, such as Lipnowski and Mathevet (2018), Bloedel and Segal (2018), Galperti (2019), and Lipnowski et al. (2020), endogenize the receiver's (non-Bayesian) belief updating rule. Beauchêne et al. (2019) study a model in which the sender can propose multiple experiments, and an ambiguity averse receiver adopts one of them to maximize the worst-case payoff. In contrast, the receiver in our model chooses between adopting the true experiment and the proposed interpretation to maximize the likelihood of the realized signal. The ex post manipulation of signals arises also in Lipnowski et al. (2019).

Second, our work relates to papers that study scientific communication in models of strategic information acquisition and communication. Di Tillio et al. (2017) study a model in which the sender can manipulate experimental outcomes. The sender's manipulation can benefit the receiver, because manipulation behavior reveals the sender's private information. We study a new kind of manipulation, i.e., the strategic interpretation of outcomes, and its interaction with information design. Libgober (forthcoming) studies a model in which the sender's choice about experiment is multi-dimensional, and examines the impact of transparency.

Finally, several papers model a sender's attempt to interpret information. We adopt model persuasion in Schwartzstein and Sunderam (2021) to formulate the strategic interpretation of signals (Section 2.2 provides a detailed discussion). Aina (2021) studies model persuasion in which the sender commits to a set of experiments at the ex ante stage. These papers assume that a signal about the state is drawn from an exogenously given experiment. In Eliaz et al. (2021), the sender sends a multi-dimensional message and provides a dictionary to interpret it. We adopt a different way to model strategic interpretation, which could be relevant to scientific communication.

2 The Model

We study strategic communication between a sender (she) and a receiver (he). The receiver has a payoff function $u(a, \theta)$ that depends on his action $a \in A$ and the binary state $\theta \in \Theta := \{\theta_0, \theta_1\}$. The action space A is a compact subset of $\mathbb{R}$ and $u(a, \theta)$ is continuous in a. The sender's payoff $v(a)$ depends only on the receiver's action. We identify any $\mu \in \Delta\Theta$ with $\mu(\theta_1)$, the probability it places on θ_1.[1] Thus $\mu' \in [0, 1]$ denotes the distribution over Θ that puts probability μ' on θ_1.

[1] We have ΔX denotes the set of all probability distributions on set X.

The common prior is $\mu_0 \in (0,1)$.

An *experiment* $(S, \pi : \Theta \to \Delta S)$ maps each state to a distribution over signals, where S is the space of possible signals. We write $\pi(s|\theta)$ for the probability of signal $s \in S$ given a realized state $\theta \in \Theta$ under experiment π. Let $\Pr(s|\pi) := \sum_{\theta\in\Theta} \mu_0(\theta)\pi(s|\theta)$ denote the ex ante probability that experiment π draws signal s. Abusing notation, we write $\pi(\theta|s) := \frac{\mu_0(\theta)\pi(s|\theta)}{\Pr(s|\pi)}$ for the posterior probability of state θ conditional on signal s given experiment π.

Let Π denote the set of the experiments the sender can use. The sender chooses from Π twice in the game: In the first stage, the sender chooses from Π to generate a signal about the state. In the second stage, the sender chooses from Π to provide the receiver with an interpretation of the realized signal. For clarity, we call $\pi \in \Pi$ an *experiment* when we refer to a generic element of Π or the sender's choice in the first stage. We also call π the *true experiment* to emphasize that it is the sender's choice in the first stage. We call $\pi \in \Pi$ an *interpretation* for the sender's choice in the second stage. We restrict Π as follows:

Assumption 1 (No Relabeling)**.** *The set* Π *of feasible experiments is*

$$\Pi = \{(S,\pi) : S = \{0,1\} \textit{ and } \pi(1|\theta_1) \geq \pi(1|\theta_0)\}. \tag{2.1}$$

Any experiment $\pi \in \Pi$ satisfies both $\pi(1|\theta_1) \geq \pi(1|\theta_0)$ and $\pi(0|\theta_0) \geq \pi(0|\theta_1)$. In terms of the Bayesian posteriors that π induces, the assumption means that π belongs to Π if and only if $\pi(\theta_1|1) \geq \mu_0$ and $\pi(\theta_0|0) \geq 1-\mu_0$, i.e., signals 1 and 0 indicate that states θ_1 and θ_0 are weakly more likely than at the prior, respectively. For example, θ_1 means that a new treatment is effective, and signal 1 or 0 is a positive or negative result, respectively. The assumption means that after observing a positive result, a Bayesian believes that the treatment is (weakly) more likely to be effective than at the prior. Generally, Assumption 1 represents a commonly understood relation between the state and the statistical signal that the sender's experiment can generate. We believe that such a restriction is natural in the context of scientific communication. The assumption also restricts the kinds of interpretation the sender can use to influence the receiver's beliefs.

The timing of the game is as follows. First, the sender publicly chooses an experiment $\pi^* \in \Pi$. Nature draws state $\theta \sim \mu_0$ and signal $s \sim \pi^*(\cdot|\theta) \in \Delta S$. The state θ is unobservable, but signal s is publicly observable. The sender then provides an interpretation $\pi_s \in \Pi$ of signal s. The receiver adopts the proposed interpretation π_s instead of the true experiment π^* if and only if π_s renders signal s more likely than π^* does: Formally, the receiver adopts π_s if and only if $\Pr(s|\pi_s) \geq \Pr(s|\pi^*)$. Finally, the receiver updates his belief according to the adopted interpretation, and then chooses an action: If the receiver adopts π_s, he chooses action $a_s \in \operatorname{argmax}_a \sum_\theta \pi_s(\theta|s)u(a,\theta)$; if the receiver adopts π^*, he chooses action $a_s \in \operatorname{argmax}_a \sum_\theta \pi^*(\theta|s)u(a,\theta)$. The receiver breaks ties in favor of the sender upon choosing an action. We study the sender's optimal persuasion strategy.

Definition 1. The *sender's optimal strategy* consists of the true experiment $\pi^* \in \Pi$ and interpretations (π_0, π_1) that maximize her expected payoff $\sum_{\theta\in\{\theta_0,\theta_1\}} \mu_0(\theta) \sum_{s\in\{0,1\}} \pi^*(s|\theta)v(a_s)$, where a_0 and a_1 are induced by the receiver's optimal behavior described above.

2.1 The Role of Interpretation and Assumption 1

We illustrate how interpretation expands the set of outcomes the sender can achieve. We also motivate Assumption 1 and explain how it restricts the attainable outcomes.

