Strategically Controlling Worldviews

Cuimin Ba[*], Danil Dmitriev[†], Victoria Hang[‡], and Freddie Papazyan[§]

February 3, 2025

Abstract

This paper studies how strategic information disclosure can consistently lead rational agents to abandon their initially correct model of the world in favor of a misspecified one. We study a dynamic game between a biased sender and an agent. Over an infinite horizon, the agent chooses between two "bandit arms" – representing alternative policies, projects, etc. – with uncertain success rates, while the sender discloses verifiable information to sway the agent towards the sender's preferred (inferior) arm. The agent initially assumes that the sender is biased but also entertains an alternative (incorrect) model where the sender is unbiased. The agent updates their beliefs and switches models when the Bayes Factor is sufficiently high. We show how the sender can successfully mislead the agent and convince them to choose the sender-preferred arm in the long run. Moreover, we characterize when the sender can achieve this outcome with certainty.

Keywords: misspecified learning, information design, model switching, dynamic games, verifiable information, bandit problems.

[*]University of Pittsburgh, Department of Economics. Email: cub8@pitt.edu.
[†]Yale University, Institution for Social and Policy Studies. Email: danil.dmitriev@yale.edu.
[‡]Texas Tech University, Department of Economics. Email: victoria.hang@ttu.edu
[§]Texas Tech University, Department of Economics. Email: freddie.papazyan@ttu.edu

1 Introduction

Shaping perceptions often hinges on the appearance of impartiality. News outlets, think tanks, car sellers, and even AI recommendation systems are more influential if they are perceived as unbiased sources of information. Yet, persuaders frequently have strategic incentives to bias their communication when they favor one outcome over another. This fact does not escape people's awareness, which is why impartiality is so valued.[1] A suspicious listener can discern the persuader's bias from the information they provide, particularly if the interaction happens over a long time span. Swaying their opinion while continuing to appear impartial in this situation seems difficult at best and contradictory at worst.

Our central question is as follows: when and how can a biased sender persistently mislead a suspicious decision-maker? We find that the answer depends on how much control the sender exerts over information flow. Appearing impartial is most successful when the sender has access to exclusive information that the decision-maker cannot acquire elsewhere. This allows the sender to avoid long-run contradictions between what the agent has directly seen and what he has been told. In fact, it may be possible to *guarantee* that the decision-maker is misled in the long run, depending on his prior beliefs. Understanding how persistent misspecification can be strategically induced has far-reaching implications ranging from political media influencing voters to think tanks skewing government policies.

Using a stylized two-armed bandit model, we explore a dynamic environment where an agent must choose between two projects with uncertain success rates.[2] A biased sender selectively discloses her private (verifiable) information with the goal of inducing the agent to pick her preferred project in the long run.[3] We assume that the agent is at least somewhat suspicious of the sender and entertains two competing subjective models. One model states that the sender is unbiased and shares everything she sees, whereas the other states (correctly) that the sender is sharing information selectively. The agent may switch between these models every period based on the Bayes factor rule and may exhibit some switching stickiness (see Ba (2024) for the framework). In essence, the agent commits to the more likely model, which then informs his action that period.

We assume that the agent cannot learn the success rates of both projects without the sender. Specifically, at the end of each period, the agent observes the outcome of one project, and the sender observes the outcome of the other. This feature serves two purposes. First, it makes the sender's information valuable to the agent so she cannot be simply ignored. Second, as becomes apparent in our results,

[1]For example, the YouGov (2024, p. 2) survey on media trust suggests that Americans tend to trust and consume news which they perceive to be "strictly factual" rather than news involving humor or commentary.

[2]Formally, these projects are Bernoulli bandit arms with unknown probabilities of success.

[3]To make the problem interesting, the sender-preferred project has a lower success probability.

this feature is crucial in allowing the misspecified subjective model to survive in the face of infinite signals. If the agent can disregard the sender's information and still perfectly learn both projects' success rates, there is nothing the sender can do. There also are many reasons why the agent may be constrained or choose not to observe the outcomes of both projects in each period, ranging from rational inattention to institutional constraints (more details in section 2.2).

We measure the sender's success by the probability of the agent's long-run action converging to sender-preferred project, which depends on the long-run model choice. We show in Proposition 1 that the conditional likelihood ratio of the two models almost surely converges to a ratio of prior beliefs of the agent. One is placed on the true success rate of the sender's project, and the other is put on the success rate implied by the sender's sharing strategy. For example, if the true success rate is 40% and the sender shares his signals so as to fake a 70% success rate, then the long-run plausibility of the models boils down to the relative likelihood of these success rates.

In Proposition 2, we show that if the sender's implied success rate is ex-ante more likely than the true success rate and the agent's switching is not too sticky, the agent almost surely adopts the sender-preferred project in the long run. Note that the sender faces an important tradeoff: while suppressing failures of sender-preferred project makes it appear better, the sender has to ensure that the implied success rate remains plausible. Successful senders exercise moderation in their selectivity. Aside from the tradeoff between appeal and plausibility, another take-away from these results is how much the sender gains from having control over her preferred project's information. Despite the agent's initial suspicion and rationality, the sender is able to *guarantee* that her persuasion succeeds. This advantage would become even more striking if the agent had no private signals and relied on the sender to learn about both projects.

To illustrate that our theoretical framework is plausible and has real-world implications, we would like to briefly discuss the case of the Dickey Amendment in U.S. politics. In the 1997, the U.S. Congress passed an omnibus federal spending bill that included a provision that the Centers for Disease Control and Prevention could not use their funds for advocating gun control. Reasons for this amendment aside, it brought about an effective freeze into all federally funded gun violence research. As a result, studying the impact of government policies on gun safety and gun violence was left in the hands of private agents (e.g., think tanks). If we think of the government as the decision-maker and the projects as different laws regulating gun ownership, our results suggest that the Dickey Amendment could have indirectly made government policy in this area vulnerable to manipulation.

At this point, the reader may wonder whether our results are completely dependent on the agent's inability to observe any information about one of the projects. To alleviate this concern, we consider an extension where the agent observes the outcome of the arm that he picked that period, and the sender observes the other.

This information structure preserves the necessity of the sender for complete learning while allowing the agent to get private signals from both arms. We show in Proposition 3 that the sender can no longer mislead the agent with certainty. The limit of the conditional likelihood ratio between the models (and the long-run action) is sensitive to the initial sample the agent collects before making the final switch between arms. However, the sender still succeeds in misleading the agent with non-trivial probability. Characterizing the probability in closed form is analytically infeasible, so we illustrate this insight through simulations in Figures 4 and 5.

Related Literature This paper explores a rapidly evolving area of research at the intersection of misspecified learning, information design, and multi-armed bandit literatures in economics, while also connecting to recent work on artificial intelligence. To our knowledge, this is the first paper that studies how a biased sender can strategically induce long run misspecification in a rational agent. We provide an explicit characterization of how this can be achieved in a robust manner. In doing so, we uncover a novel trade-off between making the state of the world implied by the misspecified model *appealing* for the sender and making it seem *plausible* for the decision-maker. The related literature is now discussed.

Misspecified learning has received a fervent resurgence[4] of attention among economists, following Esponda and Pouzo's (2016) landmark paper on Berk-Nash equilibrium, a solution concept they developed for analyzing misspecified agents in strategic settings.[5] The first wave of research that followed largely focused on settings with *dogmatic* agents who operate under an exogenous, misspecified model of the world without question.[6] More recently, a second wave of research has emerged which has focused on endogenizing misspecified learning.[7,8]

[4]Foundational work in economics related to misspecified learning includes Arrow and Green (1973), Kirman (1975), Sobel (1984), Kagel and Levin (1986), Nyarko (1991), Sargent (1999), Sandroni (2000), Eyster and Rabin (2010), Schwartzstein (2014), and Acemoglu et al. (2016).

[5]Berk (1966) is the seminal work on asymptotic misspecified (passive) learning, which was later considered in more general settings by Zhang (2006), Kleijn and van der Vaart (2006), Lian (2009), and Shalizi (2009). Berk-Nash equilibrium most closely relates to self-confirming equilibrium (Fudenberg and Tirole, 1991), cursed equilibrium (Eyster and Rabin, 2005), analogy-based expectation equilibrium (Jehiel, 2005, 2022), behavioral equilibrium (Esponda, 2008), and self-observation equilibrium (Heidhues et al., 2023).

[6]See Ortoleva and Snowberg (2015), Bohren (2016), Spiegler (2016, 2021, 2020), Fudenberg et al. (2017), Heidhues et al. (2018, 2021), Jehiel (2018), Molavi (2019), Frick et al. (2020, 2022), Spiegler (2020), Bohren and Hauser (2021), Esponda et al. (2021), Esponda and Pouzo (2021), Fudenberg et al. (2021), Murooka and Yamamoto (2021), He (2022), Ba and Gindin (2023), Bowen et al. (2023), Molavi et al. (2024), and He and Libgober (2024a).

[7]Cho and Kasa (2015, 2017), Galperti (2019), Massari (2020), Schwartzstein and Sunderam (2021), Lanzani (2022), Montiel Olea et al. (2022), Levy et al. (2022), Fudenberg and Lanzani (2023), Frick et al. (2024), He and Libgober (2024b), and Ba (2024).

[8]In addition to the theoretical work mentioned above, there has also been experimental work on

Our paper joins this second wave by precisely characterizing how senders can robustly induce agents to abandon their initially correct model of the world. Similarly to Ba (2024), we allow agents to switch between models using a (sticky) Bayes factor rule, since it is a standard, microfounded Bayesian model selection procedure. Our paper complements Ba (2024) in two ways. First, in contrast to her active learning setting, we operationalize her results in a strategic communication setting featuring a sender who has an incentive to misspecify the agent through selective sharing of information. Second, our paper focuses on an entirely different mechanism through which persistent misspecification can arise. Ba's (2024) focus was on providing a high-level characterization focused on two key features of information environments (model asymptotic accuracy and prior tightness). In contrast, our paper focuses on explicitly characterize how senders can robustly induce model misspecification through selective disclosure.

Also related is Bowen et al. (2023), who show how long-run belief polarization can emerge from (dogmatic) misperception regarding how much selective sharing they are exposed to. The results of the present paper shows that it is possible to intentionally and robustly induce such misperception. Understanding how it arises is as critical for understanding the drivers of belief polarization as the misperception's impact on long-run beliefs.

Galperti (2019) is another paper that lies at the intersection of the information design and misspecified learning.[9] He also considers the problem of persuading an agent to (reluctantly) change their model of the world, but in a qualitatively different, static setting. Eliaz et al. (2021a,b), Senkov and Kerman (2024) study the persuasion of misspecified agents. Our paper considers persuasion via selectively disclosing verifiable information, which was studied by Titova (2022). Shishkin (2021) studies the role played by credibility in persuasion. Persuasion has also been studied in dynamic settings (Sher, 2014; Honryo, 2018; Bizzotto et al., 2021; Escudé and Sinander, 2023).

Our paper considers a Bernoulli two-armed bandit setup and is thus related to vast literature on bandit experimentation, where we highlight several recent papers.[10] Sun (2024) studies dynamic, strategic censorship in an exponential bandit setting (Keller et al., 2005) without misspecification. One important feature of typical exponential bandit problems is that one arm is "safe" (yielding deterministic flow rewards) while the other is "risky" (either being "bad," never delivering a reward, or "good," delivering lump sum rewards according to a Poisson arrival process). The presence of conclusive good news ("breakthroughs") is another typ-

misspecified learning (Enke, 2020; Fan, 2024; Bolte and Fan, 2024; Acemoglu et al., 2024; Chatelain et al., 2024; Esponda et al., 2024).

[9]Kamenica and Gentzkow (2011) is the seminal work of the informaton design/persuasion literature, which is surveyed by Kamenica (2019) and Bergemann and Morris (2019).

[10]See Bergemann and Välimäki (2010, 2017) for recent surveys on bandit problems.

ical feature.[11] Bandit arms in our model are *explicitly* symmetric ex ante and have uncertain success rates, and learning about these rates does not stop. Kudinova (2023) considers a variation of Keller et al. (2005) where agents can switch risky arms from bad to good via investment. Interestingly, this can cause agents to get stuck on inferior arms indefinitely. In our model, this can happen if (and only if) the sender successfully convinces the agent that they are unbiased. These both bear resemblance to the "learning traps" studied by Liang and Mu (2020), who focus on a social learning setting with short-lived agents.

Finally, this paper relates to an emerging literature in artificial intelligence and machine learning studying how users can fool AI models (Wiyatno et al., 2019; Koren et al., 2021), how AI models can fool one another (Guo et al., 2019), and how they can fool users (Panisson et al., 2018; Smith et al., 2021; Hagendorff, 2024). Our results indicate that the prior information or sample given to AI models play an important role in whether they can be successfully manipulated.

2 Model

2.1 Preliminaries

Objective environment Time $t \in \{1, 2, \ldots\}$ is discrete and has an infinite horizon. There are two arms $i \in \{1, 2\}$, which each yield payoff $z_{it} \sim \text{Bernoulli}(\mu_i)$. Each arm's probability of success μ_i is ex ante unknown; at time $t = 0$ each μ_i is drawn independently according to probability mass function p which has support $\mathcal{M} \subset [0, 1]$.

Players There are two players: Agent and Sender. The agent is assumed to be myopic. Agent faces two bandit arms, a_1 and a_2, and chooses which arm to activate in each period $t = 1, 2, \ldots$. When activated in period t, arm i gives an i.i.d. payoff $x_{i,t} \sim \text{Bernoulli}(\mu_i^*)$:

$$\Pr(x_{i,t} = 1) = \mu_i^*, \text{ and } \Pr(x_{i,t} = 0) = 1 - \mu_i^*.$$

Prior information Agent knows the prior distribution $p(\cdot)$ and its support $\mathcal{M}$ but does not know μ_1^* and μ_2^*. Sender knows this and also observes μ_1^* and μ_2^* at $t = 0$.

Preferences Agent is myopic and chooses his action (which arm to activate) to maximize the probability of obtaining a success that period. Sender is biased towards arm a_2 and wishes to convince Agent to adopt that arm in the long run (asymptotically). In other words, she is maximizing the probability that Agent

[11] Keller and Rady (2015) consider the case with "breakdowns."

eventually picks a_2 and never switches to a_1 again. To make Sender's problem non-trivial, we will focus on cases with $\mu_1^* > \mu_2^*$, so that a_2 is objectively worse.

Agent signals In the baseline model, Agent's actions and signals are decoupled from each other. Regardless of which arm Agent activates in period t, Agent observes $x_{1,t}$, and Sender may observe and share $x_{2,t}$ in the way described below. This decoupling allows us to precisely characterize asymptotic learning outcomes and to demonstrate the stark impact of Sender's control over information. We consider a non-decoupled version of the model further in the paper (section 3.2) and obtain similar results.

Sender signals In period t, Sender observes payoff $x_{2,t}$ with probability $\gamma \in (0, 1]$. If she observes $x_{2,t}$, she may choose to reveal it to Agent or conceal it.[12] Recall that her objective is to convince Agent to adopt a_2 in the long run. To accomplish that goal, Sender commits to a (stationary) sharing strategy at $t = 0$.[13] A strategy is characterized by two probabilities: successes are shared with probability $d_{2,s} \in [0, 1]$, and failures are shared with probability $d_{2,f} \in [0, 1]$.

