Bayesian Persuasion with Confirmation Bias

Mengxi Sun

August 2022

Abstract

This paper studies a behavioral model of Bayesian Persuasion, where the Receiver with confirmation bias may misinterprets the signal realizations depending on the prior beliefs. As opposed to other erroneous inferences or distortions of the rational updating rule, confirmation bias distorts the information structure and hence the posterior beliefs and distribution. With a binary pure persuasion example, I find that if the Receivers correctly incorporate the bias in belief updating process, the sophistication serves as a buffer protecting the disadvantaged side due to the confirmation bias: (1) when the Receiver holds a prior with bias beneficial to the Sender, the Receivers is harder to convince, and hence the optimal signal reveals more information; and (2) when the Receiver holds a prior with bias detrimental to the Sender, the Receiver is easier to convince. This result of strategic adjustment is bounded by the relative position of the prior and posterior beliefs and the confirmation bias.

1 Introduction

When we theoretically study communications between players, we assume that they at least agree on how to interpret the messages. This consensus enables us to label the messages in a consistent way that simplifies the setting. However, oftentimes the signals in real-world communication have various degrees of flexibility in interpretation, and we evaluate information always relative to something else. The information consumers more or less can choose the reference points depending on their previous experience when making such judgments. There are several other perspectives to look at the deviation from rationality depending on priors. I want to focus on one–the confirmation bias–in this paper.

The persuasion models grant the information designer (henceforth, the Sender) the power to induce desirable beliefs based on the predictability of the information consumers' (henceforth, the Receiver) behaviors. What if the Sender can only anticipate a distribution of behaviors rather than deterministic actions? The confirmation bias of the Receivers introduces a behavioral uncertainty into the information design. On the one hand, since we conventionally praise rationality, the Sender could take advantage of the Receiver's irrationality. On the other hand, it could also be that the irrationality somehow protects the Receiver from the powerful manipulation of the Sender because the informational incentives change through the lens of confirmation bias. Even after removing the strategic components, it is still unclear how the effects of confirmation bias work in persuasion.

To analyze the strategic responses of the Sender, think about a scenario where a proposer tries to convince a decision maker to approve her proposal. The proposer and the decision maker share common prior beliefs but have conflicts in interest. A decision maker has confirmation bias as defined in Rabin and Schrag (1999). If the decision maker believes that the proposal is less likely to be good, he is skeptical and may misinterpret a positive signal as a negative one when receiving new information; vice versa, if the decision maker has prior beliefs that the proposal is more likely to be a good one, he is optimistic and may update beliefs in favor of the proposer despite a signal intended to be negative. Compared to rationality and mistakes in inference, the confirmation bias distorts the perception of the designed information structure due to asymmetric misinterpretation of the signal realizations. Consequently, the strategic effect works against the force of confirmation bias when the decision maker adjusts to the fact that he sees a signal realization more or less than fully rational. Can the Sender still achieve the best Bayesian persuasion value? Does the Receiver make ex-post correct decision more or less often?

1.1 Related Literature

I want to emphasize that the Receiver with confirmation bias is still Bayesian and updates beliefs according to Bayes' rule. My model studies irrationality not from the non-Bayesian perspective as in the behavioral models of erroneous inference because confirmation bias is a mistake in perception of the information structure, which occurs before inference. In general, the non-Bayesian models investigate discrepancies only in posterior beliefs but not in posterior distributions. Most representatively, de Clippel and Zhang (2022) theorizes a group of inference mistakes as the systematic distortion of the Receivers' beliefs and characterizes the underlying belief updating rules to be homogeneous of degree zero belief distortion functions. The authors also identified the sufficient and necessary conditions of the systematic distortions for the revelation principle to hold and elaborated on when the Sender benefits from persuasion. The confirmation bias in this paper is out of the scope of de Clippel and Zhang (2022) because it cannot be reduced to a simple belief updating rule that maps a signal realization to a posterior belief. More substantially, the posterior beliefs of a Receiver with confirmation bias deviates from rationality attributed solely to the misperception of the information structure. Thus, the posterior outcome departs not only in the beliefs as under the de Clippel and Zhang (2022)'s model but also in the distributions.