Suppose that the two states are equally likely, and the sender prefers the receiver to have a larger $\mu = \Pr(\theta_1)$ as a posterior belief. For example, the sender wants the receiver to choose a higher action, and the receiver optimally chooses a higher action when he believes that state θ_1 is more likely.

We make three observations. First, the sender can attain a weakly greater payoff than in Bayesian persuasion. To see this, take any experiment $\pi^* \in \Pi$ and signal s. The sender can induce the receiver to adopt π^* as the interpretation by setting $\pi_s = \pi^*$. The receiver's posterior after each signal s will be the Bayesian posterior calculated from π^*. Thus the sender in our model can attain the same outcome as Bayesian persuasion. Note that Assumption 1 restricts feasible experiments. However, in Bayesian persuasion with binary states, the sender has an optimal policy that induces a two-point or degenerate distribution of posterior beliefs, and she can induce any such distribution under Assumption 1.[2]

Second, in some cases the sender prefers to distort the receiver's posterior. Suppose that the

[2] Take any two-point Bayes-plausible distribution of posteriors that induces $\mu_L < \mu_0$ and $\mu_H > \mu_0$. The sender can induce it with $\pi \in \Pi$ such that $\pi(1|\theta_1) = \frac{\mu_H}{\mu_0}\left(\frac{\mu_0-\mu_L}{\mu_H-\mu_L}\right)$ and $\pi(1|\theta_0) = \frac{1-\mu_H}{1-\mu_0}\left(\frac{\mu_0-\mu_L}{\mu_H-\mu_L}\right)$, which satisfies Assumption 1.

sender has chosen the fully informative experiment π^*, i.e., $\pi^*(0|\theta_0) = \pi^*(1|\theta_1) = 1$. Signal 0 leads to the Bayesian posterior $\mu = 0$. If the sender wants the receiver to have a greater μ, she can propose the uninformative experiment as an interpretation, i.e., π_0 such that $\pi_0(0|\theta_0) = \pi_0(0|\theta_1) = 1$. The receiver adopts π_0 because it better fits the observation: $\Pr(0|\pi_0) = 1 > 0.5 = \Pr(0|\pi^*)$. Because the receiver updates his belief as if signal $s = 0$ is drawn from π_0, his posterior is 0.5. In contrast, if signal 1 is realized, the sender can induce the receiver to maintain $\mu = 1$. As a result, from the ex ante perspective, the receiver will hold subjective posterior 0.5 or 1 with equal probability. The outcome violates the Bayes-plausibility condition and cannot arise under Bayesian persuasion. In scientific communication, interpretation π_0 that follows $s = 0$ corresponds to a "spin," i.e., a reporting strategy that puts negative findings in a more palatable way to readers (Boutron et al., 2010). For example, a researcher might downplay a negative result (signal 0) by arguing that a certain aspect of the experiment renders the false negative likely.

Finally, there are some outcomes that the sender cannot attain. For example, suppose again that the sender chooses the fully informative experiment π^*, and signal 0 is realized. Suppose the sender tries to propose interpretation $\pi_0(0|\theta_0) = 2\epsilon$ and $\pi_0(0|\theta_1) = 1$. The ex ante likelihood of signal 0 under π_0 is $0.5\pi_0(0|\theta_1) + 0.5\pi_0(0|\theta_0) = 0.5 + \epsilon$, which is strictly greater than $\Pr(0|\pi^*) = 0.5$. If the receiver adopts π_0, he would hold posterior $\frac{0.5}{0.5+\epsilon}$, which could be arbitrarily close to 1 for a small ϵ. However, Assumption 1 prevents the sender from choosing π_0 as an interpretation, because π_0 violates inequality (2.1). Intuitively, π_0 reverses the meaning of signals—e.g., it treats a negative signal, such as an insignificant result, as the indication of a positive effect. Under Assumption 1 the receiver dismisses such an interpretation as unreasonable. In the example of a clinical intervention, the researcher might downplay negative results, but it would be difficult to convince the audience that for some reason, a negative result indicates a positive effect of the intervention.

2.2 Discussion of Assumptions

Comparison to "Model Persuasion." If we fix the true experiment π^* and signal s, but allow the sender to choose an interpretation, the model becomes "model persuasion" in Schwartzstein and Sunderam (2021). In particular, we follow their paper and assume that the receiver adopts the sender's interpretation whenever it makes the realized outcome more likely than the chosen experiment. The main difference between the two papers is that in our model, the experiment from which the signal is drawn is endogenous and chosen by the sender. In terms of the receiver's belief formation, we assume that the receiver's "default model" is equal to the true experiment. The assumption implies that in the absence of strategic interpretation, the receiver interprets the outcome of the chosen experiment as it is. This assumption on the default model is a tractable way to connect the sender's choice of the true experiment and the receiver's default model.

A Model Without the No Relabeling Assumption. Assumption 1 is not only natural for our intended applications, but also necessary to conduct a meaningful analysis. To see this, suppose we drop the assumption and allow the sender to choose any experiment and interpretation with a signal space $\{1, 2, \ldots, M\}$ for a large M. Take any belief μ, and consider the following $M + 1$ experiments: First, experiment $\pi_\emptyset$ draws one of the M signals uniformly randomly, independently of θ. Second, for each $m \in \{1, ..., M\}$, experiment $\pi_{\mu,m}$ draws signal m with the ex ante probability of at least $\frac{1}{M}$, and following signal m, $\pi_{\mu,m}$ induces posterior μ.[3] Suppose the sender chooses $\pi_\emptyset$ as the true experiment, and after each signal m, she chooses $\pi_{\mu,m}$ as an interpretation. After every signal m, the receiver adopts $\pi_{\mu,m}$ and holds posterior μ. Thus the sender can induce the receiver to have any posterior with the ex ante probability of 1.

3 The Sender's Optimal Strategy

We characterize the sender's optimal strategy and payoff. Let $V(\mu)$ denote the sender's payoff when the receiver has a (possibly non-Bayesian) posterior of μ and chooses an action optimally, breaking ties in favor of the sender. Because the sender's payoff is state-independent, V depends only on the receiver's posterior. Recall that $\mu \in [0, 1]$ denotes the distribution on $\Theta = \{\theta_0, \theta_1\}$ that puts probability μ on θ_1.