Subjective models Agent is not certain that Sender is biased towards a_2 and entertains two subjective models, m_b and m_{ub}. Model m_b states that Sender is *biased* — she receives information with probability γ and shares it selectively with strategy $(d_{2,s}, d_{2,f})$. Model m_{ub} states that Sender is *unbiased* — she receives information with probability $\hat{\gamma} < \gamma$ and shares all signals. As we can see, m_b is correctly specified, and m_{ub} is misspecified.[14]

Model switching Let $m(t)$ denote Agent's choice of model in period t. In line with Ba (2024), we assume that Agent employs a Bayes factor rule in deciding which model he adopts in every period. In period 1, Agent chooses $m(1)$ at random due to lack of signals. In every period $t > 1$, he calculates the conditional likelihood ratio (CLR) of the two models, also known as the Bayes factor. He switches models ($m(t) \neq m(t-1)$) if and only if the CLR of *alternative model versus current model* exceeds a switching threshold $\alpha \geq 1$. This is equivalent to switching if and only if the CLR of *current versus alternative models* falls below $\frac{1}{\alpha}$. If $\alpha > 1$, Agent's switching

[12] Signals are verifiable, so Sender cannot modify the signal before sharing.

[13] As we will see, stationary strategies are sufficient for Sender to achieve her desired outcome. Additionally, we focus on asymptotic learning in this paper and wish to consider strategies that are conducive to "stable" asymptotic outcomes.

[14] Alternative way to conceptualize this setup is to assume that Sender has two types — biased and unbiased. Agent does not have a prior over how Sender's type is drawn and relies on models m_{ub} and m_b. If Sender is unbiased, she can simply share all her information, and Agent will asymptotically learn μ_1^* and μ_2^*. We focus on the biased case.

is *sticky* (as in Ba (2024)). While we do not require switching to be sticky for our results, it will be helpful to consider the implications of this behavioral feature.

Timing At the start of each period, Agent commits to one of the models via the Bayes factor rule. Given that model and the history of signals, he evaluates his beliefs regarding μ_1 and μ_2 and chooses one arm to activate. After that, he observes $x_{1,t}$, and Sender observes $x_{2,t}$ with probability γ and chooses whether to share it with Agent. At the end of the period, Agent observes payoff $x_{1,t}$ and the counterfactual payoff $x_{2,t}$ if it was shared by Sender.

We will now discuss and justify various elements of this model. The following subsection is not required for understanding our results.

2.2 Discussion of the Model

Decoupling signals from actions. One key assumption of the baseline model is that Agent does not observe signals from a_2 and has to rely on Sender. There are a few ways in which such information structure could arise. First, Sender may have expertise in the area pertaining to analyzing a_2. For instance, citizens often rely on news media and political commentators to learn about government performance and the quality of political platforms. In this example, citizens play the role of Agent that has to learn about their electoral choices from media sources, which play the role of Sender.

Second, Agent may be prevented from directly analyzing a_2 by institutional constraints. As a notable example, the Dickey Amendment in the 1997 U.S. federal spending bill significantly inhibited the federal government's ability to fund research into gun violence. That resulted in an effective freeze in public funding of related research for two decades, leaving research about this policy area in the hands of private think tanks.[15] Government administrations here play the role of Agent, and think tanks doing research about gun violence act as Sender.

Subjective models. We assume m_b is absolutely correct about Sender's information structure and her strategy $(d_{2,s}, d_{2,f})$. Intuitively, this gives the correctly specified model a better chance against the misspecified model. We want to stack the environment against Sender in as many ways as possible in order to illustrate the key mechanisms driving the results.

On the other hand, model m_{ub} is as naïve about Sender as possible. Under that model, Agent must believe that Sender gets information with probability less than γ in order to explain the artificially low signal sharing rate induced by selective

[15] For more information about the Dickey amendment and the gun research freeze, see the American Psychological Association's summary (Jamieson, 2013) the BMJ's report on the end of the freeze (Dyer, 2019), and Brady Campaign's commentary on the freeze's impact.

sharing. Our main results would be similar if m_{ub} instead underestimated the degree of selective sharing by misunderstanding Sender's strategy $(d_{2,s}, d_{2,f})$.

Model switching. As Ba (2024) notes, the Bayes factor is "a common model selection criterion in Bayesian statistics" and has several attractive features, such as simplicity of formulation and close relation to other model selection rules. In the case of $\alpha = 1$, it can be supported endogenously by assuming that Agent has uniform prior over the two models and myopically maximizes the probability of picking the correct model in each period. We allow $\alpha > 1$ because conservative sticking to the current worldview is a well-documented psychological bias, and it has an impact on Agent's long-run learning.

Myopic agent This assumption is made primarily for expositional simplicity. Similar to the analysis in Ba (2024), we can relax it and make Agent forward-looking when exploring arms. As long as model switching remains myopic (which is reasonable, given the complexity of changing one's worldview)

Parameters γ and $\hat{\gamma}$. Model m_{ub} is naïve about Sender's strategy, and thus has to explain her less-than-always sharing of signals through other means. Misspecifying this model with respect to γ is the simplest and most natural way to accomplish that. For illustration, suppose that $\gamma = 1$ and $\hat{\gamma} = 0.7$. Suppose that Sender's strategy is such that she shares a signal in 70% of periods. Note that her strategy must then satisfy

$$\gamma \cdot \left[\mu_2^* d_{2,s} + (1 - \mu_2^*) d_{2,f}\right] = \hat{\gamma}.$$

The correct model explains Sender's occasional silence via selective sharing. The misspecified model naïvely explains this with $\hat{\gamma} = 0.7$: whenever Sender is silent, it is because she legitimately observed no signal in that period.

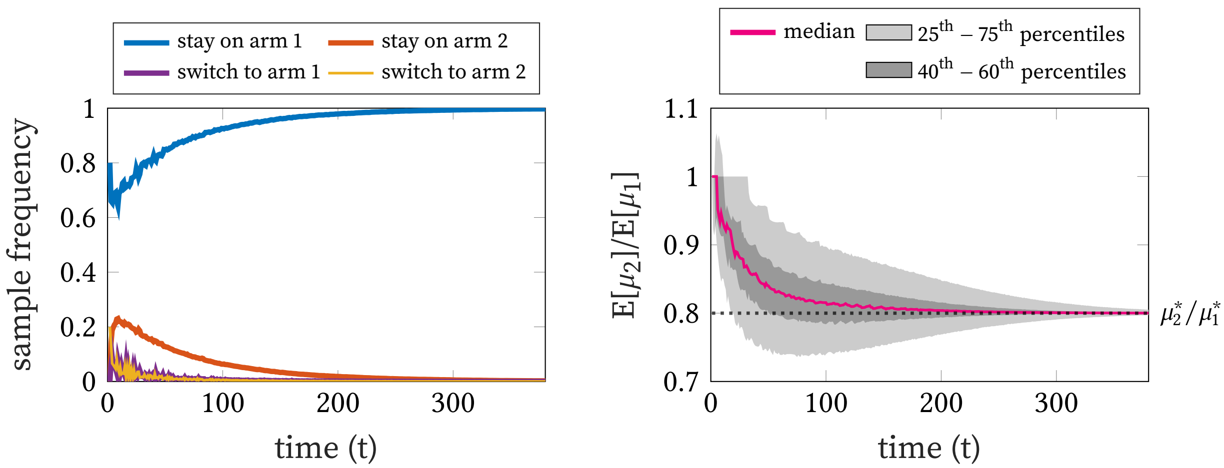
We could obtain the same structure of models in the long run by endogenizing both models' beliefs about γ. Suppose that the value of γ is drawn from some distribution on $(0, 1]$ and is known by Sender before she chooses her strategy. Since m_b is correctly specified, Agent under that model will learn the true value γ in the long run (via standard learning arguments). Under m_{ub}, Agent believes that Sender always shares her signals, which implies that her belief will eventually converge to the degenerate belief on $\hat{\gamma} = \gamma\left(\mu_2^* d_{2,s} + (1 - \mu_2^*) d_{2,f}\right)$. Whenever $d_{2,s} < 1$ or $d_{2,f} < 1$, we have $\hat{\gamma} < \gamma$. Note that in this case, Sender effectively chooses $\hat{\gamma}$ that the misspecified model will eventually converge to.

3 Analysis

We will begin by outlining the long-run outcome when there is no Sender bias. Suppose that Sender shares all her signals, success or failure. Then Agent gets an unfiltered signal about each arm each period. The two models agree on what Sender is doing, so it does not matter which one Agent adopts. His learning is then equivalent to that of a standard Bayesian agent, and by standard arguments, he eventually learns the true μ_i^* and settles on the better arm.

Remark 1. If Sender chooses $d_{2,s} = d_{2,f} = 1$, Agent's belief converges almost surely to the degenerate belief on (μ_1^*, μ_2^*) under both subjective models.

Figure 1 illustrates this observation visually.



(a) Arm transition frequencies over time.

(b) Relative expected payoffs over time.

Figure 1: Simulated case without a sender. This simulation assumes the following: each μ_i takes values in $\mathcal{M} = \{0.4, 0.5, 0.6, 0.7\}$ with probabilities $\{p(\mu)\}_{\mu \in \mathcal{M}} = \{0.1, 0.2, 0.3, 0.4\}$; $(\mu_1^*, \mu_2^*) = (.5, 0.4)$ and $(\hat{\mu}_1, \hat{\mu}_2) = (0.3, 0.6)$. Simulations were ran for 100,000 trials over 380 periods.

3.1 Environment: Agent Only Sees a_1

Let $\bar{x}_{1,\tau}^s$ $(\bar{x}_{1,\tau}^f)$ be the total number of successes (failures) on arm a_1 which were directly observed by Agent up to period τ. Let $\bar{y}_{2,\tau}^s$ $(\bar{y}_{2,\tau}^f)$ be the number of successes (failures) on arm a_2 shared by Sender up to period τ. Recall that Sender shares a success with probability $d_{2,s} \in [0, 1]$ and a failure with probability $d_{2,f} \in [0, 1]$. Note that model m_b is correctly specified about Sender's strategy.

The conditional likelihood ratio (CLR) of the two models is given by:

$$\mathrm{CLR}_{b,ub}^{\tau} = \frac{\sum\limits_{\mu_1 \in \mathcal{M}} \sum\limits_{\mu_2 \in \mathcal{M}} p(\mu_1) p(\mu_2) \mu_1^{\bar{x}_{1,\tau}^s} (1-\mu_1)^{\bar{x}_{1,\tau}^f} \left(\gamma \mu_2 d_{2,s}\right)^{\bar{y}_{2,\tau}^s} \left(\gamma (1-\mu_2) d_{2,f}\right)^{\bar{y}_{2,\tau}^f} \left(1 - \gamma \left(\mu_2 d_{2,s} + (1-\mu_2) d_{2,f}\right)\right)^{\tau - \bar{y}_{2,\tau}^s - \bar{y}_{2,\tau}^f}}{\sum\limits_{\mu_1 \in \mathcal{M}} \sum\limits_{\mu_2 \in \mathcal{M}} p(\mu_1)(\mu_2) \mu_1^{\bar{x}_{1,\tau}^s} (1-\mu_1)^{\bar{x}_{1,\tau}^f} \left(\hat{\gamma} \mu_2\right)^{\bar{y}_{2,\tau}^s} \left(\hat{\gamma}(1-\mu_2)\right)^{\bar{y}_{2,\tau}^f} \left(1-\hat{\gamma}\right)^{\tau - \bar{y}_{2,\tau}^s - \bar{y}_{2,\tau}^f}}.$$

Due to Lemma 3 in Ba (2024), we know that the CLR converges almost surely (a.s.) to some random variable with finite expectation. Since Agent's per-period information does not depend on his actions, the a.s. limit of the CLR also does not depend on them. If the limit of $\mathrm{CLR}_{b,ub}^{\tau}$ is above α, then Agent will reject model m_{ub} a.s., and if the limit is below $\frac{1}{\alpha}$, then model m_b will be rejected.

Note that model m_b is correctly specified, and thus will have perfect asymptotic accuracy. Under that model, Agent will eventually learn μ_1^* and μ_2^* a.s., in the sense that his beliefs will converge to degenerate beliefs on μ_1^* and μ_2^*. To ensure the CLR does not converge to $+\infty$, Sender must choose his strategy such that model m_{ub} also has perfect asymptotic accuracy. For that, Sender's strategy should be consistent with some $\hat{\mu}_2 \in \mathcal{M}$ under model m_{ub}, which pins down $d_{2,s}$ and $d_{2,f}$ in the following way:

$$\begin{cases} \gamma \mu_2^* d_{2,s} = \hat{\gamma} \hat{\mu}_2, \\ \gamma (1-\mu_2^*) d_{2,f} = \hat{\gamma}(1-\hat{\mu}_2), \\ 1 - \gamma \left(\mu_2^* d_{2,s} + (1-\mu_2^*) d_{2,f}\right) = 1 - \hat{\gamma}. \end{cases} \tag{1}$$

Note that the three equations are linearly dependent, so any two of them will pin down a unique pair $(d_{2,s}, d_{2,f})$:

$$\tilde{d}_s(\hat{\mu}_2) = \frac{\hat{\gamma}\hat{\mu}_2}{\gamma \mu_2^*}, \text{ and } \tilde{d}_f(\hat{\mu}_2) = \frac{\hat{\gamma}(1-\hat{\mu}_2)}{\gamma(1-\mu_2^*)}. \tag{2}$$

We will call state $\hat{\mu}_2 \in \mathcal{M}$ *feasible* if there exists a strategy $(d_{2,s}, d_{2,f})$ that induces a signal distribution *consistent* with $\hat{\mu}_2$.

Definition 1. A strategy $(d_{2,s}, d_{2,f})$ is *consistent with* $\hat{\mu}_2 \in (0,1)$ if $d_{2,s} = \tilde{d}_s(\hat{\mu}_2)$ and $d_{2,f} = \tilde{d}_f(\hat{\mu}_2)$.

Definition 2. A state $\hat{\mu}_2$ is *feasible* if $\hat{\mu}_2 \in \mathcal{M}$ and $\tilde{d}_s(\hat{\mu}_2) \in [0,1]$ and $\tilde{d}_f(\hat{\mu}_2) \in [0,1]$.

Another way to express the set of feasible states is the intersection between $\mathcal{M}$ and the following interval:

$$\left[\max\left\{0, 1 - \frac{\gamma(1-\mu_2^*)}{\hat{\gamma}}\right\}, \min\left\{\frac{\gamma\mu_2^*}{\hat{\gamma}}, 1\right\}\right].$$

The left-hand (right-hand) bound of the interval is the lowest (highest) implied arm 2 success rate ($\hat{\mu}_2$) the Sender can induce while sharing information as frequently as the agent would expect from the unbiased type (i.e. $\hat{\gamma} = \left[\mu_2^* d_{2,s} + (1-\mu_2^*)d_{2,f}\right] \cdot \gamma$).

Obviously, μ_2^* is always feasible and can be reached by setting $d_{2,s} = d_{2,f} = 1$. For any other $\hat{\mu}_2 \in \mathcal{M}$, there exists a sufficiently small $\hat{\gamma} > 0$ such that $\hat{\mu}_2$ is feasible. Thus, Sender weakly benefits from having a lower $\hat{\gamma}$ because it expands the set of feasible states. Intuitively, this is due to the fact that $\gamma - \hat{\gamma}$ can be thought as a censoring budget. The larger that difference, the more signals Sender can safely suppress without giving her selective sharing away.

We can now characterize the asymptotic limit of $\mathrm{CLR}^{\tau}_{b,ub}$.