As a result, the systematic distortions of the Receiver's beliefs in de Clippel and Zhang (2022) are equivalent to transforming the Sender's value function. However, the effect of confirmation bias cannot be represented merely by the change of the Sender's value function since the outcome depends on the relative position of the prior and posterior beliefs.

Although a belief distortion function cannot fully characterized the confirmation bias, the mapping between the rational beliefs and the confirmatorily biased beliefs is projective, consistent with de Clippel and Zhang (2022)'s result on the revelation principle.

Another related paper Lee et al. (2021) studies the prior biased inference in a cheap talk game. The prior biased belief updating behavior is covered in the systematic distortion in de Clippel and Zhang (2022). It captures the distrust of the Receiver to the Sender's information as a prior-biased belief updating rule. The Receiver mistakenly thinks there is some positive probability where the Sender obtains no information but sends a message according to the unconditional distribution of the equilibrium signal. The prior biased inference treats all signal realizations with skepticism. Hence, the Receiver with prior bias always update beliefs midway between the prior and the rational posterior. The authors find that prior bias may be welfare-improving due to this dampening effect of the prior bias in communication. As a contrast, the confirmation bias treats signal realization differently and the posterior beliefs spread out more than the prior biased beliefs. So, the confirmation bias in communication is motivated by a fundamentally distinct behavior from the prior biased inference, which not only preserves current beliefs but also may polarize beliefs further. The effect of confirmation bias cannot be inferred from Lee et al. (2021)'s intuition and requires specific analysis.

Lockwood (2017) models a political agency problem in the presence of confirmation bias and finds that confirmation bias reduces the incumbent's pandering behavior and, in turn, improves voter welfare. The paper analyzes two types of voters. One type suffers confirmation bias but is unaware of the error, which correspond to the Naive type in my model; the other type treats the error rationally as if receiving a biased noisy signal, which correspond to the Sophisticated type in my model. The author finds that the politician panders less with the presence of confirmation bias due to misinterpretation; when the voter is unaware of the error, two opposing effects come into play. I decompose the effect of confirmation bias in another way. I find that both types move to the posterior belief closer to the prior more easily comparing to the fully rational type, but the Sophisticated type triggers the strategic effect that work against the bias.

Lastly, since I interpret the viability of confirmation bias to be flexible interpretations, Eliaz et al. (2021) provides some context in this environment. It models the variant interpretation of the evidence driven by a decipher from the Sender or a third party interpreter. The control over interpretation of messages expands the Sender's informational power since the Receiver only partially understand the Sender's strategy constrained by the decipher. As a result, full persuasion is possible. The authors also discuss the case of a third party interpreter with different motives. Here my model discusses a situation where no player controls the interpretation of the signal since confirmation bias is an inherent cognitive limitation.

In the following sections, I will first characterize the confirmation bias for the Kamenica and Gentzkow (2011) (henceforth, KG)'s concavification technique to work and then illustrate my findings with a binary pure persuasion example. Then, I will .

2 Model

The general framework follows KG's persuasion setup. There is a finite state space Ω and a compact action space A. Let $u(a, \omega)$ and $v(a, \omega)$ denote the utility functions of the Receiver and the Sender respectively, where $a \in A$ and $\omega \in \Omega$. Both players share a common prior about the state, $\mu_0 \in int(\Delta\Omega)$. The Sender removes the strategic component from the communication environment by committing to a signal $\{\pi(s|\omega)\}_{\omega\in\Omega}$ over a finite signal realization space S.

The only deviation from the full rationality of my model is the Receiver's confirmation bias. In this behavioral model, the Receiver might see a different signal realization from the one designed by the Sender. Specifically, for each Receiver with a prior belief μ_0, given a signal π over a finite signal realization S that corresponds to a set of posterior beliefs $\{\mu(s)\}_{s\in S}$, there is a probability of misinterpreting a signal realization s as another signal realization s' if the induced posterior belief $\mu(s')$ is closer to the prior belief μ_0 than $\mu(s)$ is. One immediate result of this bias is that even the public messages polarize opinions because of individual interpretive variations. Depending on whether the Receiver is aware of his bias or not, the posterior beliefs can deviate from the rational ones even though the belief updating rule remains to be the Bayes' rule. This is because the Receiver perceive the information differently.