Kamenica and Gentzkow (2011) use concavification to solve Bayesian persuasion. We may think that the same approach works—i.e., we concavify $V(g(\mu'))$, where $g(\mu')$ is the receiver's subjective posterior conditional on the Bayesian posterior μ' and the sender's optimal interpre-

[3] For example we can use $M > \max_{\theta \in \Theta} \dfrac{\mu(\theta)}{\mu_0(\theta)}$ and $\pi_{\mu,m}(m|\theta) = \dfrac{\mu(\theta)}{\mu_0(\theta)} \cdot \left[\max_{\theta \in \Theta} \dfrac{\mu(\theta)}{\mu_0(\theta)}\right]^{-1}$. Experiment $\pi_{\mu,m}$ induces posterior μ after signal m, and the ex ante probability of signal m is $\left[\max_{\theta \in \Theta} \dfrac{\mu(\theta)}{\mu_0(\theta)}\right]^{-1} > \frac{1}{M}$.

tation strategy. The approach does not work because $g(\mu')$ is not well-defined: Whether the receiver adopts the sender's interpretation depends not on the Bayesian posterior μ', but on the likelihood of each signal under the true experiment π. Nonetheless, we provide an analogous characterization result.

Definition 2. Given prior μ_0, the μ_0-*monotone hull* of V is defined as function V^M such that for all $\mu \in [0,1]$,

$$V^M(\mu) := \max_{x\in[0,1]} V(x) \quad \text{subject to} \quad \min\{\mu, \mu_0\} \leq x \leq \max\{\mu, \mu_0\}.$$

Unlike the usual concavification, V^M depends on prior μ_0 (see Figure 1). On the interval $[\mu_0, 1]$, V^M is the smallest (weakly) increasing function that is always above V. On the interval $[0, \mu_0]$, V^M is the smallest (weakly) decreasing function that is always above V.

To see the intuition for V^M, suppose that the sender chooses the true experiment that induces posterior $\mu_L < \mu_0$ with probability $\frac{\mu_H-\mu_0}{\mu_H-\mu_L}$ (following signal 0) and posterior $\mu_H > \mu_0$ with probability $\frac{\mu_0-\mu_L}{\mu_H-\mu_L}$ (following signal 1). If the sender interprets signal 0 as it is, the receiver's posterior is μ_L. However, for any $\mu \in [\mu_L, \mu_0]$, the sender can interpret signal 0 by proposing interpretation π_0 that induces posterior μ with probability $\frac{\mu_H-\mu_0}{\mu_H-\mu}$ (following signal 0) and posterior μ_H with probability $\frac{\mu_0-\mu}{\mu_H-\mu}$ (following signal 1). The receiver updates his beliefs according to the proposed interpretation π_0, because signal 0 is more likely under π_0 than π, i.e., $\frac{\mu_H-\mu_0}{\mu_H-\mu} \geq \frac{\mu_H-\mu_0}{\mu_H-\mu_L}$. Thus $V^M(\mu_L)$ is a lower bound of the sender's payoff given the Bayesian posterior μ_L. The same argument holds for $\mu_H \geq \mu_0$. As a result, the sender's expected payoff is at least the concavification of V^M evaluated at prior μ_0.

The following result shows that the concavification of V^M at μ_0 is indeed the sender's optimal outcome. We say that the sender *honestly interprets signal* s if, after signal s is realized, the sender proposes an interpretation that equals the true experiment.

Proposition 1. *The sender's ex ante payoff at the optimal strategy is the concavification of the* μ_0*-monotone hull* V^M *of* V*, evaluated at prior* μ_0*. There is an optimal strategy that satisfies one of the following:*

1. *The sender induces the receiver to maintain prior* μ_0 *with probability* 1.

2. *The true experiment induces Bayesian posteriors* 0 *and* $\mu \in (\mu_0, 1]$ *after signals* 0 *and* 1*, respectively, and the sender honestly interprets signal* 1.

3. *The true experiment induces Bayesian posteriors* $\mu \in [0, \mu_0)$ *and* 1 *after signals* 0 *and* 1*, respectively, and the sender honestly interprets signal* 0.

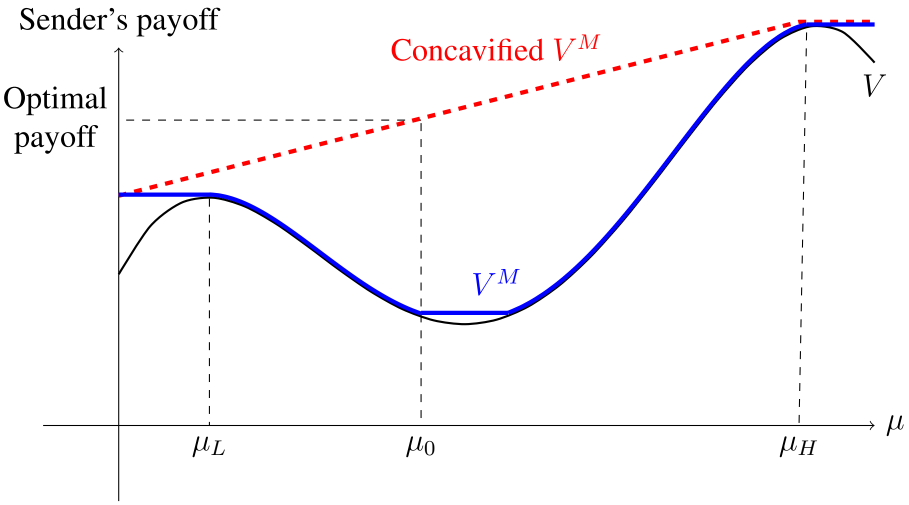

Figure 1: The Sender's Optimal Payoff

Figure 1 illustrates Part 2 of the result. The black curve is the sender's payoff V as a function of the receiver's posterior. The blue thick graph is the μ_0-monotone hull of V, and the red-dashed graph is the concavification of V^M. The figure also shows the sender's optimal strategy. The concavification of V^M splits prior μ_0 into posteriors $\mu = 0$ and $\mu = \mu_H$, which implies that the sender chooses an experiment that induces these posteriors. The sender honestly interprets signal 1 so that the receiver maintains μ_H. Following signal 0, which leads to the Bayesian posterior 0, the sender proposes an interpretation that induces posteriors μ_L and μ_H. The receiver adopts the interpretation and holds posterior μ_L, because signal 0 is more likely under the proposed interpretation. For example, the sender may say that the experiment has some problems and we cannot interpret the negative signal (signal 0) to conclude that the state is 0 for sure.

Figure 2 illustrates the sender's optimal payoff in three more cases. The thick black lines depict the sender's value function V, and the red thick dashed lines depict the concavification of V^M. The left panel is a prosecutor-judge example of Kamenica and Gentzkow (2011) in which the sender's payoff is 0 if the receiver's posterior is below $1/2$ and is positive if it is above

$1/2$. The μ_0-monotone hull coincides with the original value function, so the sender-optimal outcome coincides with that of Bayesian persuasion.