Proposition 1. *If Sender uses sharing strategy $(d_{2,s}, d_{2,f})$ that is consistent with some feasible $\hat{\mu}_2 \in \mathcal{M}$, the CLR of the two models converges almost surely (as $\tau \to \infty$):*

$$CLR^{\tau}_{b,ub} \overset{a.s.}{\to} \frac{p(\mu_2^*)}{p(\hat{\mu}_2)}.$$

If Sender uses a sharing strategy $(d_{2,s}, d_{2,f})$ that is not consistent with any $\hat{\mu}_2 \in \mathcal{M}$, then

$$CLR^{\tau}_{b,ub} \overset{a.s.}{\to} +\infty.$$

This result has a nice intuitive interpretation. In the limit, Agent observes infinitely many unfiltered signals on arm a_1, which implies by standard arguments that his belief converges almost surely to certainty of $\mu_1 = \mu_1^*$. With regard to a_2, if the value of $\hat{\mu}_2$ that is consistent with the Sender's strategy is not in $\mathcal{M}$, then m_{ub} has imperfect accuracy and loses in the long run. If instead $\hat{\mu}_2 \in \mathcal{M}$, both models have perfect predictive accuracy, though they induce different beliefs. Under model m_b, Agent is eventually convinced that $\mu_2 = \mu_2^*$, whereas under model m_{ub}, Agent is eventually convinced that $\mu_2 = \hat{\mu}_2$. Since both models explain the data perfectly, their relative likelihood in Agent's eyes boils down to the ratio of prior probabilities placed on μ_2^* and $\hat{\mu}_2$.

Using Proposition 1, we can characterize which model-arm pairings can arise in the long run as function of $\hat{\mu}_2$ and α. Note that if Agent eventually adopts m_b, he will learn $\mu_1 = \mu_1^*$ and $\mu_2 = \mu_2^*$, which implies that he also eventually picks a_1 (recall $\mu_1^* > \mu_2^*$). On the other hand, if Agent eventually adopts m_{ub}, he will learn $\mu_1 = \mu_1^*$ and $\mu_2 = \hat{\mu}_2$, which implies that he eventually picks a_2 if and only if $\hat{\mu}_2 > \mu_1^*$.[16]

Proposition 2. *Suppose Sender's strategy is consistent with $\hat{\mu}_2 \in \mathcal{M}$. The following statements are true:*

(i) *If $\hat{\mu}_2 < \mu_1^*$, Agent eventually picks a_1 almost surely.*

(ii) *If $\hat{\mu}_2 > \mu_1^*$ and $\frac{p(\mu_2^*)}{p(\hat{\mu}_2)} > \alpha$, Agent eventually picks m_b and a_1 almost surely.*

[16] If $\hat{\mu}_2 = \mu_1^*$, Agent eventually becomes close to being indifferent between the two arms.

(iii) *If $\hat{\mu}_2 > \mu_1^*$ and $\frac{p(\mu_2^*)}{p(\hat{\mu}_2)} \in \left(\frac{1}{\alpha}, \alpha\right)$, Agent eventually picks either (m_b, a_1) or (m_{ub}, a_2) with positive probabilities.*

(iv) *If $\hat{\mu}_2 > \mu_1^*$ and $\frac{p(\mu_2^*)}{p(\hat{\mu}_2)} < \frac{1}{\alpha}$, Agent eventually picks m_{ub} and a_2 almost surely.*

The limit of the CLR pins down the model Agent eventually converges to, and that model in turn determines Agent's long-run action choice. We can see that in case (iv), Sender succeeds almost surely. For that to occur, she must find a success rate $\hat{\mu}_2$ that is both higher than μ_1^* and that seems more plausible to Agent than the true μ_2^*. The existence of such $\hat{\mu}_2$ depends in turn on Agent's prior $p(\cdot)$ and on $\hat{\gamma}$ being sufficiently low.

If no feasible $\hat{\mu}_2$ satisfies the two conditions in case (iv), the next-best thing Sender can hope for is case (iii). In that case, there exists a feasible $\hat{\mu}_2 > \mu_1^*$, but it is not sufficiently likely to induce Agent to reject m_b in the long run. Both models are asymptotically stable, and which one Agent eventually settles on depends the initial periods. With a sufficiently extreme starting sample, Agent may get thoroughly stuck on one of the models and never switch away from it. That model then determines Agent's long-run action.

It is noteworthy that Agent's learning is not improved by having a lower α. Having less sticky switching only makes it easier for Sender to sway Agent away from the correct model. However, Agent does benefit from having a larger α. If Agent's switching is very sticky and he starts with model m_b, Sender is not able to always induce $\hat{\mu}_2$ that seems sufficiently more plausible than μ_2^* to induce a switch. Both models remain robust and arise with positive probability in the limit. This argument is formalized in the following proposition.

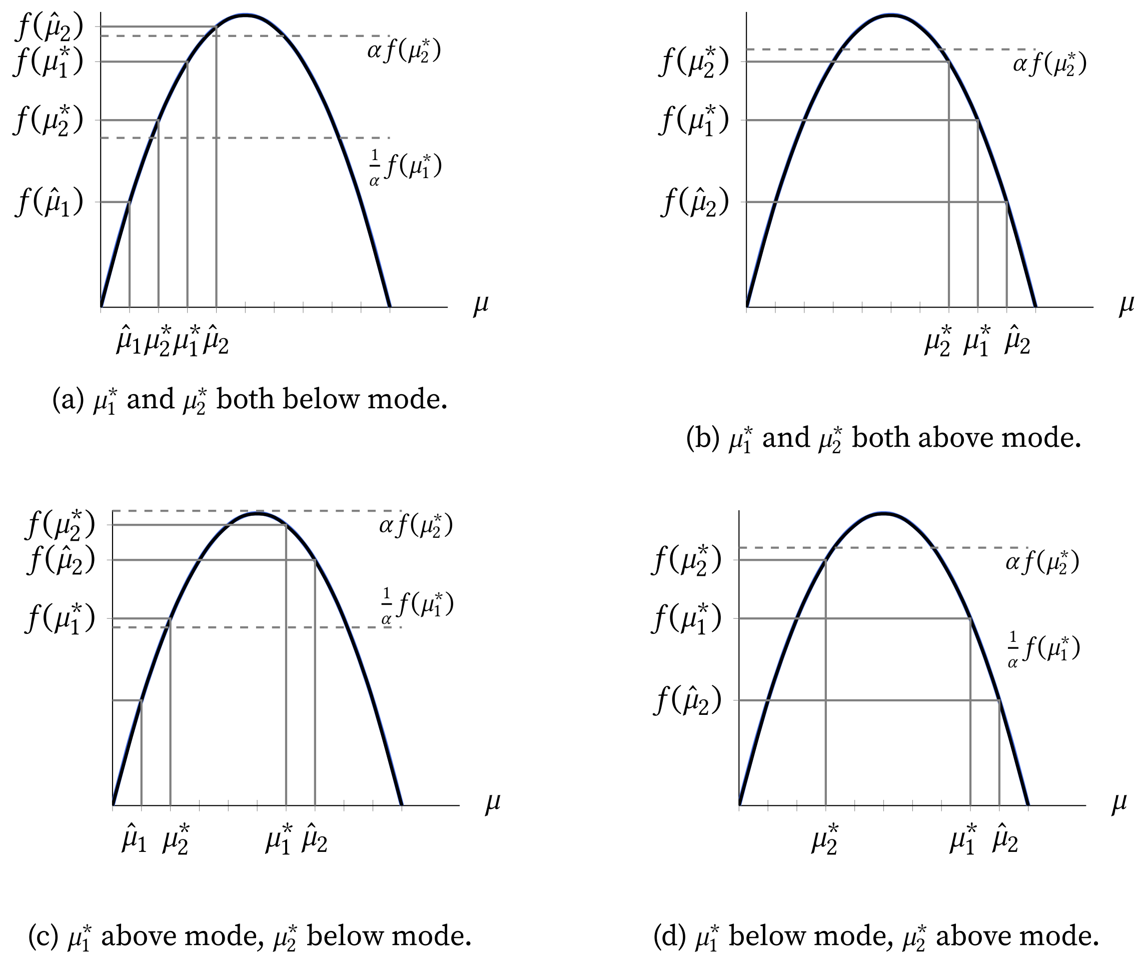
Corollary 1. *There exists $\bar{\alpha} > 1$ such that for any $\alpha \in [1, \bar{\alpha})$, there exists a strategy for Sender such that Agent eventually adopts model m_{ub} and arm a_2 almost surely. For $\alpha > \bar{\alpha}$, both m_b and m_{ub} (and corresponding arms) are eventually adopted with positive probabilities.*

It is informative to consider the implications of Proposition 2 under certain shapes of the prior $p(\cdot)$. Recall that we assume $\mu_1^* > \mu_2^*$ throughout the paper.

Example 1. Suppose $p(\cdot)$ is **single-peaked**, i.e., there exists $\bar{\mu} \in \mathcal{M}$ such that for any $\mu', \mu'' \in \mathcal{M}$ we have

$$\begin{aligned} \mu' < \mu'' \leq \bar{\mu} &\implies p(\mu') < p(\mu'') \leq p(\bar{\mu}), \\ \bar{\mu} \leq \mu' < \mu'' &\implies p(\mu'') < p(\mu') \leq p(\bar{\mu}). \end{aligned}$$

Proposition 2 implies that if $\bar{\mu} \leq \mu_2^*$, then there does not exist a Sender strategy under which Agent eventually picks a_2 almost surely. For that outcome, Sender's strategy must be consistent with some $\hat{\mu}_2 \in \mathcal{M}$ such that $\hat{\mu}_2 > \mu_1^*$ and $\frac{p(\mu_2^*)}{p(\hat{\mu}_2)} < \frac{1}{\alpha}$.



(a) μ_1^* and μ_2^* both below mode.

(b) μ_1^* and μ_2^* both above mode.

(c) μ_1^* above mode, μ_2^* below mode.

(d) μ_1^* below mode, μ_2^* above mode.

Figure 2: Four possible cases when $f(\cdot)$ is single-peaked.

However, the first inequality coupled with $\bar{\mu} \leq \mu_2^* < \mu_1^*$ implies that $p(\hat{\mu}_2) < p(\mu_1^*) < p(\mu_2^*)$. For graphical illustration, see Figures 2(b) and 2(d) below.

In contrast, if $\bar{\mu} \geq \mu_1^*$, then there may exist $\hat{\mu}_2 > \mu_1^*$ (including $\bar{\mu}$) such that $\frac{p(\mu_1^*)}{p(\hat{\mu}_2)} < \frac{1}{\alpha}$ holds. The details depend on the exact shape of $p(\cdot)$ and values μ_1^* and μ_2^*. For example, in Figure 2(c) we can see a case where μ_2^* is too likely ex-ante, or α is too large, to satisfy $f(\hat{\mu}_2) > \alpha p(\mu_2^*)$ even if Sender chooses $\hat{\mu}_2 = \bar{\mu}$. On the other hand, in Figure 2(a) we can see a case where Sender is able to find a sufficiently plausible $\hat{\mu}_2$ and to eventually convince Agent of m_{ub}.

Curiously, this example uncovers a fundamental tradeoff for Sender. On the one hand, Sender wants his project to appear as good as possible, so she wants to make the lie $\hat{\mu}_2$ high.

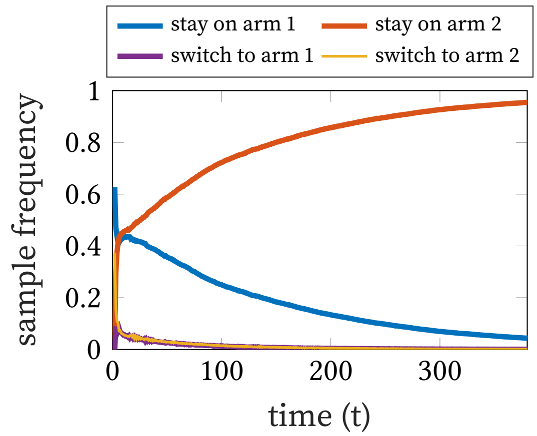
Overall, this example demonstrates that when Agent's prior is (strictly) single-peaked, Sender has a chance to almost surely bias Agent's learning only when the true expected payoff of Sender-preferred project is worse than the mode payoff.

This is caused by the following tradeoff: Sender wants to make her preferred arm appear better, but she also has to make the expected payoff appear (considerably) more plausible than the truth.

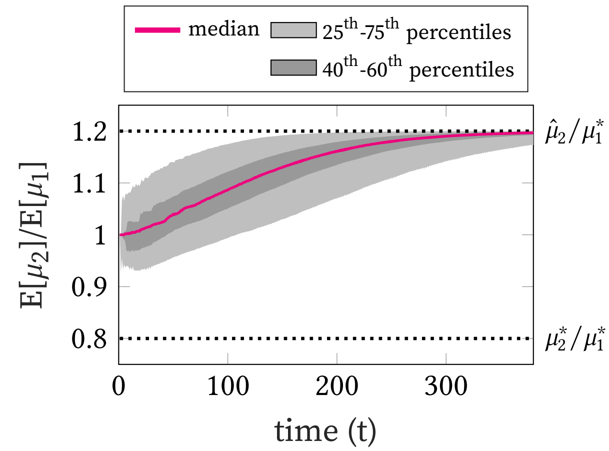
Example 2. Suppose $p(\cdot)$ is **uniform**, i.e., for any $\mu', \mu'' \in \mathcal{M}$, we have $p(\mu') = p(\mu'')$. In that case, $\frac{p(\mu_2^*)}{p(\hat{\mu}_2)} = 1$ for any Sender strategy consistent with some feasible state. As a result, Sender can never convince Agent with certainty that a_2 is the better arm. However, it may still happen with positive probability, as long as the starting sample is sufficiently indicative of model m_{ub}.

This example demonstrates that in order for Sender to mislead Agent to the fullest extent, there must be exploitable differences in Agent's prior over $\mu \in \mathcal{M}$.

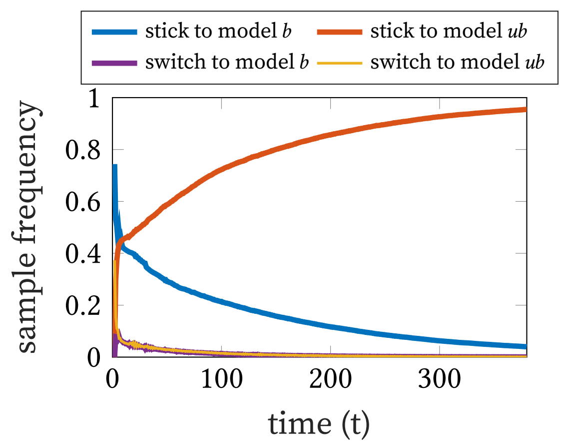
We conclude this section with a visualziation of 1 and 2. As we can see, Agent starts fairly unsure of which model is correct, but relatively quickly Sender's sharing convinces him of Sender's impartiality. Agent then quickly adopts Sender-preferred arm with rapidly climbing probability.