2.1 Sophisticated Receiver

If the Receiver is aware of the bias, he incorporates his irrationality in a way that he updates beliefs according to the probability of seeing a signal realization instead of the probability of a signal realization being sent. The confirmation bias connects the Sender-designed information structure π to the probability of a Receiver seeing a signal realization ϕ as the following:

$$\phi(s|\omega, \pi) = \sum_{s' \in supp(\pi)} \pi(s'|\omega) \times \gamma(s|s'). \tag{1}$$

$\gamma(s|s')$ measures the probability of a Receiver seeing a signal s when a signal s' is designed to be sent publicly by the Sender. When $s \neq s'$, $\gamma(s|s')$ represents the confirmation bias; $\gamma(s|s) = 1 - \sum_{s' \in supp(\pi) \backslash s} \gamma(s'|s)$ captures the probability of the signal passing through as designed. For the sake of representation, I assume confirmation bias is relatively small such that $\gamma(s|s') < 0.5$ for all $s \neq s'$ and $\gamma(s|s) = 1 - \sum_{s' \in supp(\pi) \backslash s} \gamma(s'|s) > 0.5$ for all s. This assumption ensures that the signal is interpreted as designed by the Sender most of the time.

Hence, by replacing π with ϕ, the posterior belief $\mu(\cdot|s)$ and the signal π induced posterior distribution $\tau(\cdot)$ in my model are

$$\mu^{R_S}(\omega|s) = \frac{\phi(s|\omega, \pi)\mu_0(\omega)}{\sum_{\omega' \in \Omega} \phi(s|\omega', \pi)\mu_0(\omega')} \tag{2}$$

and

$$\tau^R(\mu) = \tau^S(\mu) = \sum_{s: \mu^{R_S}(\cdot|s) = \mu} \sum_{\omega' \in \Omega} \phi(s|\omega', \pi)\mu_0(\omega'). \tag{3}$$

The Sender and the Receiver understand the Receiver's beliefs equally. [To establish that any Bayes plausible beliefs can be implemented by some information structure, I need to solve the condition on the confirmation bias parameters]

2.2 Naive Receiver

If the Receiver is oblivious about the bias, he will treat what he sees as the designed information structure. However, the Sender knows the distinction between ϕ and π. As a result, the Receiver's belief updating rule is the same as the Bayes' rule in the fully rational case, but the distribution of induced posterior is calculated with ϕ instead of π:

$$\mu^{R_N}(\omega|s) = \frac{\pi(s|\omega)\mu_0(\omega)}{\sum_{\omega'\in\Omega}\pi(s|\omega')\mu_0(\omega')} \tag{4}$$

$$\tau^R(\mu) = \sum_{s:\mu^{R_N}(\cdot|s)=\mu}\sum_{\omega'\in\Omega}\pi(s|\omega')\mu_0(\omega') \tag{5}$$

$$\tau^S(\mu) = \sum_{s:\mu^{R_N}(\cdot|s)=\mu}\sum_{\omega'\in\Omega}\phi(s|\omega',\pi)\mu_0(\omega'). \tag{6}$$

Despite misunderstanding the true posterior distribution, the Naive Receiver's posterior beliefs are still Bayes plausible with respect to his understanding of the beliefs. So, any Bayes plausible Naive Receiver's posterior belief can be implemented by some information structure.

Proposition 1 *(Bayes Plausibility)*

1. *If the Receiver correctly takes his confirmation bias γ into account when updating beliefs, the distribution of posteriors is Bayes -plausible with respect to the probability of the Receiver seeing the signal realizations ϕ, which relates to the Sender-designed information structure π via γ; that is, $\sum_{supp(\tau^R)}\mu^{R_S}\cdot\tau^R(\mu^{R_S}) = \mu_0$.*

2. *If the Receiver is naive that he does not incorporate his confirmation bias γ when updating beliefs, the distribution of posteriors is Bayes-plausible with respect to the Sender-designed information structure π; that is, $\sum_{supp(\tau^R)}\mu^{R_N}\cdot\tau^R(\mu^{R_N}) = \mu_0$.*

To summarize, KG's requirement on Bayes plausibility approximately holds with respect to the Receiver's knowledge as long as the the Receiver updates beliefs and estimates posterior distribution consistently.