In the middle panel, the sender's payoff is a convex function of the receiver's posterior, so the solution of Bayesian persuasion is full disclosure. In our model the sender also chooses a fully informative experiment. However, the sender now earns a strictly higher payoff than in Bayesian persuasion because she can interpret signal 0 (that led to the Bayesian posterior $\mu = 0$) as a pure noise. To do so, the sender proposes an uninformative experiment that sends signal 0 with probability 1 after every state.

In the right panel the sender's payoff is strictly concave, so the Bayesian persuasion solution is no disclosure. In our setting the sender chooses a partially informative experiment that generates Bayesian posteriors 0 and μ_H, following which she interprets $\mu = 0$ as a pure noise.

The three panels show that the sender's interpretation may divert the receiver's posterior away from the posterior of the Bayesian decision maker, but it also affects the sender's choice of the true experiment. The next section examines how the two channels affect the receiver's payoffs.

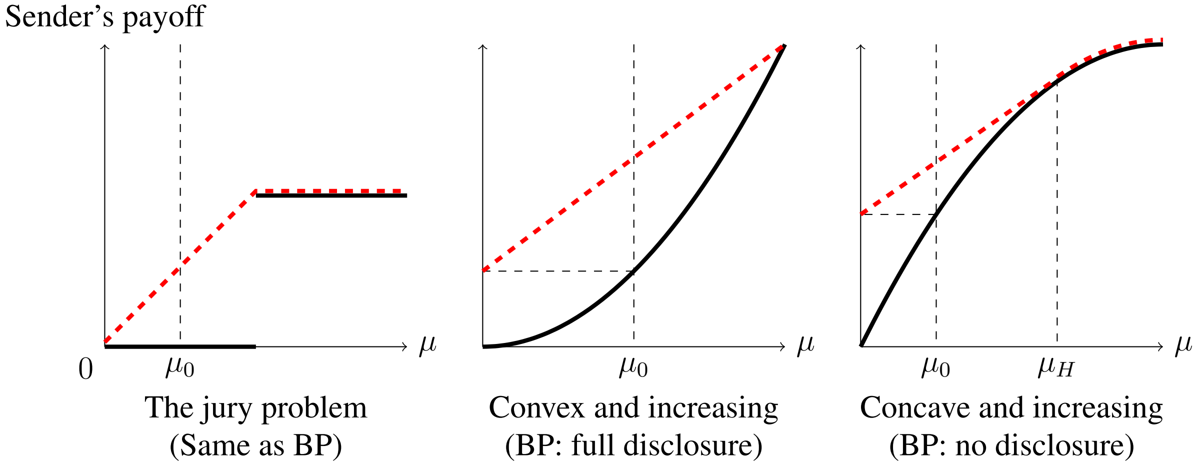

Figure 2: Comparisons to Bayesian Persuasion (BP)

4 The Interpretation and Information Effects

We examine how the sender's strategic interpretation affects the receiver's payoff compared to Bayesian persuasion. Let $\tilde{\mu}^{BP}$ denote the receiver's posterior under the optimal experiment in

Bayesian persuasion. We view $\tilde{\mu}^{BP}$ as a random variable from the ex ante perspective. Let $(\tilde{\mu}^T, \tilde{\mu}^R)$ denote the outcome of our model: $\tilde{\mu}^T$ is the Bayesian posterior under the sender's true experiment, and $\tilde{\mu}^R$ is the receiver's posterior given the sender's optimal interpretation. We view $(\tilde{\mu}^T, \tilde{\mu}^R)$ as a joint random variable.

Let $U(\mu, \mu')$ denote the receiver's interim payoff when the true posterior is μ but he acts optimally based on posterior μ'.[4] We use $\mathbb{E}[\cdot]$ as the expectation with respect to the random posteriors $\tilde{\mu}^{BP}$ or $(\tilde{\mu}^T, \tilde{\mu}^R)$ induced by the sender's strategy. We can decompose the receiver's gain from the sender's strategic interpretation as follows:

$$\begin{aligned}&\mathbb{E}[U(\tilde{\mu}^T, \tilde{\mu}^R)] - \mathbb{E}[U(\tilde{\mu}^{BP}, \tilde{\mu}^{BP})]\\ =&\underbrace{\mathbb{E}[U(\tilde{\mu}^T, \tilde{\mu}^R)] - \mathbb{E}[U(\tilde{\mu}^T, \tilde{\mu}^T)]}_{\text{Interpretation Effect}} + \underbrace{\mathbb{E}[U(\tilde{\mu}^T, \tilde{\mu}^T)] - \mathbb{E}[U(\tilde{\mu}^{BP}, \tilde{\mu}^{BP})]}_{\text{Information Effect}}.\end{aligned}$$

The first term, which we call the *interpretation effect*, captures the downside of strategic interpretation. Given a chosen experiment, the sender's interpretation distorts the receiver's posterior and makes it more likely that he takes a suboptimal action. The second term, which we call the *information effect*, captures the effect of strategic interpretation on the ex ante choice of experiment: Anticipating that the sender can influence the receiver's posterior via interpretation, the sender may choose an experiment that is different from Bayesian persuasion. The following result shows that the interpretation effect is non-positive, but the information effect is non-negative—i.e., the sender who can strategically interpret signals chooses a more informative experiment in the ex ante stage.

Proposition 2. *The interpretation effect is non-positive. In contrast, the information effect is non-negative. Specifically, for any solution of Bayesian persuasion, we can find an experiment that is a part of the sender's optimal strategy and is weakly more (Blackwell) informative than the Bayesian persuasion solution.*

The interpretation effect harms the receiver, and the information effect benefits the receiver. The following example shows that either effect can dominate. Suppose $\mu_0 = \frac{1}{2}$ and $A = [0, 1]$. The receiver's payoff is $1 - (a - \theta)^2$, so he optimally chooses $a = \mu$ at posterior $\mu = \Pr(\theta_1)$. The sender's payoff $V(\mu)$, as a function of the receiver's posterior μ, is piece-wise linear and

[4] Formally, let $a(\mu')$ denote receiver's chosen action when her posterior is μ'. We have $U(\mu, \mu') = \sum_\theta \mu(\theta) \cdot u(a(\mu'), \theta)$.