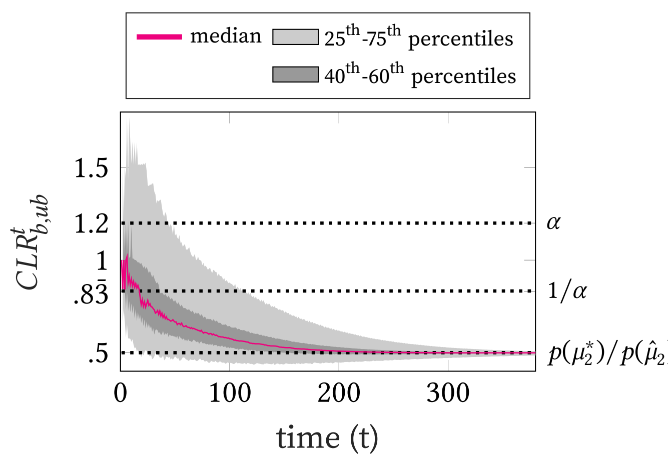
(a) Arm transition frequencies.



(b) Relative expected payoffs.



(c) Model transition frequencies.



(d) Conditional likelihood ratio.

Figure 3: Simulated case with a sender with full informational control over arm 2. This simulation assumes the following: $\alpha = 1.2$, $\gamma = 1$, $\hat{\gamma} = 0.65$; each μ_i takes values in $\mathcal{M} = \{0.3, 0.4, 0.5, 0.6\}$ with probabilities $\{p(\mu)\}_{\mu \in \mathcal{M}} = \{0.1, 0.2, 0.3, 0.4\}$; $(\mu_1^*, \mu_2^*) = (0.5, 0.4)$ and $(\hat{\mu}_1, \hat{\mu}_2) = (0.3, 0.6)$. Simulations were ran for 100,000 trials over 380 periods.

3.2 Environment: Agent sees Current Arm

The almost-sure nature of Propositions 1 demonstrates the danger of allowing Sender to control the flow of information about her preferred project. It is natural to ask what changes when Sender no longer has that exclusive control. This section covers the analysis of such environment.

Suppose that Agent observes the period-t payoff from an arm if and only if he activates that arm in period t. Sender observes the counterfactual period-t payoff of the other arm (with probability γ) and can choose to share it, as previously.

Agent's information is no longer independent of his actions, and which model-arm pair he eventually adopts may depend on the path of his actions. Let $\bar{x}_{1,\tau}^{s}$ ($\bar{x}_{i,\tau}^{f}$) be the total number of successes (failures) on arm a_i which were directly observed by Agent up to period τ. Let $\bar{y}_{i,\tau}^{s}$ ($\bar{y}_{i,\tau}^{f}$) be the number of successes (failures) on arm a_i observed and shared by Sender up to period τ. Recall that Sender shares a success with probability $d_{i,s} \in [0,1]$ and a failure with probability $d_{i,f} \in [0,1]$.[17] Note that in this case, each Sender strategy is characterized by four probabilities: $(d_{1,s}, d_{1,f}, d_{2,s}, d_{2,f})$. Finally, recall that model m_b is correctly specified with respect to Sender's strategy.

The CLR of the two models up to period τ is given by

$$\mathrm{CLR}_{b,ub}^{\tau} := \frac{\sum\limits_{\mu_1 \in \mathcal{M}} \sum\limits_{\mu_2 \in \mathcal{M}} p(\mu_1)p(\mu_2)S_{1,\tau}(\mu_1, \bar{x}_{1,\tau}, \bar{y}_{1,\tau})S_{2,\tau}(\mu_2, \bar{x}_{2,\tau}, \bar{y}_{2,\tau})}{\sum\limits_{\mu_1 \in \mathcal{M}} \sum\limits_{\mu_2 \in \mathcal{M}} p(\mu_1)p(\mu_2)\hat{S}_{1,\tau}(\mu_1, \bar{x}_{1,\tau}, \bar{y}_{1,\tau})\hat{S}_{2,\tau}(\mu_2, \bar{x}_{2,\tau}, \bar{y}_{2,\tau})},$$

where

$$S_{i,\tau}\left(\mu_i, \bar{x}_{i,\tau}, \bar{y}_{i,\tau}\right) = \mu_i^{\bar{x}_{i,\tau}^{s}}(1-\mu_i)^{\bar{x}_{i,\tau}^{f}}\left(\gamma d_{i,s}\mu_i\right)^{\bar{y}_{i,\tau}^{s}}\left(\gamma(1-\mu_i)d_{i,f}\right)^{\bar{y}_{i,\tau}^{f}}\left(1-\gamma\left(\mu_i d_{i,s}+(1-\mu_i)d_{i,f}\right)\right)^{\tau-\bar{x}_{i,\tau}^{s}-\bar{x}_{i,\tau}^{f}-\bar{y}_{i,\tau}^{s}-\bar{y}_{i,\tau}^{f}},$$

and

$$\hat{S}_{i,\tau}\left(\mu_i, \bar{x}_{i,\tau}, \bar{y}_{i,\tau}\right) = \mu_i^{\bar{x}_{i,\tau}^{s}}(1-\mu_i)^{\bar{x}_{i,\tau}^{f}}(\hat{\gamma}\mu_i)^{\bar{y}_{i,\tau}^{s}}(\hat{\gamma}(1-\mu_i))^{\bar{y}_{i,\tau}^{f}}(1-\hat{\gamma})^{\tau-\bar{x}_{i,\tau}^{s}-\bar{x}_{i,\tau}^{f}-\bar{y}_{i,\tau}^{s}-\bar{y}_{i,\tau}^{f}}.$$

Referring back to Lemma 3 in Appendix A.1 in Ba (2024), we know that the CLR converges almost surely to a random variable with finite expectation. That implies that the CLR eventually settles on some limiting value (somewhere in $[0,\infty)\cup\{+\infty\}$), which means that Agent eventually sticks to one model and does not switch again.

If Agent eventually adopts model m_b, he learns μ_1^* and μ_2^* almost surely as $t \to \infty$. That happens regardless of Sender sharing strategy, since the model is correctly specified with respect to it. On the other hand, if Agent eventually adopts m_{ub}, then what he learns depends on which arm he settles on. If he eventually settles on arm a_1, he almost surely learns μ_1^* and $\hat{\mu}_2$, where $\hat{\mu}_2$ is the state consistent with Sender's strategy $(d_{2,s}, d_{2,f})$. Similarly, if Agent settles on arm a_2, he almost surely learns $\hat{\mu}_1$ and μ_2^*. Aside from this caveat, one might expect that Proposition 1 should hold without any other changes based on similar intuition. That is not the case.

For the following result, we will assume that Sender's strategy is consistent with some $\hat{\mu}_1, \hat{\mu}_2 \in \mathcal{M}$. That is, Sender chooses a strategy that fulfills perfect asymptotic accuracy for model m_{ub} regardless of which arm Agent eventually adopts.

[17] We allow the sharing strategy to differ across arms, since Sender wishes to oversell a_2 and undersell a_1 and can observe Agent's arm choices.

Proposition 3. *Let T be the period in which Agent makes the final switch between arms. If Agent eventually sticks to arm a_1, the CLR of the two models converges almost surely to the following:*

$$CLR^{\tau}_{b,ub} \overset{a.s.}{\to} \frac{p(\mu_2^*)}{p(\hat{\mu}_2)} \times \frac{{\mu_2^*}^{\bar{x}^s_{2,T}}(1-\mu_2^*)^{\bar{x}^f_{2,T}}}{\hat{\mu}_2^{\bar{x}^s_{2,T}}(1-\hat{\mu}_2)^{\bar{x}^f_{2,T}}} \times \frac{\hat{\mu}_1^{\bar{y}^s_{1,T}}(1-\hat{\mu}_1)^{\bar{y}^f_{1,T}}}{{\mu_1^*}^{\bar{y}^s_{1,T}}(1-\mu_1^*)^{\bar{y}^f_{1,T}}}.$$

If Agent eventually sticks to arm a_2, the CLR instead converges almost surely to:

$$CLR^{\tau}_{b,ub} \overset{a.s.}{\to} \frac{p(\mu_1^*)}{p(\hat{\mu}_1)} \times \frac{{\mu_1^*}^{\bar{x}^s_{1,T}}(1-\mu_1^*)^{\bar{x}^f_{1,T}}}{\hat{\mu}_1^{\bar{x}^s_{1,T}}(1-\hat{\mu}_1)^{\bar{x}^f_{1,T}}} \times \frac{\hat{\mu}_2^{\bar{y}^s_{2,T}}(1-\hat{\mu}_2)^{\bar{y}^f_{2,T}}}{{\mu_2^*}^{\bar{y}^s_{2,T}}(1-\mu_2^*)^{\bar{y}^f_{2,T}}}.$$

Intuitive interpretation of the result becomes more challenging but can still be constructed. Suppose Agent eventually adopts arm a_2 and never switches back to a_1. In that case, he will learn μ_2^* under both models, whereas his belief about μ_1 will converge to μ_1^* and $\hat{\mu}_1$ under m_b and m_{ub}, respectively. The CLR eventually converges to

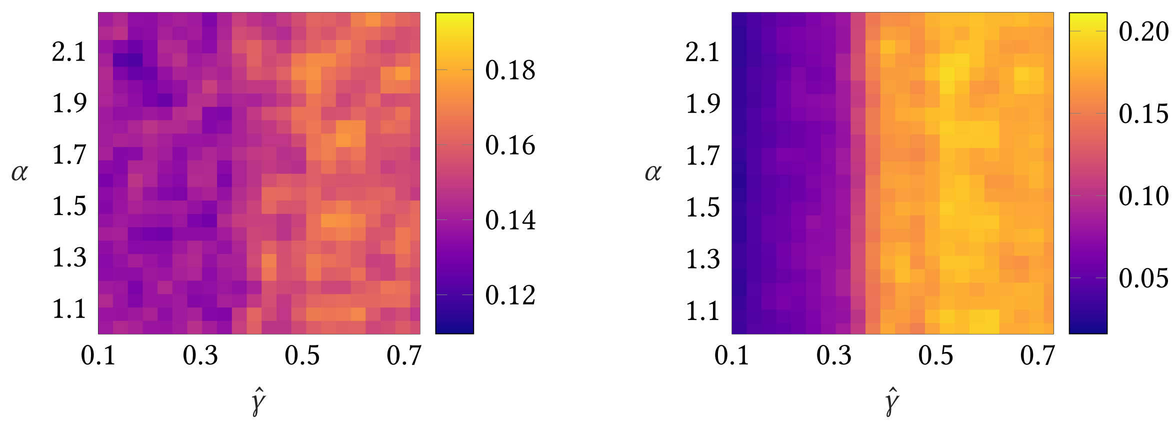
$$\frac{p(\mu_1^*)}{p(\hat{\mu}_1)} \times \frac{{\mu_1^*}^{\bar{x}^s_{1,T}}(1-\mu_1^*)^{\bar{x}^f_{1,T}}}{\hat{\mu}_1^{\bar{x}^s_{1,T}}(1-\hat{\mu}_1)^{\bar{x}^f_{1,T}}} \times \frac{\hat{\mu}_2^{\bar{y}^s_{2,T}}(1-\hat{\mu}_2)^{\bar{y}^f_{2,T}}}{{\mu_2^*}^{\bar{y}^s_{2,T}}(1-\mu_2^*)^{\bar{y}^f_{2,T}}}.$$

The first term captures the relative likelihood of μ_1^* and $\hat{\mu}_1$. The second term reflects the relative likelihood of Agent's own observations of a_1, given his beliefs under the two models. The third term is the hardest to interpret. It exists because the sample of Sender's signals about a_2 has different likelihoods under m_b and m_{ub}. In the proof, we show that their ratio is equal to the inverse relative likelihood if, hypothetically, Agent learned $\hat{\mu}_2$ under m_b and μ_2^* under m_{ub}. In summary, the second and third terms reflect the relative likelihood of the sample accumulated before the final switch, given Agent's limit beliefs.

In contrast to Proposition 1, the limit of the CLR in Proposition 3 crucially depends on the sample Agent accumulates before the final switch. Aside from the second and third terms in the limit, the initial sample also affects which arm Agent will adopt in the initial phase. That means Sender can no longer guarantee that Agent is eventually convinced of model m_{ub} or that he eventually sticks to a_2. Analytically characterizing the probability of misleading the agent is infeasible. Given that, we wish to demonstrate the effect Sender's presence has on Agent learning through a simulation, which is presented in Figures 4 and 5.

Given the chosen parameters, the CLR without the pre-final-switch sample converges to $\frac{11}{9}$ in case Agent settles on a_1 and to $\frac{4}{5}$ in case he settles on a_2. Note that for low α, neither model survives in the long run by the limit alone, and the outcome depends solely on the sample before the final switch (the simulations were done with $\alpha = 1.01$). There are two ways in which the initial sample can work against Sender. First, Agent may initially choose a_1 in several periods and accumulate a sample with CLR less than $\frac{9\alpha}{11}$. In this case, Agent may never get convinced to switch away from a_1. Second, Agent may initially choose a_2 and accumulate a sample with CLR lower than $\frac{5}{4\alpha}$. In this case, Agent is guaranteed to eventually abandon a_2 and may fall back into the first case.

However, even when Sender can only hope for a lucky initial sample, he succeeds in the long run with non-trivial probability. Simulations show that for sufficiently high $\hat{\gamma}$ the probability Agent picks m_{ub} and a_2 remains between 0.15 and 0.2 even after 400 periods. If α was higher or Agent's prior was more favorable, Sender would mislead him with a higher probability.

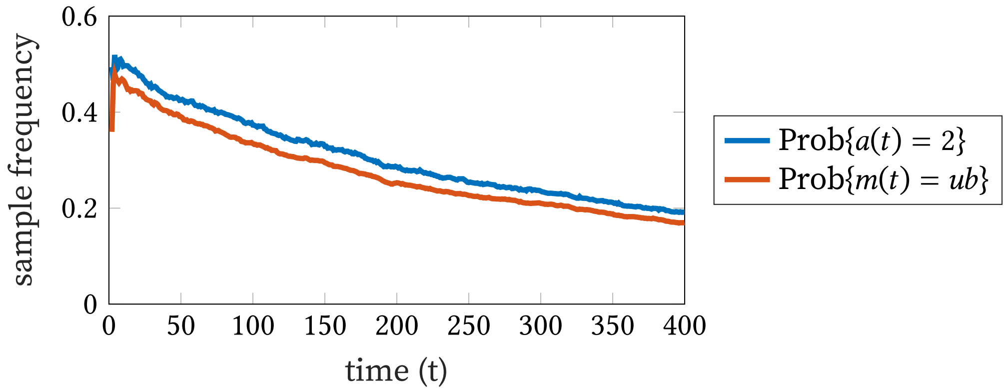


(a) Frequency of arm 2 being chosen at the end of the simulation.

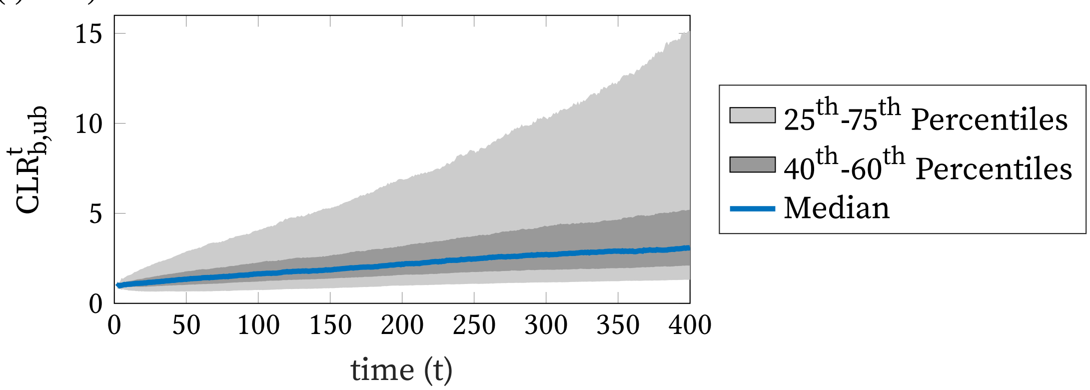
(b) Frequency of unbiased model being adopted at the end of the simulation.

