

Whose Bias?

Vasudha Jain* Mark Whitmeyer[†]

November 19, 2021

Abstract

Law enforcement acquires costly evidence with the aim of securing the conviction of a defendant, who is convicted if a decision-maker's belief exceeds a certain threshold. Either law enforcement or the decision-maker is biased and is initially overconfident that the defendant is guilty. Although an innocent defendant always prefers an unbiased decision-maker, he may prefer that law enforcement have some bias to none. Nevertheless, fixing the level of bias, an innocent defendant may prefer that the decision-maker, not law enforcement, is biased.

Keywords: Bias, Information Acquisition, Bayesian Persuasion, Misspecified Beliefs, Overconfidence

JEL Classifications: D02, D83, D90, K42

*Indian Institute of Technology Kanpur, Email: vasudhaj@iitk.ac.in

[†]Arizona State University, Email: mark.whitmeyer@gmail.com

We thank Gregorio Curello, Claudia Herresthal, Tatiana Mayskaya, Franz Ostrizek, Mallesh Pai, Joseph Whitmeyer and Thomas Wiseman for their comments. We are also grateful to the participants in the Bonn Postdoc Seminar for their feedback. This work was begun while the first author was at UT Austin and the second author was at the Institute for Microeconomics and the Hausdorff Center for Mathematics at the University of Bonn, where he was generously supported by the DFG under Germany's Excellence Strategy-GZ 2047/1, Projekt-ID 390685813.

1 Introduction

How does bias shape law enforcement's information gathering protocol, as it attempts to secure the conviction of a defendant? We explore a simple three party model. There is a defendant, suspected of a crime; law enforcement, who wants the defendant to be convicted; and a decision-maker, who decides whether to convict the defendant based on the evidence provided by law enforcement.

In our model, the defendant is either guilty or innocent, and the true prior probability of his guilt is μ . We suppose one of either law enforcement or the decision-maker is biased and has an incorrectly high prior probability that the defendant is guilty—either μ_L or μ_{DM} is strictly greater than μ . We address three questions: 1. Must this belief-based bias always hurt an innocent defendant and make it more likely that he is (wrongfully) convicted? 2. Does it matter which party, decision-maker or law enforcement, is biased? 3. Fixing the level of bias, whose bias is worse for an innocent defendant?

These are important questions. Indeed, there is an enormous literature that looks at bias in the legal system. [Fryer Jr \(2019\)](#) explores racial differences in police use of force and finds that there are significant differences in how police interact with civilians. [Devi and Fryer Jr \(2020\)](#) discover that well-publicized deadly encounters lead to a subsequent change in police investigations. These relate to our model in that they are illustrations of law enforcement's (L 's) prejudice. A large portion of the literature documents evidence of the decision-maker's (DM 's) bias: [Shayo and Zussman \(2011\)](#), [Anwar, Bayer, and Hjalmarsson \(2012\)](#), [Abrams, Bertrand, and Mullainathan \(2012\)](#), and [Cohen and Yang \(2019\)](#) all investigate racial and gender disparities in sentencing.

In our framework, conviction occurs if the decision-maker's posterior belief about the defendant's guilt exceeds an exogenously specified threshold. Law enforcement is single-minded and only wishes to maximize the probability that the defendant is convicted. In order to obtain a conviction, law enforcement may gather evidence at a cost, which

we micro-found via a continuous time Wald problem (Wald (1945)). Importantly, all evidence must be disclosed, both positive and negative, as law enforcement is prohibited from hiding exculpatory findings.¹

When law enforcement is biased, we assume that it is sophisticated and is, therefore, aware that the decision-maker's belief about the defendant's guilt is lower than her own. Crucially, it believes that the decision-maker is incorrectly pessimistic, although it is actually law enforcement that is mistaken about the prior probability of guilt.² Thus, when it collects evidence, law enforcement knows that it needs more evidence than should be required to exceed the conviction threshold (from her point of view) because of the decision-maker's lower prior. Similarly, when the decision-maker is biased, law enforcement is aware of this bias and knows that it needs less evidence to secure a conviction.

We find that an innocent defendant is always hurt by a biased decision-maker—he always prefers the decision-maker's bias to be as low as possible. On the other hand, an innocent defendant may prefer that law enforcement have a slight bias. A biased decision-maker affects law enforcement's information acquisition by lowering the conviction threshold—which always hurts an innocent defendant—whereas a biased law enforcement misperceives the defendant's guilt throughout her dynamic problem, which causes her either to over- or under-acquire information. It is this under-acquisition that benefits the innocent defendant. Nevertheless, despite the potential for beneficial bias when law enforcement is biased, an innocent defendant may still be better off when the decision-maker, not law enforcement, is biased. The over-investigation that arises when law enforcement is prejudiced may be worse than the lower conviction requirements insisted upon by the biased decision-maker.

Although the majority of the paper is focused on belief-based bias, in our penultimate section (Section 4) we explore *preference-based* bias and investigate how varying the incentives, rather than the beliefs, of the agents affects an innocent defendant's welfare. We

¹Thus, in our framework, although law enforcement may be biased, it is always *honest*: it must disclose exculpatory evidence, and it may not misrepresent evidence or falsify evidence.