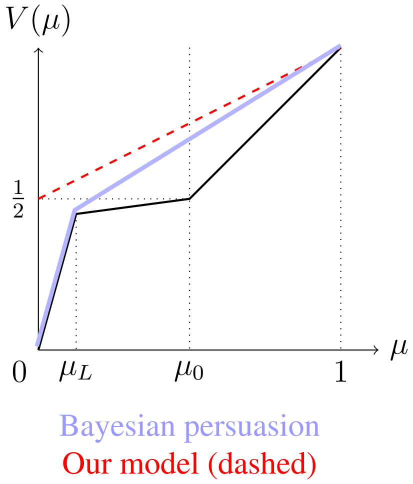

Figure 3: Sender's payoffs

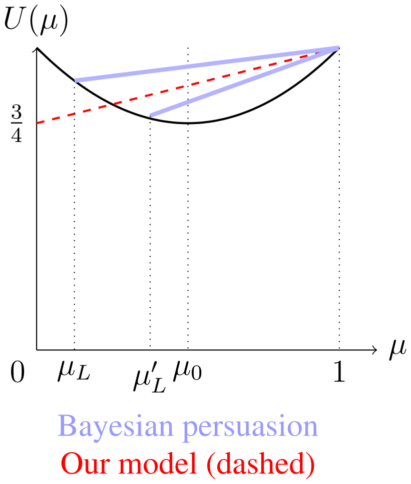

Figure 4: Receiver's payoffs

connects $(0, 0)$, $(\mu_L, 0.49)$, $(0.5, 0.5)$, and $(1, 1)$, where $\mu_L < 0.49$. The black solid line in Figure 3 depicts V.

The solution of Bayesian persuasion follows from the concavification of V (the thick light blue line in Figure 3). The sender induces posteriors μ_L and 1. When the receiver has a Bayesian posterior of μ, his payoff is $U(\mu) = 1 - [(1 - \mu) \cdot (\mu - 0)^2 + \mu \cdot (\mu - 1)^2]$, as depicted by the black line in Figure 4. The receiver's payoff under Bayesian persuasion is the value of the light blue line that connects $U(\mu_L)$ and $U(1)$, evaluated at μ_0.

In our model the sender chooses a fully informative experiment, which induces a Bayesian posterior 0 or 1 after signal 0 or 1, respectively. Ex post when signal 0 is realized, the sender interprets it as a pure noise, inducing the receiver to maintain prior μ_0. As a result the sender's interpretation induces the receiver to have a posterior of $\frac{1}{2}$ or 1 with equal probability. Hence when the true posterior is 0, the receiver takes $a = \frac{1}{2}$, and her actual payoff $1 - (0 - \frac{1}{2})^2 = \frac{3}{4}$. When the true posterior is 1, the receiver takes $a = 1$ and obtains a payoff of 1. The receiver's payoff is thus the line that connects $(0, \frac{3}{4})$ and $(1, 1)$ evaluated at μ_0. Figure 4 shows the receiver's payoff as the average of $\frac{3}{4}$ and 1, which is the value of the red dashed line at μ_0.

Depending on μ_L, strategic interpretation can benefit or hurt the receiver. The magnitude of the interpretation effect is large if μ_L is close to 0, because the receiver's distorted belief μ_0 is further away from the Bayesian posterior μ_L. Indeed in Figure 4 the red line (i.e, the receiver's payoff in our model) is independent of μ_L, but the blue line (i.e., the receiver's payoff under Bayesian persuasion) gets flatter as μ_L goes to 0. In particular if we replace μ_L with μ'_L where

μ'_L is close to μ_0, then the receiver's payoff is lower in Bayesian persuasion. Thus the receiver benefits from the sender's strategic interpretation if and only if μ_L is close enough to μ_0.

5 Extensions

5.1 Persuading the Public via Strategic Interpretation

This section considers a sender interested in persuading a population of receivers. We provide a simple condition under which the information effect dominates the interpretation effect. The sender could be a public agency trying to encourage people to take some private costly action, such as wearing a mask. To persuade the receivers, the agency conducts an experiment, interprets the result, and then issues a public report.

We modify our model as follows. The sender faces a unit mass of receivers. Each receiver $i \in [0, 1]$ cares only about his action, $a_i \in \{0, 1\}$. Receiver i's payoff is $a_i(\theta - c_i)$, where the state $\theta \in \{0, 1\}$ is binary and the cost c_i of taking $a_i = 1$ is distributed across receivers according to some cumulative distribution function F that has a positive density f on its support $[0, 1]$. Given the receivers' (common) posterior $\mu \in [0, 1]$ on $\theta = 1$, the mass of receivers who take $a_i = 1$ is $F(\mu)$. The sender publicly chooses an experiment and interprets the realized signal to maximize her payoff, which is the mass of receivers who choose $a_i = 1$.

Suppose that the true posterior is μ, but the receivers have posterior μ'. The sender's payoff is $F(\mu')$. The receivers' welfare, which is defined as the average payoff of all receivers, is $F(\mu')[\mu - \mathbb{E}_F(c|c \leq \mu')]$, because receiver i chooses $a_i = 1$ whenever $c_i \leq \mu'$.

Proposition 3. *If F is strictly concave, the sender's payoff and the receivers' welfare are strictly higher in our model than in Bayesian persuasion. If F is strictly convex, the sender's payoff is strictly greater, but the receivers' welfare is strictly lower in our model than in Bayesian persuasion.*

If the Bayesian persuasion solution is full disclosure, the information effect is zero, so the receivers are worse off when the sender can interpret signals. In contrast, if the Bayesian persuasion solution is no disclosure, the information can be positive. The result shows that the information effect is indeed positive and dominates the interpretation effect for a concave F.

5.2 Negative Information Effect for a General State Space

We extend the model to accommodate a general state space and then show that the information effect can be negative. The receiver's payoff $u(a, \theta)$ depends on his action $a \in A$ and the state $\theta \in \Theta$, where Θ is finite. The sender's payoff $v(a)$ depends only on the receiver's action. They share a common prior $\mu_0 \in \Delta\Theta$ with full support. Let Π denote the set of all experiments the sender can use. We extend Assumption 1 as follows.

Assumption 2. *The set Π of feasible experiments is*

$$\Pi = \left\{(S, \pi) : S = \{s_\theta\}_{\theta \in \Theta} \text{ and } \forall \theta, \theta' \in \Theta, \pi(s_\theta|\theta) \geq \pi(s_\theta|\theta')\right\}. \tag{5.1}$$

The assumption generalizes Assumption 1. Figure 5 considers $\Theta = \{\theta_1, \theta_2, \theta_3\}$ and shows the set of feasible experiments in terms of the posteriors each signal can induce. The triangle represents the belief space. The right and top vertices represent degenerate beliefs $\Pr(\theta_1) = 1$ and $\Pr(\theta_2) = 1$, respectively. The dashed lines partition the triangle into three areas. Assumption 2 means that the sender can choose an experiment π if and only if posterior $\pi(\cdot|s_{\theta_1})$ is in the right bottom (gray) area, $\pi(\cdot|s_{\theta_2})$ is in the top area, and $\pi(\cdot|s_{\theta_3})$ is in the left bottom area.