3 A Binary Example

To further analyze the effects of confirmation bias on the persuasion outcome, I illustrate the findings with a simple binary example. Imagine a pure persuasion problem of one Sender and one Receiver. The Sender propose a plan to the Receiver and the Receiver can either approve or reject, $A = \{a, r\}$. The state of the world is either high or low, $\omega \in \Omega = \{H, L\}$. The ex-ante probability of high state μ_0 is common knowledge. The Receivers want to approve the proposal in the high state and reject the proposal in the low state. The Sender maximizes the probability of getting approved, independent of the state. She gets 1 if the Receiver approves and 0 if the Receiver rejects.

The baseline without confirmation bias can be solved just like the prosecutor and judge example in KG. Since the payoffs of the Sender is a step function of beliefs, the optimal signal can be simplified as π^* over $S = \{s_L, s_H\}$ inducing either a revealing low posterior belief of $\mu(s_L) = 0$ or an approving high posterior belief $\mu(s_H) = \tilde{\mu}$, where $\tilde{\mu}$ is the threshold belief that the Receiver is indifferent between approving and rejecting. Thus, the baseline optimal signal is that $\pi^*(s_H|H) = 1$ and $\pi^*(s_H|L) = \frac{\mu_0(1-\tilde{\mu})}{\tilde{\mu}(1-\mu_0)}$. Thus, the maximized probability of inducing the approving posterior belief is $\tau^* = \frac{\mu_0}{\tilde{\mu}}$, which equals to the Sender's ex-ante maximized expected payoff.

By the arguments of Proposition **??** and **??**, if the Receiver is consistent when updating beliefs and estimating posterior distribution of his beliefs, his posterior beliefs are Bayes-plausible with respect to his knowledge. Then the Sender can choose the posterior distribution she desire instead of solve the optimal information structure since she only concerns the Receiver's beliefs. So, the optimal persuasion with confirmation bias can also be supported by a signal of binary realizations with s_L and s_H. I parameterize the confirmation bias as $\{\gamma_s\}_{s \in \{s_H, s_L\}}$ with γ_H being the probability of misinterpreting s_H as s_L and γ_L being the probability of misinterpreting s_L as s_H. Then, the probability of seeing s in state ω given a Sender's chosen signal $\pi(s|\omega)_{s \in \{s_L, s_H\}}$ is that $\pi(s|\omega) \times (1 - \gamma_s) + \pi(-s|\omega) \times \gamma_{-s}$. With binary signal realizations, when the prior belief μ_0 is closer to the targeting high posterior belief, only γ_L takes effect and vice versa. In the next subsections, I will identify when each γ_s comes into force and discuss the cases where γ_L or γ_H takes effect separately.

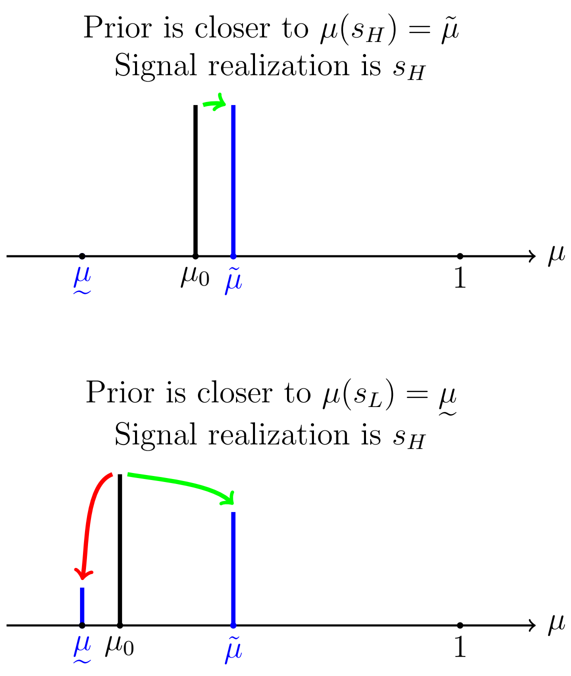


Figure 1: Direction of Misinterpretation
Interpret as designed
Misinterpret

3.1 Direction of confirmation bias

First of all, I need to know when the prior belief is closer to the targeting low posterior belief and when it is closer to the targeting approval posterior belief. The threshold depends on both the relative position of the prior and the posterior beliefs and the probability of misinterpretation.