Figure 4: This simulation assumes the following: $\gamma = 1$; each μ_i takes values in $\mathcal{M} = \{0.4, 0.45, 0.5, 0.55\}$ with probabilities $\{p(\mu)\}_{\mu\in\mathcal{M}} = \{0.1, 0.2, 0.3, 0.4\}$; $(\mu_1^*, \mu_2^*) = (.5, 0.45)$ and $(\hat{\mu}_1, \hat{\mu}_2) = (0.4, 0.55)$. For each $(\hat{\gamma}, \alpha)$ combination above, 2000 simulation trials (each lasting 400 periods) were ran.[18]

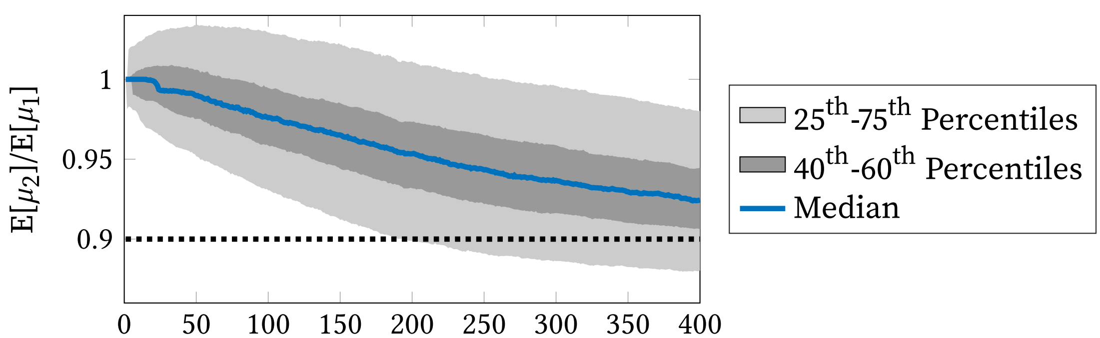
[18] Specifically $(\hat{\gamma}, \alpha) \in$ `linspace(0.1,0.8,20)` $\times$ `linspace(1.01,2.3,20)`, where `linspace(z1,z2,n)` is the MATLAB command for defining the set of $n \in \{2, 3, \ldots\}$ equally-spaced values between $z_1 \in (-\infty, \infty)$ and $z_2 \in (z_1, \infty)$ (i.e. $\{z_1 + (z_2 - z_1)(i-1)/(n-1)\}_{i=1}^{n}$).



(a) Sample frequencies of choosing arm 2 ($a(t) = 2$) and believing the unbiased model ($m(t) = ub$) over time.



(b) Median and percentile ranges for how the conditional likelihood ratio of the biased model relative to the unbiased model ($CLR^t_{b,ub}$) over time.



(c) Median and percentile ranges for how the expected payoff of arm 2 relative to arm 1 ($\mathbb{E}[\mu_2]/\mathbb{E}[\mu_1]$) evolves over time.

Figure 5: This simulation assumes the following: $\gamma = 1$ and $\hat{\gamma} = 0.8$; each μ_i takes values in $\mathcal{M} = \{0.4, 0.45, 0.5, 0.55\}$ with probabilities $\{p(\mu)\}_{\mu \in \mathcal{M}} = \{0.1, 0.2, 0.3, 0.4\}$; $(\mu_1^*, \mu_2^*) = (.5, 0.45)$ and $(\hat{\mu}_1, \hat{\mu}_2) = (0.4, 0.55)$. Simulations were ran for 10,000 trials over 400 periods.

4 Concluding Remarks

This paper demonstrates how biased senders can robustly induce rational agents to eventually abandon their initially correct worldview in favor of a false model of the world. This can be achieved with non-trivial probability, and sometimes with certainty.

Alarmingly, these results imply that a post-truth world can be robustly induced despite assumptions that favor correct learning. Our verifiable information setup shows that senders can robustly induce misspecification without lying[19] (they just need to omit key truths). Moreover, senders succeed even against sophisticated agents who learn and switch models via Bayesian reasoning and observe an infinite sequence of verifiable data.

This paper also contributes to the highly-active literatures on misspecified learning and information design. We contribute to these literatures by providing a systematic understanding of how rational agents can be systematically fooled in long run settings.

Amid the looming threat of disinformation, our results seem to underscore a serious threat: even sophisicated agents can be misled with disconcerting precision through simple omissions of the truth. We hope that future research will to investigate countermeasures against such control over worldviews. Another worthwhile line of follow-up research would investigate whether robust misdirection emerges in other information environments and settings with multiple competing senders.

References

Acemoglu, Daron, Cevat Giray Aksoy, Ceren Baysan, Carlos Molina, and Gamze Zeki, "Misperceptions and Demand for Democracy under Authoritarianism," *NBER Working Paper* № *33018*, 2024.

_, Victor Chernozhukov, and Muhamet Yildiz, "Fragility of asymptotic agreement under Bayesian learning," *Theoretical Economics*, 2016, *11* (1), 187–225.

Arrow, Kenneth J. and JR Green, "Notes on expectations equilibria in Bayesian settings," Technical Report 1973. Working Paper № 33.

Ba, Cuimin, "Robust Misspecified Models and Paradigm Shift," Working Paper, 2024. https://www.cuiminba.com/working-papers/robust_models/.

_ and Alice Gindin, "A multi-agent model of misspecified learning with overconfidence," *Games and Economic Behavior*, 2023, *142*, 315–338.

[19] See Sobel (2020) for formal definitions of lying and deception in strategic communication.

Bergemann, Dirk and Juuso Välimäki, "Bandit Problems," *Cowles Foundation Paper No. 1292*, 2010.

— **and** — , *Bandit Problems*, London: Palgrave Macmillan UK,

— **and Stephen Morris**, "Information design: A unified perspective," *Journal of Economic Literature*, 2019, *57* (1), 44–95.

Berk, Robert H., "Limiting Behavior of Posterior Distributions when the Model is Incorrect," *The Annals of Mathematical Statistics*, 1966, *37* (1), 51 – 58.

Bizzotto, Jacopo, Jesper Rüdiger, and Adrien Vigier, "Dynamic Persuasion with Outside Information," *American Economic Journal: Microeconomics*, 2021, *13* (1), pp. 179–194.

Bohren, J Aislinn, "Informational herding with model misspecification," *Journal of Economic Theory*, 2016, *163*, 222–247.

Bohren, J. Aislinn and Daniel N. Hauser, "Learning With Heterogeneous Misspecified Models: Characterization and Robustness," *Econometrica*, 2021, *89* (6), 3025–3077.

Bolte, Lukas and Tony Q. Fan, "Motivated mislearning: The case of correlation neglect," *Journal of Economic Behavior Organization*, 2024, *217*, 647–663.

Bowen, T Renee, Danil Dmitriev, and Simone Galperti, "Learning from Shared News: When Abundant Information Leads to Belief Polarization," *The Quarterly Journal of Economics*, 01 2023, *138* (2), 955–1000.

Chatelain, Marc-Antoine, Paul Han, En Hua Hu, and Xiner Xu, "Updating and Misspecification: Evidence from the Classroom," Working Paper, 2024.

Cho, In-Koo and Kenneth Kasa, "Gresham's law of model averaging," *American Economic Review*, 2017, *107* (11), 3589–3616.

Dyer, Owen, "US set to raise smoking age to 21 and to fund gun violence research after 20 year freeze," *BMJ*, 2019, *367.*

Eliaz, Kfir, Ran Spiegler, and Heidi C. Thysen, "Persuasion with endogenous misspecified beliefs," *European Economic Review*, 2021, *134*, 103712.

— , — **, and Heidi C Thysen**, "Strategic interpretations," *Journal of Economic Theory*, 2021, *192*, 105192.

Enke, Benjamin, "What you see is all there is," *The Quarterly Journal of Economics*, 2020, *135* (3), 1363–1398.

Escudé, Matteo and Ludvig Sinander, "Slow persuasion," *Theoretical Economics*, 2023, *18* (1), 129–162.

Esponda, Ignacio, "Behavioral equilibrium in economies with adverse selection," *American Economic Review*, 2008, *98* (4), 1269–1291.

— **and Demian Pouzo**, "Berk–Nash equilibrium: A framework for modeling agents with misspecified models," *Econometrica*, 2016, *84* (3), 1093–1130.

— **and** — , "Equilibrium in misspecified Markov decision processes," *Theoretical Economics*, 2021, *16* (2), 717–757.

— , — , **and Yuichi Yamamoto**, "Asymptotic behavior of Bayesian learners with misspecified models," *Journal of Economic Theory*, 2021, *195*, 105260.

— **, Emanuel Vespa, and Sevgi Yuksel**, "Mental models and learning: The case of base-rate neglect," *American Economic Review*, 2024, *114* (3), 752–782.

Eyster, E. and M. Rabin, "Naive Herding in Rich-Information Settings," *American Economic Journal: Microeconomics*, 2010, *2*, 221–243.

Eyster, Erik and Matthew Rabin, "Cursed equilibrium," *Econometrica*, 2005, *73* (5), 1623–1672.

Fan, Tony Qiaofeng, "Forming Misspecified Mental Models: The Power of Choice," Working Paper, 2024. https://www.dropbox.com/scl/fi/gb9jm3e2qd3jpzt6fuv0z/manuscript_with_appendix.pdf?rlkey=p2y2gwf4qs12ei08c967ssjev&dl=0.

Frick, Mira, Ryota Iijima, and Yuhta Ishii, "Misinterpreting Others and the Fragility of Social Learning," *Econometrica*, 2020, *88* (6), 2281–2328.

— , — , **and** — , "Dispersed Behavior and Perceptions in Assortative Societies," *American Economic Review*, September 2022, *112* (9), 3063–3105.

— , — , **and** — , "Welfare Comparisons for Biased Learning," *American Economic Review*, June 2024, *114* (6), 1612–49.

Fudenberg, Drew and Giacomo Lanzani, "Which misspecifications persist?," *Theoretical Economics*, 2023, *18* (3), 1271–1315.

— **and Jean Tirole**, "Perfect Bayesian equilibrium and sequential equilibrium," *Journal of Economic Theory*, apr 1991, *53* (2), 236–260.

— **, Giacomo Lanzani, and Philipp Strack**, "Limit Points of Endogenous Misspecified Learning," *Econometrica*, 2021, *89* (3), 1065–1098.

— , **Gleb Romanyuk, and Philipp Strack**, "Active learning with a misspecified prior," *Theoretical Economics*, 2017, *12* (3), 1155–1189.

Galperti, Simone, "Persuasion: The Art of Changing Worldviews," *American Economic Review*, March 2019, *109* (3), 996–1031.

Guo, Haoqiang, Lu Peng, Jian Zhang, Fang Qi, and Lide Duan, "Fooling AI with AI: An Accelerator for Adversarial Attacks on Deep Learning Visual Classification," in "2019 IEEE 30th International Conference on Application-specific Systems, Architectures and Processors (ASAP)," Vol. 2160-052X July 2019, pp. 136–136.

Hagendorff, Thilo, "Deception abilities emerged in large language models," *Proceedings of the National Academy of Sciences*, 2024, *121* (24), e2317967121.

He, Kevin, "Mislearning from censored data: The gambler's fallacy and other correlational mistakes in optimal-stopping problems," *Theoretical Economics*, 2022, *17* (3), 1269–1312.

— **and Jonathan Libgober**, "Higher-Order Beliefs and (Mis) learning from Prices," Technical Report 2024. https://kevinhe.net/papers/higher.pdf.

— **and** — , "Misspecified Learning and Evolutionary Stability," Technical Report 2024. https://kevinhe.net/papers/evo.pdf.

Heidhues, Paul, Botond Kőszegi, and Philipp Strack, "Unrealistic Expectations and Misguided Learning," *Econometrica*, 2018, *86* (4), 1159–1214.

— , — , **and** — , "Convergence in models of misspecified learning," *Theoretical Economics*, 2021, *16* (1), 73–99.

— , — , **and** — , "Misinterpreting Yourself," Working Paper, 2023. https://elischolar.library.yale.edu/cgi/viewcontent.cgi?article=3761&context=cowles-discussion-paper-series.

Honryo, Takakazu, "Dynamic persuasion," *Journal of Economic Theory*, 2018, *178*, 36–58.

Jamieson, Christine, "Gun violence research: History of the federal funding freeze," Technical Report, American Psychological Association 2013. Archived webpage found at https://web.archive.org/web/20130218012644/http://www.apa.org/science/about/psa/2013/02/gun-violence.aspx.

Jehiel, Philippe, "Analogy-based expectation equilibrium," *Journal of Economic theory*, 2005, *123* (2), 81–104.

— , "Investment strategy and selection bias: An equilibrium perspective on overoptimism," *American Economic Review*, 2018, *108* (6), 1582–1597.

— , “Analogy-Based Expectation Equilibrium and Related Concepts:Theory, Applications, and Beyond,” July 2022. working paper or preprint.

Kagel, John H and Dan Levin, “The winner’s curse and public information in common value auctions,” *American economic review*, 1986, *76* (5), 894–920.

Kamenica, Emir, “Bayesian persuasion and information design,” *Annual Review of Economics*, 2019, *11* (1), 249–272.

— **and Matthew Gentzkow**, “Bayesian Persuasion,” *American Economic Review*, October 2011, *101* (6), 2590–2615.

Keller, Godfrey and Sven Rady, “Breakdowns,” *Theoretical Economics*, 2015, *10* (1), 175–202.

— **, — , and Martin Cripps**, “Strategic experimentation with exponential bandits,” *Econometrica*, 2005, *73* (1), 39–68.

Kirman, Alan P., “Learning by Firms about Demand Conditions,” in Richard H. Day and Theodore Groves, eds., *Adaptive Economic Models*, Academic Press, 1975, pp. 137–156.

Kleijn, B. J. K. and A. W. van der Vaart, “Misspecification in infinite-dimensional Bayesian statistics,” *The Annals of Statistics*, 2006, *34* (2), 837 – 877.

koo Cho, In and Kenneth Kasa, “Learning and Model Validation,” *The Review of Economic Studies*, 2015, *82* (1 (290)), 45–82.

Koren, Nodens, Xingjun Ma, Qiuhong Ke, Yisen Wang, and James Bailey, “Adversarial Interaction Attacks: Fooling AI to Misinterpret Human Intentions,” in “ICML 2021 Workshop on Adversarial Machine Learning” 2021.

Kudinova, Evgeniya, “Exploration, Exploitation, Amelioration: Experimentation with Endogenously Changing Arms,” Working Paper, November 2023. `https://sites.google.com/view/evgeniya-kudinova/research`.

Lanzani, Giacomo, “Dynamic concern for misspecification,” *Available at SSRN 4454504*, 2022.

Levy, Gilat, Ronny Razin, and Alwyn Young, “Misspecified Politics and the Recurrence of Populism,” *American Economic Review*, March 2022, *112* (3), 928–62.

Lian, Heng, “On Rates of Convergence for Posterior Distributions Under Misspecification,” *Communications in Statistics - Theory and Methods*, 2009, *38* (11), 1893–1900.