The timing of the game is the same as before. First, the sender publicly chooses an experiment $\pi^* \in \Pi$. Nature draws a state $\theta \sim \mu_0$ and a signal $s \sim \pi^*(\cdot|\theta) \in \Delta S$. The sender then provides an interpretation $\pi_s \in \Pi$ of signal s. The receiver adopts interpretation π_s if and only if the ex ante probability of s is greater under π_s than π^*. The receiver updates his belief according to the adopted interpretation, and then chooses an action, breaking ties in favor of the sender.

The following example shows that under a general state space, the sender in our model may choose the true experiment that is less informative than in Bayesian persuasion. Thus Proposition 2 may not hold for a general state space.

Example 1. Suppose $\Theta = \{\theta_1, \theta_2, \theta_3\}$ and $\mu_0 = (1/3, 1/3, 1/3)$. We construct $V(\mu)$ in Figure 5. Let ϵ be a small positive number. The three vertices have values $V(1, 0, 0) = \epsilon$, $V(0, 1, 0) = 100 + \epsilon$, and $V(0, 0, 1) = 98 + \epsilon$. The midpoints of the three edges have values $V(1/2, 0, 1/2) = 49$, $V(1/2, 1/2, 0) = 50$, and $V(0, 1/2, 1/2) = 99$. We also have $V(\mu_0) = 66$. Define function

$V(\cdot)$ as the lower convex hull of the given points. Function V is linear within each of the three quadrilaterals of Figure 5. If $\epsilon = 0$, then V is linear in its entire domain, but if $\epsilon > 0$, then V is convex.

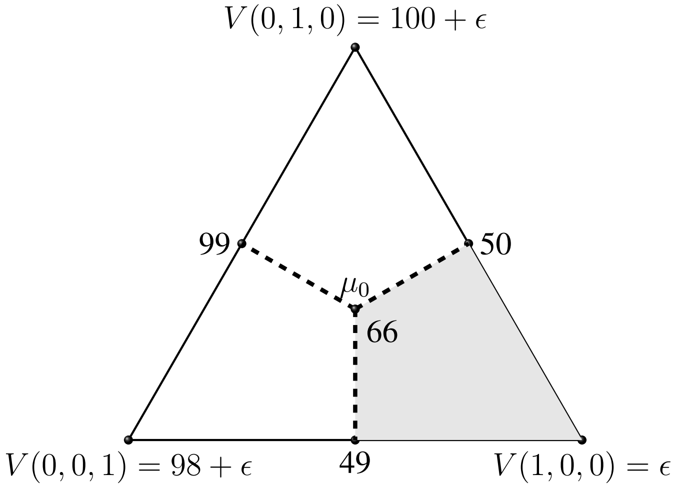

Figure 5: A Convex Payoff Function with Three States: The triangle represents the belief space, and the value at each vertex and midpoint is the sender's payoff. The lines that connect μ_0 and the midpoints delineate the set of posteriors that each signal can induce under Assumption 2, e.g, the gray area is the posteriors that signal s_{θ_1} can induce across all $\pi \in \Pi$.

Because $V(\cdot)$ is convex, the sender chooses the fully informative experiment in Bayesian persuasion. We show that the sender in our model does not choose the fully informative experiment as the true experiment. The fully informative experiment induces posteriors $(1,0,0)$, $(0,1,0)$, or $(0,0,1)$ with equal probability. Thus the sender's payoff following the optimal interpretation is at most

$$\frac{1}{3} \cdot 66 + \frac{1}{3} \cdot (100 + \epsilon) + \frac{1}{3} \cdot 99. \tag{5.2}$$

For example, if signal s_{θ_1} is realized, any posterior the sender can induce via interpretation belongs to the gray area in Figure 5. Thus 66 is the maximum payoff given s_{θ_1}. In contrast, suppose that the sender chooses experiment π^* such that $\pi^*(s_{\theta_2}|\theta_2) = 1$ and $\pi^*(s_{\theta_3}|\theta_1) = \pi^*(s_{\theta_3}|\theta_3) = 1$. This experiment induces (Bayesian) posterior $(0,1,0)$ with probability $1/3$, and posterior $(1/2,0,1/2)$ with probability $2/3$. The sender can then interpret signal s_{θ_3} according to π_3 such that $\pi_3(s_{\theta_3}|\theta_2) = \pi_3(s_{\theta_3}|\theta_3) = 1$, and $\pi_3(s_{\theta_3}|\theta_1) = 0$. The receiver adopts π_3 and interprets s_{θ_3} according to π_3. As a result, the receiver's posterior becomes $(0,1/2,1/2)$. Thus the sender's optimal payoff is at least

$$\frac{1}{3} \cdot (100 + \epsilon) + \frac{2}{3} \cdot 99,$$

which is strictly larger than (5.2), the best payoff conditional on choosing the fully informative experiment.

This example shows that even if the sender's payoff is convex, the true experiment could differ from full disclosure. Therefore the sender may choose a less informative experiment than in Bayesian persuasion when $|\Theta| > 2$, i.e., the information effect can be negative. The key driver is that the sender's payoff from state θ_1 is small ($V(1, 0, 0) = \epsilon$), so the true experiment pools states θ_1 and θ_3 together. By pooling them, the sender can "better" interpret signal s_{θ_3}.

6 Conclusion

We study a model in which the sender designs an experiment that generates a signal and then provides an interpretation of the realized signal. Strategic interpretation distorts the receiver's belief, but encourages the sender to generate more information ex ante. Under binary states, strategic interpretation could benefit both the sender and the receiver by improving the quality of information generated. A natural future direction is to allow an arbitrary state space and examine the potential tension between the information and interpretation effects more generally.

References

Aina, Chiara (2021), "Tailored stories."

Beauchêne, Dorian, Jian Li, and Ming Li (2019), "Ambiguous persuasion." *Journal of Economic Theory*, 179, 312–365.

Bloedel, Alexander W and Ilya R Segal (2018), "Persuasion with rational inattention." *Available at SSRN 3164033*.

Boutron, Isabelle, Douglas G Altman, Sally Hopewell, Francisco Vera-Badillo, Ian Tannock, and Philippe Ravaud (2014), "Impact of spin in the abstracts of articles reporting results of randomized controlled trials in the field of cancer: the spiin randomized controlled trial." *Journal of Clinical Oncology*, 32, 4120–4126.

Boutron, Isabelle, Susan Dutton, Philippe Ravaud, and Douglas G Altman (2010), "Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes." *JAMA*, 303, 2058–2064.

de Clippel, Geoffroy and Xu Zhang (2020), "Non-bayesian persuasion."