Lemma 2 *For* $\gamma_H < \frac{1}{2}$*,*

- *When* $0 < \mu_0 < \frac{\tilde{\mu}}{2}\frac{1}{1-\gamma_H}$*, p is closer to the low targeting posterior belief.*

- *When $\frac{\tilde{\mu}}{2}\frac{1}{1-\gamma_H} < \mu_0 < \tilde{\mu}$, p is closer to the high targeting posterior belief.*

Since $\frac{\tilde{\mu}}{2}\frac{1}{1-\gamma_H}$ is increasing in γ_H, Lemma 2 indicates that, in additional to the probability of misinterpretation, γ_H also hurts the Sender by increasing the interval of priors that may misinterpret the high signal realization s_H as low signal realization s_L. This is because the targeting low posterior belief is on the boundary so that s_L can no longer be fully revealing when the Receiver misinterpret s_H.

3.2 High prior

When μ_0 is closer to the targeting approval posterior belief $\tilde{\mu}$, only the low signal s_L may be misinterpreted. If the Receivers are correctly aware of their confirmation bias and consider this irrationality when they update belief, the Sender's optimal signal becomes $\pi^{*R_S}(s_H|H) = 1$, which is same as the baseline, and $\pi^{*R_S}(s_H|L) = \frac{\mu_0(1-\tilde{\mu})}{\tilde{\mu}(1-\mu_0)}\frac{1}{1-\gamma_L} - \frac{\gamma_L}{1-\gamma_L}$, which is lower than the baseline. Consequently, the maximized probability of sending the high signal s_H decreases to $\eta^{*R_S} = \frac{\mu_0}{\tilde{\mu}}\frac{1}{1-\gamma_L} - \frac{\gamma_L}{1-\gamma_L}$ from the baseline's $\tau^* = \eta^* = \frac{\mu_0}{\tilde{\mu}}$. Therefore, the Sender's ex-ante maximized probability of winning is $\tau^{*R_S} = \eta^{*R_S} + (1-\eta^{*R_S}) \times \gamma_L = \frac{\mu_0}{\tilde{\mu}}$.

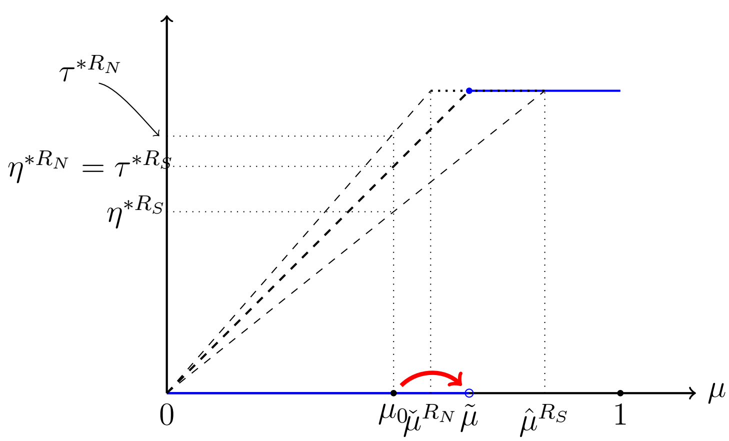


Figure 2: One Receiver with high prior misinterprets s_L w.p. γ_L

Result 3 *For high μ_0, comparing to the full rationality case,*

- *the Sophisticated Receiver is harder to convince and the Sender can achieve the concavification value by revealing more information;*
- *to persuade a Naive Receiver, the optimal signal remains unchanged as in the full rationality case and the Sender can achieve a higher maximized due to confirmation bias in favor of the Sender.*

The sophisticated Receivers are more difficult to convince because they know that they see the high signal realization more often than the signal realization is actually sent. Hence, to support the same approval posterior belief $\tilde{\mu}$, the Sender has to reveal more information and reduce the probability of sending the high signal s_H when the state is low $\omega = L$. So, the optimal signal with confirmation bias equates the probability of the sophisticated Receivers seeing the high signal realization s_H to the level of the probability of sending s_H without confirmation bias.