Liang, Annie and Xiaosheng Mu, "Complementary information and learning traps," *The Quarterly Journal of Economics*, 2020, *135* (1), 389–448.

Massari, Filippo, "Under-reaction: Irrational behavior or robust response to model misspecification?," *Available at SSRN 3636136*, 2020.

Molavi, Pooya, "Macroeconomics with Learning and Misspecification: A General Theory and Applications," Working Paper, 2019. https://pooyamolavi.com/CREE.pdf.

_ , Alireza Tahbaz-Salehi, and Andrea Vedolin, "Model complexity, expectations, and asset prices," *Review of Economic Studies*, 2024, *91* (4), 2462–2507.

Murooka, Takeshi and Yuichi Yamamoto, "Misspecified Bayesian Learning by Strategic Players: First-Order Misspecification and Higher-Order Misspecification," Technical Report, Osaka School of International Public Policy, Osaka University 2021.

Nyarko, Yaw, "Learning in mis-specified models and the possibility of cycles," *Journal of Economic Theory*, 1991, *55* (2), 416–427.

Olea, José Luis Montiel, Pietro Ortoleva, Mallesh M Pai, and Andrea Prat, "Competing Models," *The Quarterly Journal of Economics*, 04 2022, *137* (4), 2419–2457.

Ortoleva, Pietro and Erik Snowberg, "Overconfidence in political behavior," *American Economic Review*, 2015, *105* (2), 504–535.

Panisson, Alison R., Stefan Sarkadi, Peter McBurney, Simon Parsons, and Rafael H. Bordini, "Lies, Bullshit, and Deception in Agent-Oriented Programming Languages," in "20th International Workshop on Trust in Agent societies (TRUST 2018)" 2018.

Sandroni, Alvaro, "Do markets favor agents able to make accurate predictions?," *Econometrica*, 2000, *68* (6), 1303–1341.

Sargent, Thomas J, *The conquest of American inflation*, Princeton University Press, 1999.

Schwartzstein, Joshua, "Selective attention and learning," *Journal of the European Economic Association*, 2014, *12* (6), 1423–1452.

_ and Adi Sunderam, "Using models to persuade," *American Economic Review*, 2021, *111* (1), 276–323.

Senkov, Maxim and Toygar T Kerman, "Changing Simplistic Worldviews," *arXiv preprint arXiv:2401.02867*, 2024.

Shalizi, Cosma Rohilla, "Dynamics of Bayesian updating with dependent data and misspecified models," *Electronic Journal of Statistics*, January 2009, *3* (none).

Sher, Itai, "Persuasion and dynamic communication," *Theoretical Economics*, 2014, *9* (1), 99–136.

Shishkin, Denis, "Evidence acquisition and voluntary disclosure," in "Proceedings of the 22nd ACM Conference on Economics and Computation" 2021, pp. 817–818.

Smith, Wally, Michael Kirley, Liz Sonenberg, and Frank Dignum, "The Role of Environments in Affording Deceptive Behaviour: Some Preliminary Insights from Stage Magic," in Stefan Sarkadi, Benjamin Wright, Peta Masters, and Peter McBurney, eds., *Deceptive AI*, Springer International Publishing Cham 2021, pp. 17–26.

Sobel, Joel, "Non-linear prices and price-taking behavior," *Journal of Economic Behavior & Organization*, 1984, *5* (3-4), 387–396.

— , "Lying and deception in games," *Journal of Political Economy*, 2020, *128* (3), 907–947.

Spiegler, Ran, "Bayesian networks and boundedly rational expectations," *The Quarterly Journal of Economics*, 2016, *131* (3), 1243–1290.

— , "Behavioral implications of causal misperceptions," *Annual Review of Economics*, 2020, *12*, 81–106.

— , "Can agents with causal misperceptions be systematically fooled?," *Journal of the European Economic Association*, 2021, *18* (2), 583–617.

Sun, Yiman, "A dynamic model of censorship," *Theoretical Economics*, 2024, *19* (1), 29–60.

Titova, Mariia, "Persuasion with verifiable information," Working Paper, 2022. https://maria-titova.com/papers/PVI-titova.pdf.

Wiyatno, Rey Reza, Anqi Xu, Ousmane Dia, and Archy de Berker, "Adversarial Examples in Modern Machine Learning: A Review," 2019.

YouGov, "YouGovS urvey: Media Trust, Ideology, & Usage," Technical Report, YouGov May 2024. https://d3nkl3psvxxpe9.cloudfront.net/documents/crosstabs_Media_Trust_Ideology_Usage.pdf.

Zhang, Tong, "From ε-entropy to KL-entropy: Analysis of minimum information complexity density estimation," *The Annals of Statistics*, 2006, pp. 2180–2210.

A Proofs

A.1 Proof of Proposition 1

Proof. Let $\bar{x}_\tau$ denote the number of observed successes on arm a_1 and $\bar{y}_\tau^s$ ($\bar{y}_\tau^f$) denote the number of observed and shared successes (failures) on arm a_2. Suppose that Sender shares a success (failure) on arm a_2 with probability $d_{2,s}$ ($d_{2,f}$). Recall that model m_b is correctly specified w.r.t. this sharing behavior. The CLR of the two models is then given by

$$
\begin{aligned}
\mathrm{CLR}_{b,ub}^{\tau} &= \frac{\sum_{\mu_1\in\mathcal{M}}\sum_{\mu_2\in\mathcal{M}} p(\mu_1)p(\mu_2)\mu_1^{\bar{x}_\tau}(1-\mu_1)^{\tau-\bar{x}_\tau}\left(\gamma\mu_2 d_{2,s}\right)^{\bar{y}_\tau^s}\left(\gamma(1-\mu_2)d_{2,f}\right)^{\bar{y}_\tau^f}\left(1-\gamma\left(\mu_2 d_{2,s}+(1-\mu_2)d_{2,f}\right)\right)^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}{\sum_{\mu_1\in\mathcal{M}}\sum_{\mu_2\in\mathcal{M}} p(\mu_1)(\mu_2)\mu_1^{\bar{x}_\tau}(1-\mu_1)^{\tau-\bar{x}_\tau}\left(\hat{\gamma}\mu_2\right)^{\bar{y}_\tau^s}\left(\hat{\gamma}(1-\mu_2)\right)^{\bar{y}_\tau^f}(1-\hat{\gamma})^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}} \\
&= \cancel{\frac{\sum_{\mu_1\in\mathcal{M}} p(\mu_1)\mu_1^{\bar{x}_\tau}(1-\mu_1)^{\tau-\bar{x}_\tau}}{\sum_{\mu_1\in\mathcal{M}} p(\mu_1)\mu_1^{\bar{x}_\tau}(1-\mu_1)^{\tau-\bar{x}_\tau}}} \times \frac{\sum_{\mu_2\in\mathcal{M}} p(\mu_2)\left(\gamma\mu_2 d_{2,s}\right)^{\bar{y}_\tau^s}\left(\gamma(1-\mu_2)d_{2,f}\right)^{\bar{y}_\tau^f}\left(1-\gamma\left(\mu_2 d_{2,s}+(1-\mu_2)d_{2,f}\right)\right)^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}{\sum_{\mu_2\in\mathcal{M}} p(\mu_2)\left(\hat{\gamma}\mu_2\right)^{\bar{y}_\tau^s}\left(\hat{\gamma}(1-\mu_2)\right)^{\bar{y}_\tau^f}(1-\hat{\gamma})^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}} \\
&= \frac{\sum_{\mu_2\in\mathcal{M}} p(\mu_2)\left(\gamma\mu_2 d_{2,s}\right)^{\bar{y}_\tau^s}\left(\gamma(1-\mu_2)d_{2,f}\right)^{\bar{y}_\tau^f}\left(1-\gamma\left(\mu_2 d_{2,s}+(1-\mu_2)d_{2,f}\right)\right)^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}{\sum_{\mu_2\in\mathcal{M}} p(\mu_2)\left(\hat{\gamma}\mu_2\right)^{\bar{y}_\tau^s}\left(\hat{\gamma}(1-\mu_2)\right)^{\bar{y}_\tau^f}(1-\hat{\gamma})^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}.
\end{aligned}
$$

Recall that the consistency requirement on $\hat{\mu}_2$ implies

$$
\begin{cases}
\gamma\mu_2^* d_{2,s} = \hat{\gamma}\hat{\mu}_2, \\
\gamma(1-\mu_2^*)d_{2,f} = \hat{\gamma}(1-\hat{\mu}_2), \\
1-\gamma\left(\mu_2^* d_{2,s}+(1-\mu_2^*)d_{2,f}\right) = 1-\hat{\gamma}.
\end{cases}
$$

Dividing both the numerator and denominator of $\mathrm{CLR}_{b,ub}^{\tau}$ by likelihood of the sample given μ_2^* and $\hat{\mu}_2$, respectively, we get the following:

$$\mathrm{CLR}_{b,ub}^{\tau} = \cancel{\frac{\left(\gamma\mu_2^* d_{2,s}\right)^{\bar{y}_\tau^s}\left(\gamma(1-\mu_2)d_{2,f}\right)^{\bar{y}_\tau^f}\left(1-\gamma\left(\mu_2 d_{2,s}+(1-\mu_2)d_{2,f}\right)\right)^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}{(\hat{\gamma}\hat{\mu}_2)^{\bar{y}_\tau^s}\left(\hat{\gamma}(1-\hat{\mu}_2)\right)^{\bar{y}_\tau^f}(1-\hat{\gamma})^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}} \times$$

$$\times \frac{p(\mu_2^*) + \sum_{\mu_2 \neq \mu_2^*} p(\mu_2) \frac{\left(\gamma\mu_2 d_{2,s}\right)^{\bar{y}_\tau^s}\left(\gamma(1-\mu_2)d_{2,f}\right)^{\bar{y}_\tau^f}\left(1-\gamma\left(\mu_2 d_{2,s}+(1-\mu_2)d_{2,f}\right)\right)^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}{\left(\gamma\mu_2^* d_{2,s}\right)^{\bar{y}_\tau^s}\left(\gamma(1-\mu_2)d_{2,f}\right)^{\bar{y}_\tau^f}\left(1-\gamma\left(\mu_2 d_{2,s}+(1-\mu_2)d_{2,f}\right)\right)^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}}{p(\hat{\mu}_2) + \sum_{\mu_2 \neq \hat{\mu}_2} p(\mu_2) \frac{(\hat{\gamma}\mu_2)^{\bar{y}_\tau^s}\left(\hat{\gamma}(1-\mu_2)\right)^{\bar{y}_\tau^f}(1-\hat{\gamma})^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}{(\hat{\gamma}\hat{\mu}_2)^{\bar{y}_\tau^s}\left(\hat{\gamma}(1-\hat{\mu}_2)\right)^{\bar{y}_\tau^f}(1-\hat{\gamma})^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}}$$

$$= \frac{p(\mu_2^*) + \sum_{\mu_2 \neq \mu_2^*} p(\mu_2) \frac{\left(\gamma\mu_2 d_{2,s}\right)^{\bar{y}_\tau^s}\left(\gamma(1-\mu_2)d_{2,f}\right)^{\bar{y}_\tau^f}\left(1-\gamma\left(\mu_2 d_{2,s}+(1-\mu_2)d_{2,f}\right)\right)^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}{\left(\gamma\mu_2^* d_{2,s}\right)^{\bar{y}_\tau^s}\left(\gamma(1-\mu_2)d_{2,f}\right)^{\bar{y}_\tau^f}\left(1-\gamma\left(\mu_2 d_{2,s}+(1-\mu_2)d_{2,f}\right)\right)^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}}{p(\hat{\mu}_2) + \sum_{\mu_2 \neq \hat{\mu}_2} p(\mu_2) \frac{(\hat{\gamma}\mu_2)^{\bar{y}_\tau^s}\left(\hat{\gamma}(1-\mu_2)\right)^{\bar{y}_\tau^f}(1-\hat{\gamma})^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}{(\hat{\gamma}\hat{\mu}_2)^{\bar{y}_\tau^s}\left(\hat{\gamma}(1-\hat{\mu}_2)\right)^{\bar{y}_\tau^f}(1-\hat{\gamma})^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}}.$$

For any $\mu_2 \neq \mu_2^*$, by standard arguments of Bayesian learning we have[20]

$$\frac{\left(\gamma\mu_2 d_{2,s}\right)^{\bar{y}_\tau^s}\left(\gamma(1-\mu_2)d_{2,f}\right)^{\bar{y}_\tau^f}\left(1-\gamma\left(\mu_2 d_{2,s}+(1-\mu_2)d_{2,f}\right)\right)^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}{\left(\gamma\mu_2^* d_{2,s}\right)^{\bar{y}_\tau^s}\left(\gamma(1-\mu_2^*)d_{2,f}\right)^{\bar{y}_\tau^f}\left(1-\gamma\left(\mu_2^* d_{2,s}+(1-\mu_2^*)d_{2,f}\right)\right)^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}} \xrightarrow{\text{a.s.}} 0.$$

For the denominator, let

$$\rho_\tau = \frac{\mu_2^{\bar{y}_\tau^s}(1-\mu_2)^{\bar{y}_\tau^f}}{\hat{\mu}_2^{\bar{y}_\tau^s}(1-\hat{\mu}_2)^{\bar{y}_\tau^f}}.$$

[20]For a formal argument, see the proof of Proposition 2.

Note that

$$
\begin{aligned}
\frac{1}{\tau}\ln(\rho_\tau) &= \frac{\bar{y}_\tau^s}{\tau}\ln\left(\frac{\mu_2}{\hat{\mu}_2}\right) + \frac{\bar{y}_\tau^f}{\tau}\ln\left(\frac{1-\mu_2}{1-\hat{\mu}_2}\right) \\
&\overset{\text{a.s.}}{\to} \gamma\mu_2^* d_{2,s}\ln\left(\frac{\mu_2}{\hat{\mu}_2}\right) + \gamma(1-\mu_2^*)d_{2,f}\ln\left(\frac{1-\mu_2}{1-\hat{\mu}_2}\right) \\
&= \hat{\gamma}\hat{\mu}_2\ln\left(\frac{\mu_2}{\hat{\mu}_2}\right) + \hat{\gamma}(1-\hat{\mu}_2)\ln\left(\frac{1-\mu_2}{1-\hat{\mu}_2}\right) \\
&= \hat{\gamma}\ln\left(\frac{1-\mu_2}{1-\hat{\mu}_2}\right) + \hat{\gamma}\hat{\mu}_2\ \ln\left(\frac{\mu_2}{\hat{\mu}_2}\cdot\frac{1-\hat{\mu}_2}{1-\mu_2}\right).
\end{aligned}
$$

This limit is strictly negative for all $\mu_2 \neq \hat{\mu}_2$. Therefore, $\rho_\tau \overset{\text{a.s.}}{\to} 0$ for any $\mu_2 \neq \hat{\mu}_2$. Together with the convergence of the numerator of $\text{CLR}_{b,ub}^\tau$, that implies

$$
\text{CLR}_{b,ub}^\tau \overset{\text{a.s.}}{\to} \frac{p(\mu_2^*)}{p(\hat{\mu}_2)}.
$$

■

A.2 Proof of Proposition 2

Proof. We need to establish the asymptotic limit of Agent's beliefs under both models. Let $\pi_{k,i,\tau}(\mu)$ denote Agent's period-τ posterior belief that $\mu_i = \mu$ under model m_k, for $k \in \{b, ub\}$.