Di Tillio, Alfredo, Marco Ottaviani, and Peter Norman Sørensen (2017), "Persuasion bias in science: Can economics help?" *Economic Journal*, 127.

Eliaz, Kfir, Ran Spiegler, and Heidi C Thysen (2021), "Strategic interpretations." *Journal of Economic Theory*, 192, 105192.

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

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

Libgober, Jonathan (forthcoming), "False positives and transparency." *American Economic Journal: Microeconomics*.

Lipnowski, Elliot and Laurent Mathevet (2018), "Disclosure to a psychological audience." *American Economic Journal: Microeconomics*, 10, 67–93.

Lipnowski, Elliot, Laurent Mathevet, and Dong Wei (2020), “Attention management.” *American Economic Review: Insights*, 2, 17–32.

Lipnowski, Elliot, Doron Ravid, and Denis Shishkin (2019), “Persuasion via weak institutions.” *Available at SSRN 3168103*.

Schwartzstein, Joshua and Adi Sunderam (2021), “Using models to persuade.” *American Economic Review*, 111, 276–323.

Appendix

A Proof of Proposition 1

Proof. The proof of the first part consists of three steps. First, if the sender's experiment draws signal s and induces Bayesian posterior μ, the sender can interpret signal s to induce the receiver to hold any posterior $\mu^R \in [\min\{\mu_0, \mu\}, \max\{\mu_0, \mu\}]$. To see this, take any experiment π that induces Bayesian posteriors $\mu' < \mu_0$ and $\mu > \mu_0$ following $s = 0$ and $s = 1$, respectively. The ex ante probability of signal 1 is $\Pr(1|\pi) = \frac{\mu_0 - \mu'}{\mu - \mu'}$. Following signal 1, suppose the sender proposes an interpretation that induces posteriors μ' and $\mu^R \in [\mu_0, \mu]$, respectively. The ex ante probability of μ^R is $\frac{\mu_0 - \mu'}{\mu^R - \mu'} > \Pr(1|\pi)$, so the receiver adopts the interpretation and has posterior μ^R. The same argument applies to any $\mu^R \in [\mu', \mu_0]$.

Second, the sender can attain her optimal payoff with a strategy such that her interpretation moves the receiver's posterior toward the prior, compared to the Bayesian posterior. The statement holds when the sender chooses the true experiment that discloses no information or full information. Suppose the sender chooses experiment π that induces posteriors $\mu' < \mu_0$ and $\mu > \mu_0$. Without loss of generality, suppose $\frac{\mu - \mu_0}{\mu - \mu'} > 1 - \mu_0$. Consider another experiment $\hat{\pi}$ that induces posteriors $y < \mu'$ and 1 such that $\frac{1 - \mu_0}{1 - y} = \frac{\mu - \mu_0}{\mu - \mu'}$, or equivalently, $y = 1 - \frac{(1 - \mu_0)(\mu - \mu')}{\mu - \mu_0}$. We have $y > 0$ because it reduces to $\frac{\mu - \mu_0}{\mu - \mu'} > 1 - \mu_0$. Experiments π and $\hat{\pi}$ have the identical ex ante probability of each signal $s \in \{0, 1\}$. Thus the set of possible posteriors the sender can induce is the same between π and $\hat{\pi}$. We now consider the sender's optimal interpretation under $\hat{\pi}$. We show that the sender's interpretation can only move the receiver's posterior toward the prior. The statement trivially holds when $\hat{\pi}$ induces posterior 1. The statement also holds following posterior y: The maximum ex ante probability for posterior $x < y$ (across all feasible experiments) is $\frac{1-\mu_0}{1-x} < \Pr(y|\hat{\pi}) = \frac{1-\mu_0}{1-y}$, and thus the sender cannot induce x following Bayesian posterior y. To sum up, even if the sender can only use interpretations that move Bayesian posteriors toward the prior, her optimal payoff remains the same.

Third, we show that the sender's optimal payoff comes from the concavification of the μ_0-monotone hull of V. Suppose the sender can only use interpretations that move Bayesian posteriors toward the prior. The first step implies that the sender can indeed attain any posterior between the prior and the Bayesian posterior. Thus given Bayesian posterior μ, the sender's optimal payoff in such a situation is $V^M(\mu)$. The sender can choose an experiment to induce any Bayes-plausible distribution of Bayesian posteriors (with at most two posteriors). Therefore the sender's optimal payoff is the concavification of V^M evaluated at μ_0.

We now prove the second part. Take any optimal strategy of the sender. Suppose that the associated experiment induces Bayesian posteriors $\mu^- \leq \mu_0$ and $\mu^+ \geq \mu_0$. If $\mu^- = \mu^+ = \mu_0$, Part 1 holds. Suppose $\mu^- < \mu_0$ and $\mu^+ > \mu_0$. Assume $V^M(\mu^-) \leq V^M(\mu^+)$. Suppose that the sender instead chooses an experiment with Bayesian posteriors 0 and μ^+. We have $V^M(0) \geq V^M(\mu^-)$, and the new strategy increases the probability of attaining $V^M(\mu^+)$. Thus the sender can also attain her optimal payoff with the new strategy. We now consider the sender's optimal interpretation having induced Bayesian posteriors 0 and μ^+. Suppose to the contrary that $V^M(\mu^+) > V(\mu^+)$, that is, the sender interprets posterior μ^+ to distort the receiver's belief. It implies that there is some $\mu^* \in [\mu_0, \mu^+)$ such that $V(\mu^*) = V^M(\mu^+)$. The sender could then induce Bayesian posteriors 0 and μ^* to attain payoffs $V^M(0)$ and $V(\mu^*) = V^M(\mu^+)$ so that the probability of $V^M(\mu^+)$ is now $\frac{\mu_0}{\mu^*} > \frac{\mu_0}{\mu^+}$. This is a contradiction, and thus the sender honestly intepets signal 1 (that leads to posterior μ^+), i.e., Part 2 holds. Symmetrically, if $V^M(\mu^-) \geq V^M(\mu^+)$, Part 3 holds. □

B Proof of Proposition 2

Proof. For any posteriors μ and μ', we have $U(\mu, \mu') \leq U(\mu, \mu)$. Letting $(\mu, \mu') = (\tilde{\mu}^T, \tilde{\mu}^R)$ and taking expectation, we obtain $\mathbb{E}[U(\tilde{\mu}^T, \tilde{\mu}^R)] - \mathbb{E}[U(\tilde{\mu}^T, \tilde{\mu}^T)] \leq 0$, i.e., the interpretation effect is non-positive.