As a contrast, the naive Receivers thinks that he is rational and hence the Sender treats him like one in the pure persuasion problem. However, the probability of inducing the approval posterior belief $\tau^{*R_N} = \frac{\mu_0}{\tilde{\mu}} + \frac{\tilde{\mu}-\mu_0}{\tilde{\mu}}\gamma_L$ is higher than both the Sophisticated Receiver and the Rational Receiver cases. Although the awareness of the bias does not help the Receivers more than the full rationality, the Naive Receiver make good decisions less often than the Sophisticated Receiver.

3.2.1 Low prior

Similarly, when μ_0 is closer to the targeting low posterior belief, only the high signal s_H may be misinterpreted. Then with sophisticated Receiver, the Sender's optimal signal is the same as the baseline, $\pi^{*R_S}(s_H|H) = 1$ and $\pi^{*R_S}(s_H|L) = \frac{\mu_0(1-\tilde{\mu})}{\tilde{\mu}(1-\mu_0)}$. The high realization s_H can still induce the approval posterior belief $\tilde{\mu}$. However, the low realization s_L can only move the posterior belief as low as $\underset{\sim}{\mu} = \mu(s_L;\gamma_H) = \frac{\tilde{\mu}\gamma_H}{\frac{\tilde{\mu}}{\mu_0}-(1-\gamma_H)}$, which is higher than 0 in the baseline. Since the optimal signal under the effect of γ_H remains unchanged as the baseline, the probability of sending the high signal s_H also remains the same $\eta^{*R_S} = \frac{\mu_0}{\tilde{\mu}} = \eta^*$. The Sender's maximized ex-ante expected probability of approval is $\tau^{*R_S} = \eta^{*R_S} \times (1-\gamma_H)$.

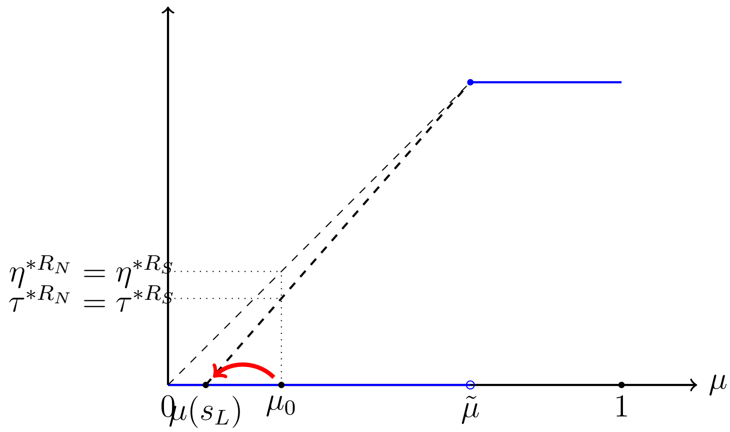


Figure 3: One Receiver with low prior misinterprets s_H w.p. γ_H

Result 4 *For low μ_0, comparing to the full rationality case,*

- *the Sophisticated Receiver is easier to convince but since the rational posterior belief is not interior, there is not room for strategic adjustment and hence the Sender gets less than the concavification value;*
- *to persuade a Naive Receiver, the Sender's optimal signal remains unchanged as in the full rationality case and the maximized payoff coincides with the Sophisticated case.*

The Sophisticated Receiver with low prior sees low signal more often than intended. From the Sender's perspective, it should be strategically easier to persuade. However, to sustain the targeting approval posterior, the Sender cannot reveal less information because the targeting low posterior is on the boundary.

With Naive Receivers, the Sender's optimal signal is also the same as in the baseline. The probability of inducing the approval posterior belief $\tau^{*R_N} = \frac{\mu_0}{\tilde{\mu}}(1-\gamma_H)$ is lower than the base line $\tau^* = \frac{\mu_0}{\tilde{\mu}}$ as the naive Receivers underestimate the value of s_H. The Sender's maximized payoffs match with the two types because the Sender loses the strategic benefit from the boundary restriction.