Model m_b is correctly specified, and thus by standard arguments we have

$$
\pi_{b,i,\tau}(\mu_i) \overset{\text{a.s.}}{\to} 1 \text{ if } \mu_i = \mu_i^*, \text{ and } \pi_{b,i,\tau}(\mu_i) \overset{\text{a.s.}}{\to} 0 \text{ otherwise.}
$$

Consider model m_{ub}. Since the model is correctly specified with respect to a_1's signals, it follows that

$$
\pi_{ub,1,\tau}(\mu_1) \overset{\text{a.s.}}{\to} 1 \text{ if } \mu_1 = \mu_1^*, \text{ and } \pi_{ub,1,\tau}(\mu_1) \overset{\text{a.s.}}{\to} 0 \text{ otherwise.}
$$

As for a_2, by Bayes rule,

$$
\begin{aligned}
\pi_{ub,1,\tau}(\mu_2) &= \frac{(\hat{\gamma}\mu_2)^{\bar{y}_\tau^s}\,(\hat{\gamma}(1-\mu_2))^{\bar{y}_{\tau]}f}\,(1-\hat{\gamma})^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}\ p(\mu_2)}{\sum_{\mu_2'\in\mathcal{M}}\left(\hat{\gamma}\mu_2'\right)^{\bar{y}_\tau^s}\,\left(\hat{\gamma}(1-\mu_2')\right)^{\bar{y}_{\tau]}f}\,(1-\hat{\gamma})^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}\ p(\mu_2')} \\
&= \frac{\mu_2^{\bar{y}_\tau^s}(1-\mu_2)^{\bar{y}_{\tau]}f}\,p(\mu_2)}{\sum_{\mu_2'\in\mathcal{M}}\mu_2'^{\bar{y}_\tau^s}(1-\mu_2')^{\bar{y}_{\tau]}f}\,p(\mu_2')}.
\end{aligned}
$$

Recall from the proof of Proposition 1 that

$$\frac{\mu_2^{\bar{y}_\tau^s}(1-\mu_2)^{\bar{y}_\tau^f}}{\hat{\mu}_2^{\bar{y}_\tau^s}(1-\hat{\mu}_2)^{\bar{y}_\tau^f}} \xrightarrow{\text{a.s.}} 0$$

for any $\mu_2 \neq \hat{\mu}_2$. Therefore,

$$\begin{aligned}
\pi_{ub,1,\tau}(\mu_2) &= \frac{\mu_2^{\bar{y}_\tau^s}(1-\mu_2)^{\bar{y}_{\tau]}^f}\, p(\mu_2)}{\sum_{\mu_2' \in \mathcal{M}} \mu_2'^{\bar{y}_\tau^s}(1-\mu_2')^{\bar{y}_{\tau]}^f}\, p(\mu_2')} \\
&= \frac{\dfrac{\mu_2^{\bar{y}_\tau^s}(1-\mu_2)^{\bar{y}_{\tau]}^f}}{\hat{\mu}_2^{\bar{y}_\tau^s}(1-\hat{\mu}_2)^{\bar{y}_\tau^f}} \cdot p(\mu_2)}{\sum_{\mu_2' \in \mathcal{M}} \dfrac{\mu_2'^{\bar{y}_\tau^s}(1-\mu_2')^{\bar{y}_{\tau]}^f}}{\hat{\mu}_2^{\bar{y}_\tau^s}(1-\hat{\mu}_2)^{\bar{y}_\tau^f}} \cdot p(\mu_2')} \xrightarrow{\text{a.s.}} 1 \text{ if and only if } \mu_2 = \hat{\mu}_2.
\end{aligned}$$

With these limit beliefs, the expected payoff of arm a_i under model m_b is eventually equal to μ_i^*, and since $\mu_1^* > \mu_2^*$, Agent will eventually only chooses a_1. Under model m_{ub}, the expected payoff of a_1 is eventually equal to μ_1^*, and the expected payoff of a_2 is eventually equal to $\hat{\mu}_2$. Thus, Agent eventually only chooses a_1 if and only if $\mu_1^* > \hat{\mu}_2$. ■

A.3 Proof of Proposition 3

Proof. Let $a_t \in \{1,2\}$ denote the arm activated by Agent in period $t \in \{1,2,\ldots\}$. Let $s_{i,t} \in \{s,f\}$ be the signal generated by arm a_i in period t (regardless of who observes it). Let $D_T = \{x_{1,t}, x_{2,t}, y_{1,t}, y_{2,t}\}_{t=1}^T$ be the sample of signals accumulated by Agent up to period T, where we define

$$x_{i,t} = \begin{cases} 1, & \text{if } a_t = i \text{ and } s_{i,t} = s \\ 0, & \text{if } a_t = i \text{ and } s_{i,t} = f \\ \varnothing, & \text{otherwise} \end{cases} \quad \text{and } y_{i,t} = \begin{cases} 1, & \text{if } a_t \neq i,\ s_{i,t} = s, \text{ and it was shared by Sender} \\ 0, & \text{if } a_t \neq i,\ s_{i,t} = f, \text{ and it was shared by Sender} \\ \varnothing, & \text{otherwise} \end{cases}$$

With slight abuse of notation, let

$$R_T = \mathrm{CLR}_{b,ub}^T(D_T)$$

be the CLR in period T given sample D_T, and let $f_{j,T}(\mu)$ be the posterior belief placed on $\mu \in \mathcal{M}$ given sample D_T and model $j \in \{b, ub\}$.

Relabel the first period after the final switch as $t = 1$ and consider a sample of $\tau \geq 1$ periods thereafter. Let

$$\hat{R}_\tau = \mathrm{CLR}_{\mathrm{b,ub}}\left(\{x_{1,t}\}_{t=1}^\tau, \{y_{2,t}\}_{t=1}^\tau\right)$$

denote the CLR in period τ after the switch, treating $f_{j,T}(\cdot)$ as model-specific priors over $\mathcal{M}$.

The remainder of the proof will assume that Agent eventually sticks to arm a_1, with the other case being completely analogous. As previously, let $\bar{x}_\tau$ denote the number of observed successes on arm a_1 up to period τ, and let $\bar{y}^s_\tau$ ($\bar{y}^f_\tau$) denote the number of observed and shared successes (failures) on arm a_2 up to period τ. Recall that Sender shares successes (failures) on arm a_i with probability $d_{i,s}$ ($d_{i,f}$) and that model m_b is correctly specified about this strategy.

Since Agent only sticks to arm a_1 after period T, $\hat{R}_\tau$ is given by

$$\begin{aligned}
\hat{R}_\tau &= \frac{\sum\limits_{\mu_1\in\mathcal{M}}\sum\limits_{\mu_2\in\mathcal{M}} f_{b,T}(\mu_1)f_{b,T}(\mu_2)\mu_1^{\bar{x}_\tau}(1-\mu_1)^{\tau-\bar{x}_\tau}\left(\gamma\mu_2 d_{2,s}\right)^{\bar{y}^s_\tau}\left(\gamma(1-\mu_2)d_{2,f}\right)^{\bar{y}^f_\tau}\left(1-\gamma\left(\mu_2 d_{2,s}+(1-\mu_2)d_{2,f}\right)\right)^{\tau-\bar{y}^s_\tau-\bar{y}^f_\tau}}{\sum\limits_{\mu_1\in\mathcal{M}}\sum\limits_{\mu_2\in\mathcal{M}} f_{ub,T}(\mu_1)f_{ub,T}(\mu_2)\mu_1^{\bar{x}_\tau}(1-\mu_1)^{\tau-\bar{x}_\tau}\left(\hat{\gamma}\mu_2\right)^{\bar{y}^s_\tau}\left(\hat{\gamma}(1-\mu_2)\right)^{\bar{y}^f_\tau}\left(1-\hat{\gamma}\right)^{\tau-\bar{y}^s_\tau-\bar{y}^f_\tau}} \\
&= \frac{\sum\limits_{\mu_1\in\mathcal{M}} f_{b,T}(\mu_1)\mu_1^{\bar{x}_\tau}(1-\mu_1)^{\tau-\bar{x}_\tau}\times\sum\limits_{\mu_2\in\mathcal{M}} f_{b,T}(\mu_2)\left(\gamma\mu_2 d_{2,s}\right)^{\bar{y}^s_\tau}\left(\gamma(1-\mu_2)d_{2,f}\right)^{\bar{y}^f_\tau}\left(1-\gamma\left(\mu_2 d_{2,s}+(1-\mu_2)d_{2,f}\right)\right)^{\tau-\bar{y}^s_\tau-\bar{y}^f_\tau}}{\sum\limits_{\mu_1\in\mathcal{M}} f_{ub,T}(\mu_1)\mu_1^{\bar{x}_\tau}(1-\mu_1)^{\tau-\bar{x}_\tau}\times\sum\limits_{\mu_2\in\mathcal{M}} f_{ub,T}(\mu_2)\left(\hat{\gamma}\mu_2\right)^{\bar{y}^s_\tau}\left(\hat{\gamma}(1-\mu_2)\right)^{\bar{y}^f_\tau}\left(1-\hat{\gamma}\right)^{\tau-\bar{y}^s_\tau-\bar{y}^f_\tau}} \\
&= \frac{\sum\limits_{\mu_1\in\mathcal{M}} f_{b,T}(\mu_1)\mu_1^{\bar{x}_\tau}(1-\mu_1)^{\tau-\bar{x}_\tau}}{\sum\limits_{\mu_1\in\mathcal{M}} f_{ub,T}(\mu_1)\mu_1^{\bar{x}_\tau}(1-\mu_1)^{\tau-\bar{x}_\tau}}\times\frac{\sum\limits_{\mu_2\in\mathcal{M}} f_{b,T}(\mu_2)\left(\gamma\mu_2 d_{2,s}\right)^{\bar{y}^s_\tau}\left(\gamma(1-\mu_2)d_{2,f}\right)^{\bar{y}^f_\tau}\left(1-\gamma\left(\mu_2 d_{2,s}+(1-\mu_2)d_{2,f}\right)\right)^{\tau-\bar{y}^s_\tau-\bar{y}^f_\tau}}{\sum\limits_{\mu_2\in\mathcal{M}} f_{ub,T}(\mu_2)\left(\hat{\gamma}\mu_2\right)^{\bar{y}^s_\tau}\left(\hat{\gamma}(1-\mu_2)\right)^{\bar{y}^f_\tau}\left(1-\hat{\gamma}\right)^{\tau-\bar{y}^s_\tau-\bar{y}^f_\tau}}.
\end{aligned}$$

The first term in the product converges to $\frac{f_{b,T}(\mu_1^*)}{f_{ub,T}(\mu_1^*)}$ almost surely as $\tau\to\infty$. That can be shown in a similar fashion as for the second term, which we will demonstrate now. Denote the second term by $\tilde{R}_\tau$:

$$\tilde{R}_\tau = \frac{\sum\limits_{\mu_2\in\mathcal{M}} f_{b,T}(\mu_2)\left(\gamma\mu_2 d_{2,s}\right)^{\bar{y}^s_\tau}\left(\gamma(1-\mu_2)d_{2,f}\right)^{\bar{y}^f_\tau}\left(1-\gamma\left(\mu_2 d_{2,s}+(1-\mu_2)d_{2,f}\right)\right)^{\tau-\bar{y}^s_\tau-\bar{y}^f_\tau}}{\sum\limits_{\mu_2\in\mathcal{M}} f_{ub,T}(\mu_2)\left(\hat{\gamma}\mu_2\right)^{\bar{y}^s_\tau}\left(\hat{\gamma}(1-\mu_2)\right)^{\bar{y}^f_\tau}\left(1-\hat{\gamma}\right)^{\tau-\bar{y}^s_\tau-\bar{y}^f_\tau}}.$$

Dividing the numerator and denominator by the likelihood of the sample under the corresponding model, we get

$$
\begin{aligned}
\tilde{R}_\tau &= \frac{\displaystyle\sum_{\mu_2 \in \mathcal{M}} f_{b,T}(\mu_2) \frac{\left(\mu_2 d_{2,s}\right)^{\bar{y}_\tau^s} \left((1-\mu_2)d_{2,f}\right)^{\bar{y}_\tau^f} \left(1-\gamma\left(\mu_2 d_{2,s} + (1-\mu_2)d_{2,f}\right)\right)^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}{\left(\mu_2^* d_{2,s}\right)^{\bar{y}_\tau^s} \left((1-\mu_2^*)d_{2,f}\right)^{\bar{y}_\tau^f} \left(1-\gamma\left(\mu_2^* d_{2,s} + (1-\mu_2^*)d_{2,f}\right)\right)^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}}{\displaystyle\left(\frac{\hat{\gamma}}{\gamma}\right)^{\bar{y}_\tau^s+\bar{y}_\tau^f} \sum_{\mu_2 \in \mathcal{M}} f_{ub,T}(\mu_2) \frac{\mu_2^{\bar{y}_\tau^s}(1-\mu_2)^{\bar{y}_\tau^f}\,(1-\hat{\gamma})^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}{\left(\mu_2^* d_{2,s}\right)^{\bar{y}_\tau^s} \left((1-\mu_2^*)d_{2,f}\right)^{\bar{y}_\tau^f} \left(1-\gamma\left(\mu_2^* d_{2,s} + (1-\mu_2^*)d_{2,f}\right)\right)^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}} \\
&= \frac{(\hat{\gamma}\hat{\mu}_2)^{\bar{y}_\tau^s}\,(\hat{\gamma}(1-\hat{\mu}_2))^{\bar{y}_\tau^f}\,(1-\hat{\gamma})^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}{\left(\gamma\mu_2^* d_{2,s}\right)^{\bar{y}_\tau^s} \left(\gamma(1-\mu_2^*)d_{2,f}\right)^{\bar{y}_\tau^f} \left(1-\gamma\left(\mu_2^* d_{2,s} + (1-\mu_2^*)d_{2,f}\right)\right)^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}} \times \\
&\quad \times \frac{\displaystyle f_{b,T}(\mu_2^*) + \sum_{\mu_2 \neq \mu_2^*} f_{b,T}(\mu_2) \frac{\left(\mu_2 d_{2,s}\right)^{\bar{y}_\tau^s} \left((1-\mu_2)d_{2,f}\right)^{\bar{y}_\tau^f} \left(1-\gamma\left(\mu_2 d_{2,s} + (1-\mu_2)d_{2,f}\right)\right)^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}{\left(\mu_2^* d_{2,s}\right)^{\bar{y}_\tau^s} \left((1-\mu_2^*)d_{2,f}\right)^{\bar{y}_\tau^f} \left(1-\gamma\left(\mu_2^* d_{2,s} + (1-\mu_2^*)d_{2,f}\right)\right)^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}}{\displaystyle f_{ub,T}(\hat{\mu}_2) + \sum_{\mu_2 \neq \hat{\mu}_2} f_{ub,T}(\mu_2) \frac{\mu_2^{\bar{y}_\tau^s}(1-\mu_2)^{\bar{y}_\tau^f}\,\cancel{(1-\hat{\gamma})^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}}{\hat{\mu}_2^{\bar{y}_\tau^s}\,(1-\hat{\mu}_2)^{\bar{y}_\tau^f}\,\cancel{(1-\hat{\gamma})^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}}} \\
&= \frac{\displaystyle f_{b,T}(\mu_2^*) + \sum_{\mu_2 \neq \mu_2^*} f_{b,T}(\mu_2) \frac{\left(\mu_2 d_{2,s}\right)^{\bar{y}_\tau^s} \left((1-\mu_2)d_{2,f}\right)^{\bar{y}_\tau^f} \left(1-\gamma\left(\mu_2 d_{2,s} + (1-\mu_2)d_{2,f}\right)\right)^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}{\left(\mu_2^* d_{2,s}\right)^{\bar{y}_\tau^s} \left((1-\mu_2^*)d_{2,f}\right)^{\bar{y}_\tau^f} \left(1-\gamma\left(\mu_2^* d_{2,s} + (1-\mu_2^*)d_{2,f}\right)\right)^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}}{\displaystyle f_{ub,T}(\hat{\mu}_2) + \sum_{\mu_2 \neq \hat{\mu}_2} f_{ub,T}(\mu_2) \frac{\mu_2^{\bar{y}_\tau^s}(1-\mu_2)^{\bar{y}_\tau^f}}{\hat{\mu}_2^{\bar{y}_\tau^s}\,(1-\hat{\mu}_2)^{\bar{y}_\tau^f}}}.
\end{aligned}
$$