Next, we show that the sender in our model chooses a more informative experiment as the true experiment than in Bayesian persuasion. The statement holds if the solution of Bayesian persuasion is to provide no information. Suppose that a solution of Bayesian persuasion induces posteriors $\mu^- < \mu_0$ and $\mu^+ > \mu_0$. We consider two cases. First, suppose there is an optimal strategy in our model such that the receiver's belief equals μ_0 with probability 1. The sender can implement such an outcome by choosing a fully informative experiment and interpreting signals so that the receiver has posterior μ_0 with probability 1. The fully informative experiment is a part of the sender's optimal strategy and is more informative than any solution of Bayesian persuasion.

Second, suppose there is no optimal strategy such that the receiver's belief equals μ_0 with probability 1. Then either Part 2 or Part 3 of Proposition 1 holds. Without loss of generality, suppose that Part 2 holds. That is, there is an optimal strategy such that the true experiment induces Bayesian posteriors 0 and $\mu \in (\mu_0, 1]$, and the sender honestly interprets signal 1, which induces μ. Suppose that $\mu \in (\mu_0, \mu^+)$. We construct another optimal strategy of the sender such that the true experiment is weakly more informative than the solution to Bayesian persuasion. First, we have

$$\frac{\mu - \mu^-}{\mu} V^M(0) + \frac{\mu^-}{\mu} V^M(\mu) \geq V^M(\mu^-) \geq V(\mu^-). \tag{6.1}$$

The first inequality holds; otherwise the concavification of V^M would include μ^-, which means that the sender would strictly prefer to induce Bayesian posteriors μ^- and μ, instead of 0 and μ. Second, we have

$$\frac{\mu^+ - \mu}{\mu^+ - \mu^-} V(\mu^-) + \frac{\mu - \mu^-}{\mu^+ - \mu^-} V(\mu^+) \geq V(\mu) = V^M(\mu). \tag{6.2}$$

The first inequality holds; otherwise the concavification of V would include μ. The equality holds because the sender honestly interprets signal 1, which leads to Bayesian posterior μ. We can write inequality (6.2) as

$$V(\mu^-) \geq \frac{\mu^+ - \mu^-}{\mu^+ - \mu} V^M(\mu) - \frac{\mu - \mu^-}{\mu^+ - \mu} V(\mu^+). \tag{6.3}$$

Inequalities (6.1) and (6.3) imply

$$\frac{\mu-\mu^-}{\mu}V^M(0)+\frac{\mu^-}{\mu}V^M(\mu)\geq\frac{\mu^+-\mu^-}{\mu^+-\mu}V^M(\mu)-\frac{\mu-\mu^-}{\mu^+-\mu}V(\mu^+).$$

Rearranging this, we have

$$\frac{\mu^+-\mu}{\mu^+}V^M(0)+\frac{\mu}{\mu^+}V(\mu^+)\geq V^M(\mu).$$

Finally, multiply both sides by $\frac{\mu_0}{\mu}$ and rearrange the terms:

$$\frac{\mu^+-\mu_0}{\mu^+}V^M(0)+\frac{\mu_0}{\mu^+}V(\mu^+)\geq\frac{\mu-\mu_0}{\mu}V^M(0)+\frac{\mu_0}{\mu}V^M(\mu).$$

Because the right-hand side is the sender's optimal payoff, this inequality implies that it is also optimal for the sender to induce Bayesian posteriors 0 (after signal 0) and μ^+ (after signal 1 and honestly interpret signal 1. Under such a strategy, the true experiment is more informative than the solution of Bayesian persuasion. Therefore the information effect is non-negative. □

C Proof of Proposition 3

Proof. Suppose F is strictly concave, so that Bayesian persuasion leads to no disclosure. As the right panel of Figure 2 shows, in our model, the sender chooses the true experiment that induces Bayesian posteriors μ_H (after signal 1) and 0 (after signal 0) with probabilities $\frac{\mu_0}{\mu_H}$ and $\frac{\mu_H-\mu_0}{\mu_H}$, respectively. The sender honestly interprets signal 1, but interprets signal 0 to induce subjective posterior μ_0 as opposed to 0. As a result we can write the receiver's ex ante expected payoff as

$$\frac{\mu_0}{\mu_H}F(\mu_H)[\mu_H-\mathbb{E}(c|c\leq\mu_H)]+\left(1-\frac{\mu_0}{\mu_H}\right)F(\mu_0)[-\mathbb{E}(c|c\leq\mu_0)].$$

We show this expression is greater than the receiver's payoff under Bayesian persuasion, which is $F(\mu_0)[\mu_0 - \mathbb{E}(c|c \leq \mu_0)]$. We obtain

$$\begin{aligned}
&\frac{\mu_0}{\mu_H} F(\mu_H)[\mu_H - \mathbb{E}(c|c \leq \mu_H)] + \left(1 - \frac{\mu_0}{\mu_H}\right) F(\mu_0)[-\mathbb{E}(c|c \leq \mu_0)] > F(\mu_0)[\mu_0 - \mathbb{E}(c|c \leq \mu_0)] \\
\iff &\frac{\mu_0}{\mu_H} F(\mu_H)[\mu_H - \mathbb{E}(c|c \leq \mu_H)] - \frac{\mu_0}{\mu_H} F(\mu_0)[-\mathbb{E}(c|c \leq \mu_0)] > F(\mu_0)\mu_0 \\
\iff &\frac{\mu_0}{\mu_H} F(\mu_H)[\mu_H - \mathbb{E}(c|c \leq \mu_H)] - \frac{\mu_0}{\mu_H} F(\mu_0)[\mu_H - \mathbb{E}(c|c \leq \mu_0)] > 0 \\
\iff &F(\mu_H)[\mu_H - \mathbb{E}(c|c \leq \mu_H)] - F(\mu_0)[\mu_H - \mathbb{E}(c|c \leq \mu_0)] > 0.
\end{aligned}$$

Function $h(x) = F(x)[\mu_H - \mathbb{E}(c|c \leq x)]$ is strictly increasing in $x < \mu_H$, because we have $h'(x) = \mu_H f(x) - x f(x) > 0$ for any $x < \mu_H$. As a result, the above sequence of inequalities holds. Therefore the receiver is strictly better off when the sender strategically interprets the signals.

If F is strictly convex, the sender discloses full information under Bayesian persuasion. In our model the sender chooses the fully informative experiment as the true experiment, but she interprets signal 0 to make the receivers believe that the posterior is $\mu = \mu_0$ instead of $\mu = 0$. On this event, the receiver is strictly worse off. Thus the receivers' welfare is strictly lower in our model than in Bayesian persuasion. □