4 Extensions

4.1 Compact action space

The results of the above binary example extend naturally to the compact action space, where the value function of the Sender has a general form.

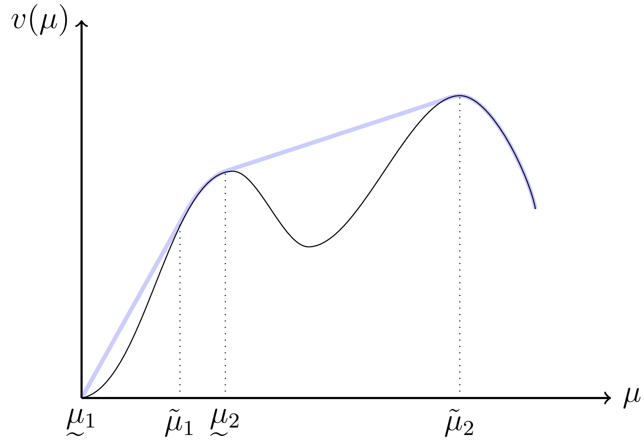


Figure 4: general value function

Proposition 5 *Assuming binary states, state-independent sender's payoff, and Sophisticated Receiver,*

1. *When the targeting posterior beliefs $\tilde{\mu}$ and $\underset{\sim}{\mu}$ are bounded away from the boundary beliefs s.t.*

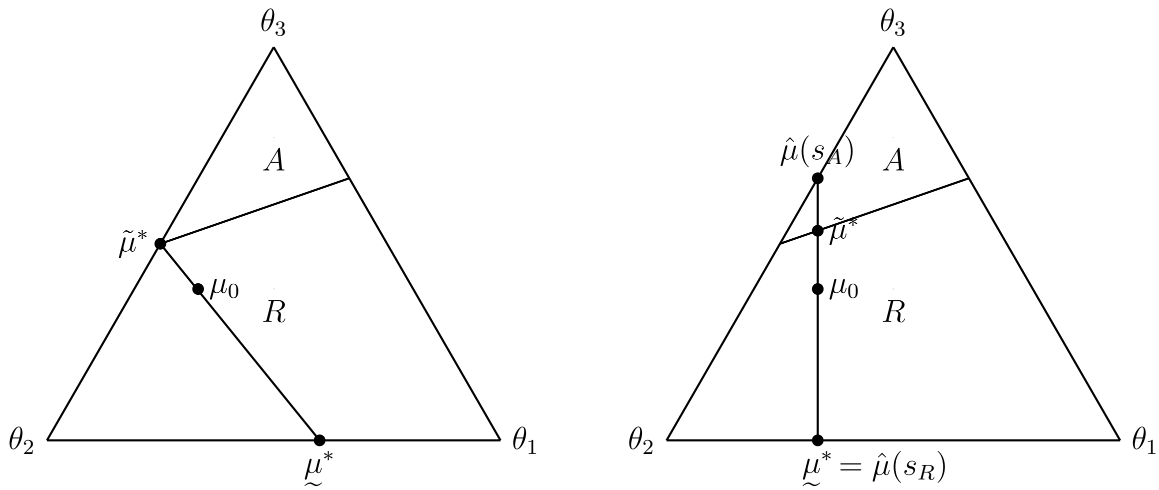
$$\frac{\tilde{\mu}(\mu_0 - \underset{\sim}{\mu})}{\mu_0(\tilde{\mu} - \underset{\sim}{\mu})} \leq 1 - \gamma_H \text{ and } \frac{\gamma_L}{1-\gamma_L} \leq \frac{\underset{\sim}{\mu}(\tilde{\mu} - \mu_0)}{\mu_0(\tilde{\mu} - \underset{\sim}{\mu})},$$

   the Sender can achieve the concavification value by adjusting the informational strategy to the bias.