For any $\mu_2 \neq \mu_2^*$, by standard arguments of Bayesian learning we have

$$
\frac{\left(\mu_2 d_{2,s}\right)^{\bar{y}_\tau^s} \left((1-\mu_2)d_{2,f}\right)^{\bar{y}_\tau^f} \left(1-\gamma\left(\mu_2 d_{2,s} + (1-\mu_2)d_{2,f}\right)\right)^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}}{\left(\mu_2^* d_{2,s}\right)^{\bar{y}_\tau^s} \left((1-\mu_2^*)d_{2,f}\right)^{\bar{y}_\tau^f} \left(1-\gamma\left(\mu_2^* d_{2,s} + (1-\mu_2^*)d_{2,f}\right)\right)^{\tau-\bar{y}_\tau^s-\bar{y}_\tau^f}} \to 0 \text{ a.s.}
$$

For the denominator, let

$$
\rho_\tau = \frac{\mu_2^{\bar{y}_\tau^s}(1-\mu_2)^{\bar{y}_\tau^f}}{\hat{\mu}_2^{\bar{y}_\tau^s}\,(1-\hat{\mu}_2)^{\bar{y}_\tau^f}}.
$$

Note that

$$\begin{aligned}
\frac{1}{\tau}\ln(\rho_\tau) &= \frac{\bar{y}_\tau^s}{\tau}\ln\left(\frac{\mu_2}{\hat{\mu}_2}\right) + \frac{\bar{y}_\tau^f}{\tau}\ln\left(\frac{1-\mu_2}{1-\hat{\mu}_2}\right) \\
&\overset{\text{a.s.}}{\to} \gamma\mu_2^* d_{2,s}\ln\left(\frac{\mu_2}{\hat{\mu}_2}\right) + \gamma(1-\mu_2^*)d_{2,f}\ln\left(\frac{1-\mu_2}{1-\hat{\mu}_2}\right) \\
&= \hat{\gamma}\hat{\mu}_2\ln\left(\frac{\mu_2}{\hat{\mu}_2}\right) + \hat{\gamma}(1-\hat{\mu}_2)\ln\left(\frac{1-\mu_2}{1-\hat{\mu}_2}\right) \\
&= \hat{\gamma}\ln\left(\frac{1-\mu_2}{1-\hat{\mu}_2}\right) + \hat{\gamma}\hat{\mu}_2\ln\left(\frac{\mu_2}{\hat{\mu}_2}\cdot\frac{1-\hat{\mu}_2}{1-\mu_2}\right).
\end{aligned}$$

This limit is strictly negative for all $\mu_2 \neq \hat{\mu}_2$. Therefore, $\rho_\tau \overset{\text{a.s.}}{\to} 0$ for any $\mu_2 \neq \hat{\mu}_2$. That also implies

$$\tilde{R}_\tau \overset{\text{a.s.}}{\to} \frac{f_{b,T}(\mu_2^*)}{f_{ub,T}(\hat{\mu}_2)}.$$

The first term in the product $\hat{R}_\tau$ can be similarly shown to converge a.s. to

$$\frac{f_{b,T}(\mu_1^*)}{f_{ub,T}(\mu_1^*)}.$$

Therefore,

$$\hat{R}_\tau \overset{\text{a.s.}}{\to} \frac{f_{b,T}(\mu_1^*)\cdot f_{b,T}(\mu_2^*)}{f_{ub,T}(\mu_1^*)\cdot f_{ub,T}(\hat{\mu}_2)}.$$

Now consider the first T periods before the final switch. Denote the sample of observations up to that period (from both arms and from both players) by $D_T = \{x_{1,t}, x_{2,t}, y_{1,t}, y_{2,t}\}_{t=1}^T$. Note that

$$\begin{aligned}
\hat{R}_\tau &\overset{\text{a.s.}}{\to} R_T\cdot\frac{f_{b,T}(\mu_1^*)\cdot f_{b,T}(\mu_2^*)}{f_{ub,T}(\mu_1^*)\cdot f_{ub,T}(\hat{\mu}_2)} \\
&= R_T\cdot\frac{\Pr_{m_b}(\mu_1^*,\mu_2^*|D_T)}{\Pr_{m_{ub}}(\mu_1^*,\hat{\mu}_2|D_T)} \\
&= \frac{\Pr_{m_b}(D_T)}{\Pr_{m_{ub}}(D_T)}\cdot\frac{\frac{\Pr_{m_b}(D_T|\mu_1^*,\mu_2^*)p(\mu_1^*)p(\mu_2^*)}{\Pr_{m_b}(D_T)}}{\frac{\Pr_{m_{ub}}(D_T|\mu_1^*,\hat{\mu}_2)p(\mu_1^*)p(\hat{\mu}_2)}{\Pr_{m_{ub}}(D_T)}} \\
&= \frac{\Pr_{m_b}\left(D_T|\mu_1^*,\mu_2^*\right)\,\cancel{p(\mu_1^*)}\,p(\mu_2^*)}{\Pr_{m_{ub}}\left(D_T|\mu_1^*,\hat{\mu}_2\right)\,\cancel{p(\mu_1^*)}\,p(\hat{\mu}_2)}.
\end{aligned}$$

Therefore,

$$\mathrm{CLR}_{b,ub} \overset{\text{a.s.}}{\to} \frac{\Pr_{m_b}(D_T|\mu_1^*,\mu_2^*)p(\mu_2^*)}{\Pr_{m_{ub}}(D_T|\mu_1^*,\hat{\mu}_2)p(\hat{\mu}_2)}.$$

For subsequent analysis, consider the following notation. Let $\bar{x}^s_{i,T}$ ($\bar{x}^f_{i,T}$) be the total number of successes (failures) on arm a_i which were observed by Agent up to period T. Let $\bar{y}^s_{i,T}$ ($\bar{y}^f_{i,T}$) be the number of successes (failures) on arm a_i shared by Sender up to period T. Suppose that Sender shares a success (failure) on arm a_i with probability $d_{i,s}$ ($d_{i,f}$). Recall that model m_b is correctly specified w.r.t. Sender's sharing behavior.

Since the two arms generate signals independently of each other, we can write the limit of the CLR as follows:

$$\mathrm{CLR}_{b,ub} \overset{\text{a.s.}}{\to} \frac{S^b_1(\mu^*_1, \bar{x}_{1,\tau}, \bar{y}_{1,\tau})S^b_2(\mu^*_2, \bar{x}_{2,\tau}, \bar{y}_{2,\tau})p(\mu^*_2)}{S^{ub}_1(\mu^*_1, \bar{x}_{1,\tau}, \bar{y}_{1,\tau})S^{ub}_2(\hat{\mu}_2, \bar{x}_{2,\tau}, \bar{y}_{2,\tau})p(\hat{\mu}_2)},$$

where

$$S^b_i\left(\mu_i, \bar{x}_{i,T}, \bar{y}_{i,T}\right) = \mu_i^{\bar{x}^s_{i,T}}(1-\mu_i)^{\bar{x}^f_{i,T}}\left(\gamma\mu_i d_{i,s}\right)^{\bar{y}^s_{i,T}}\left(\gamma(1-\mu_i)d_{i,f}\right)^{\bar{y}^f_{i,T}}\left(1-\gamma(\mu_i(1-d_{i,s})+(1-\mu_i)(1-d_{i,f}))\right)^{T-\bar{x}^s_{i,T}-\bar{x}^f_{i,T}-\bar{y}^s_{i,T}-\bar{y}^f_{i,T}},$$
$$S^{ub}_i\left(\mu_i, \bar{x}_{i,T}, \bar{y}_{i,T}\right) = \mu_i^{\bar{x}^s_{i,T}}(1-\mu_i)^{\bar{x}^f_{i,T}}\left(\hat{\gamma}\mu_i\right)^{\bar{y}^s_{i,T}}\left(\hat{\gamma}(1-\mu_i)\right)^{\bar{y}^f_{i,T}}(1-\hat{\gamma})^{T-\bar{x}^s_{i,T}-\bar{x}^f_{i,T}-\bar{x}_{i,T}-\bar{y}_{i,T}}.$$

Note that several terms in the limit ratio cancel out. We can simplify it as follows:

$$\begin{aligned}
\mathrm{CLR}_{b,ub} &\overset{\text{a.s.}}{\to} \frac{S^b_1(\mu^*_1, \bar{x}_{1,\tau}, \bar{y}_{1,\tau})S^b_2(\mu^*_2, \bar{x}_{2,\tau}, \bar{y}_{2,\tau})}{S^{ub}_1(\mu^*_1, \bar{x}_{1,\tau}, \bar{y}_{1,\tau})S^{ub}_2(\hat{\mu}_2, \bar{x}_{2,\tau}, \bar{y}_{2,\tau})}\cdot\frac{p(\mu^*_2)}{p(\hat{\mu}_2)}\\
&= \frac{\left(\gamma\mu^*_1 d_{1,s}\right)^{\bar{y}^s_{1,T}}\left(\gamma(1-\mu^*_1)d_{1,f}\right)^{\bar{y}^f_{1,T}}\left(1-\gamma(\mu^*_1(1-d_{1,s})+(1-\mu^*_1)(1-d_{1,f}))\right)^{T-\bar{x}^s_{i,T}-\bar{x}^f_{i,T}-\bar{y}^s_{1,T}-\bar{y}^f_{1,T}}}{(\hat{\gamma}\mu^*_1)^{\bar{y}^s_{1,T}}(\hat{\gamma}(1-\mu^*_1))^{\bar{y}^f_{1,T}}(1-\hat{\gamma})^{T-\bar{x}^s_{i,T}-\bar{x}^f_{i,T}-\bar{x}_{1,T}-\bar{y}_{1,T}}}\times\frac{{\mu^*_2}^{\bar{x}^s_{2,T}}(1-\mu^*_2)^{\bar{x}^f_{2,T}}}{\hat{\mu}_2^{\bar{x}^s_{2,T}}(1-\hat{\mu}_2)^{\bar{x}^f_{2,T}}}\cdot\frac{p(\mu^*_2)}{p(\hat{\mu}_2)}\\
&= \frac{(\hat{\gamma}\hat{\mu}_1)^{\bar{y}^s_{1,T}}\left(\hat{\gamma}(1-\hat{\mu}_1)\right)^{\bar{y}^f_{1,T}}(1-\hat{\gamma})^{T-\bar{x}^s_{i,T}-\bar{x}^f_{i,T}-\bar{y}^s_{1,T}-\bar{y}^f_{1,T}}}{\left(\hat{\gamma}\mu^*_1\right)^{\bar{y}^s_{1,T}}\left(\hat{\gamma}(1-\mu^*_1)\right)^{\bar{y}^f_{1,T}}(1-\hat{\gamma})^{T-\bar{x}^s_{i,T}-\bar{x}^f_{i,T}-\bar{x}_{1,T}-\bar{y}_{1,T}}}\times\frac{{\mu^*_2}^{\bar{x}^s_{2,T}}(1-\mu^*_2)^{\bar{x}^f_{2,T}}}{\hat{\mu}_2^{\bar{x}^s_{2,T}}(1-\hat{\mu}_2)^{\bar{x}^f_{2,T}}}\cdot\frac{p(\mu^*_2)}{p(\hat{\mu}_2)},
\end{aligned}$$

where the last equality follows from the consistency requirements for $\hat{\mu}_1$:

$$\begin{cases}
\hat{\gamma}\hat{\mu}_1 = \gamma\mu^*_1 d_{1,s}\\
\hat{\gamma}(1-\hat{\mu}_1) = \gamma(1-\mu^*_1)d_{1,f}\\
1-\hat{\gamma} = 1-\gamma\left(\mu^*_1 d_{1,s}+(1-\mu^*_1)d_{1,f}\right)
\end{cases}$$

Continuing:

$$
\begin{aligned}
\mathrm{CLR}_{b,ub} &\xrightarrow{\text{a.s.}} \frac{(\hat{\gamma}\hat{\mu}_1)^{\bar{y}^s_{1,T}} \left(\hat{\gamma}(1-\hat{\mu}_1)\right)^{\bar{y}^f_{1,T}} \cancel{(1-\hat{\gamma})^{T-\bar{x}^s_{i,T}-\bar{x}^f_{i,T}-\bar{y}^s_{1,T}-\bar{y}^f_{1,T}}}}{\left(\hat{\gamma}\mu_1^*\right)^{\bar{y}^s_{1,T}} \left(\hat{\gamma}(1-\mu_1^*)\right)^{\bar{y}^f_{1,T}} \cancel{(1-\hat{\gamma})^{T-\bar{x}^s_{i,T}-\bar{x}^f_{i,T}-\bar{x}_{1,T}-\bar{y}_{1,T}}}} \times \frac{{\mu_2^*}^{\bar{x}^s_{2,T}}(1-\mu_2^*)^{\bar{x}^f_{2,T}}}{\hat{\mu}_2^{\bar{x}^s_{2,T}}(1-\hat{\mu}_2)^{\bar{x}^f_{2,T}}} \cdot \frac{p(\mu_2^*)}{p(\hat{\mu}_2)} \\
&= \frac{(\hat{\gamma}\hat{\mu}_1)^{\bar{y}^s_{1,T}} \left(\hat{\gamma}(1-\hat{\mu}_1)\right)^{\bar{y}^f_{1,T}}}{\left(\hat{\gamma}\mu_1^*\right)^{\bar{y}^s_{1,T}} \left(\hat{\gamma}(1-\mu_1^*)\right)^{\bar{y}^f_{1,T}}} \times \frac{{\mu_2^*}^{\bar{x}^s_{2,T}}(1-\mu_2^*)^{\bar{x}^f_{2,T}}}{\hat{\mu}_2^{\bar{x}^s_{2,T}}(1-\hat{\mu}_2)^{\bar{x}^f_{2,T}}} \cdot \frac{p(\mu_2^*)}{p(\hat{\mu}_2)} \\
&= \frac{\hat{\mu}_1^{\bar{y}^s_{1,T}}(1-\hat{\mu}_1)^{\bar{y}^f_{1,T}}}{{\mu_1^*}^{\bar{y}^s_{1,T}}(1-\mu_1^*)^{\bar{y}^f_{1,T}}} \times \frac{{\mu_2^*}^{\bar{x}^s_{2,T}}(1-\mu_2^*)^{\bar{x}^f_{2,T}}}{\hat{\mu}_2^{\bar{x}^s_{2,T}}(1-\hat{\mu}_2)^{\bar{x}^f_{2,T}}} \cdot \frac{p(\mu_2^*)}{p(\hat{\mu}_2)}.
\end{aligned}
$$

■