2. *When the targeting posterior beliefs and the prior are close to the boundary beliefs, the informational strategy is constrained and the Sender gets less than the concavification value.*

4.2 Multiple states binary actions

The results can also extend to beyond binary states. In multiple state world, the optimal signal with binary realizations inducing the posterior beliefs $\tilde{\mu}$ and $\underset{\sim}{\mu}$ maximizes the ratio of the Euclidean distance $\frac{d(\mu_0,\underset{\sim}{\mu})}{d(\mu_0,\tilde{\mu})}$ (Alonso and Câmara, 2016). When the prior of a Sophisticated Receiver is far away from the approval set A, the optimal signal remains the same as the rational case. However, when the prior is close to the approval set such that the Receiver may misinterpret the realization inducing $\underset{\sim}{\mu}$, the Sender needs to reveal more information to sustain $\tilde{\mu}$ in the approval set. Thus the optimal signal with confirmation bias need to satisfy an additional condition. Given the optimal information structure that induces $\tilde{\mu} \in A$ and $\underset{\sim}{\mu} \in R$, the rational posterior $\hat{\mu}$ must be a valid belief in the approval set, as shown in Figure 5 (b).



(a) Rational Receiver (b) Sophisticated Receiver

Figure 5: multiple state binary actions

Proposition 6 *(Extending Alonso and Câmara (2016) Proposition 1) Let $d(\mu_0, \mu)$ be the (Euclidean) distance between the prior μ_0 and the posterior belief μ. Every optimal experiment with a binary realization space targets to induce posterior beliefs $\tilde{\mu}^*$ and $\underset{\sim}{\mu}^*$ that* *maximize the ratio of the distances*

$$\frac{d(\mu_0, \underset{\sim}{\mu}^*)}{d(\mu_0, \tilde{\mu}^*)} = \max \frac{d(\mu_0, \underset{\sim}{\mu})}{d(\mu_0, \tilde{\mu})}$$

subject to $\tilde{\mu}^* \in A$ *and* $\underset{\sim}{\mu}^* \in R$*, the rational posterior belief* $\hat{\mu}(s_A)$ *and* $\hat{\mu}(s_R)$ *in the simplex, with* $\{\underset{\sim}{\mu}^*, \mu_0, \tilde{\mu}^*, \hat{\mu}(s_A), \hat{\mu}(s_R)\}$ *collinear.*

The additional restriction on the rational posterior belief guarantees the room for strategic adjustment on the confirmation bias. Note that when the receiver approves, he must be indifferent. The targeted approving posterior belief is on the boundary between the approval set A and rejection set R. This new condition is relevant when the prior is closer to the approval set A because a Sophisticated Receiver misinterpreting the rejecting signal updates beliefs toward the approval set less than the fully rational one. Thus the rational belief $\hat{\mu}$ need to be "exaggerated" so that the induced belief $\tilde{\mu}(s_A)$ can reach the targeted approving posterior belief $\tilde{\mu}*$.

On the other hand, since the targeted rejecting posterior belief $\underset{\sim}{\mu}$ is on the boundary of the simplex, the rational posterior $\hat{\mu}(s_R)$ must coincide with it. Whether the Sender can achieve the concavification value in the rational case depends on how close the targeted posterior beliefs are to the boundary of the simplex.

5 Discussion

References

Ricardo Alonso and Odilon Câmara. Persuading voters. *The American Economic Review*, 106 (11):3590–3605, 2016. ISSN 00028282, 19447981. URL http://www.jstor.org/stable/24911317.

Geoffroy de Clippel and Xu Zhang. Non-bayesian persuasion. *Journal of Political Economy*, 130(10):2594–2642, 2022. doi: 10.1086/720464. URL https://doi.org/10.1086/720464.

Kfir Eliaz, Rani Spiegler, and Heidi Christina Thysen. Strategic interpretations. *Journal of Economic Theory*, 192:105192, 2021.

Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615, 2011. doi: 10.1257/aer.101.6.2590.

Yong-Ju Lee, Wooyoung Lim, and Chen Zhao. Cheap talk with prior-biased inferences. 2021.

Ben Lockwood. Confirmation bias and electoral accountability. *Quarterly Journal of Political Science*, 11:471–501, 02 2017. doi: 10.1561/100.00016037.

Matthew Rabin and Joel L. Schrag. First impressions matter: A model of confirmatory bias. *The Quarterly Journal of Economics*, 114(1):37–82, 1999. ISSN 00335533, 15314650. URL http://www.jstor.org/stable/2586947.