²That is, when law enforcement is biased, $\mu_{DM} = \mu < \mu_L$; but law enforcement (mistakenly) believes $\mu_L = \mu > \mu_{DM}$.

find that if it is the decision-maker who is biased, it does not matter whether his bias is preference-based or belief-based: bias effects a decrease in the *DM*'s conviction threshold, which is always bad for the defendant. In contrast, if it is law enforcement who is biased, there is a significant difference between the two types of bias. Preference-based bias shapes the incentives of law enforcement, and is similar in its effects on the defendant's welfare to bias held by the decision-maker: for several interpretations of this type of bias, the innocent defendant always prefers less bias to more bias.

1.1 Related Work

In addition to the empirical papers mentioned above, there are a number of related theoretical papers. Our paper explores a static costly persuasion problem, like that explored in [Gentzkow and Kamenica \(2014\)](#), that is equivalent to a dynamic information acquisition problem via the main result of [Morris and Strack \(2017\)](#). Other related papers include [Fernandez and Mayskaya \(2017\)](#), who look at the effects of overprecision (a mistaken overconfidence in the precision of a decision-maker's information) on a decision-maker tackling a Wald problem; and [Augias and Barreto \(2020\)](#), who investigate how one should optimally persuade a "wishful thinking" receiver. As in our paper, heterogeneous priors are important in [Che and Kartik \(2009\)](#), where such differences in opinion increase the incentives for an adviser to acquire information.

Because law enforcement must disclose all of the evidence that it finds, her information acquisition is public. [Henry and Ottaviani \(2019\)](#) explore the problem of an agent dynamically acquiring (public) information in order to persuade a decision-maker to take a binary action. They study the effects of the decision-maker's ability to commit to decision rules and the agent's ability to misrepresent her findings. Similarly, [Felgenhauer and Xu \(2021\)](#) investigate how an agent can persuade a decision-maker when the information acquisition problem is private and can, moreover, be altered (falsified) at a cost. Evidence production is also public in [Brocas and Carrillo \(2007\)](#), in which an agent acquires information to sway a decision-maker with a trio of actions at his disposal.

Both [Henry \(2009\)](#) and [Herresthal \(2017\)](#) compare public and private experimentation and encounter the surprising result that private experimentation may be better

for the decision-maker. Conversely, [Janssen \(2018\)](#) also compares public versus private experimentation—where, in contrast to [Herresthal \(2017\)](#), the experimenter is informed and has state-independent preferences—and finds that private experimentation is unambiguously bad for the decision-maker. [Felgenhauer and Schulte \(2014\)](#) and [Felgenhauer and Loerke \(2017\)](#) both look at persuasion problems in which a sender privately (and sequentially) runs experiments at a cost, the results of which she may, but need not, disclose. A striking finding of the latter paper is that the receiver prefers this private experimentation to public experimentation in all Pareto-undominated equilibria. These works suggest potentially compelling extensions of our paper: how might bias shape outcomes and the innocent defendant’s welfare when exculpatory evidence need not be disclosed or when law enforcement can misrepresent or lie? Whose bias is worse for an innocent defendant then?

2 The Model

There is a defendant (D) who is being investigated for a crime. He has a private type, $\theta \in \{-1, 1\}$ (innocent and guilty, respectively). The true probability that D is guilty, $\mathbb{P}(1)$, is $\mu \in (0, 1)$. Law enforcement (L) wants D to be convicted, and she obtains a payoff of $v > 0$ if D is convicted and a payoff normalized to 0 if he is acquitted.

In order to focus on the information/evidence acquisition problem faced by L , we simplify the conviction process greatly by assuming that there is a single decision-maker (DM) who convicts D if and only if the DM ’s belief exceeds an exogenously given threshold $a \in (\mu, 1]$. Both L and the DM have subjective priors about D ’s guilt: L ’s subjective belief that D is guilty is μ_L , and the DM ’s subjective belief that D is guilty is μ_{DM} . We assume that these priors are common knowledge.³

Prior to bringing D to trial, L may acquire information at a cost, which she must disclose. We follow [Morris and Strack \(2017\)](#): L samples sequentially *à la* [Wald \(1945\)](#) in continuous time. Each instant she observes a process $(Z_t)_{t \geq 0}$, where Z_t satisfies the

³This prevents the sort of signaling through information acquisition present in [Jiang \(2021\)](#).

following stochastic differential equation:

$$dZ_t = \theta dt + \sigma dW_t ,$$

where W_t is Brownian motion. Each instant, L may decide to keep sampling or stop acquiring information. Sampling is costly; however, and L incurs a bounded and positive flow cost $c(\cdot)$ every instant that she observes Z_t . In particular, L continuously updates her belief as a result of observing Z_t , and so her posterior belief follows the process $(\mu_t)_{t \geq 0}$. The flow cost is a function of L 's posterior belief: if she observes Z_t until time t she pays a total cost of $\int_0^t c(\mu_s) ds$. L 's strategy is a stopping time τ . However, rather than solve this optimal-stopping problem directly, we make use of Theorem 1 in [Morris and Strack \(2017\)](#), where they establish an equivalence between the dynamic Wald problem and a simpler static problem.

Indeed, consider the static problem of L choosing a distribution over posteriors $F \in \Delta([0, 1])$ subject to a posterior-separable cost $C: \Delta([0, 1]) \rightarrow \mathbb{R}_+$. Theorem 1 of [Morris and Strack \(2017\)](#) states that for any flow cost c in the (dynamic) Wald problem, there exists a unique function ϕ such that $C(F) = \int_0^1 \phi(q) dF(q)$ is an equivalent static cost function, where

$$\phi(q) := \int_{\mu_L}^q \int_{\mu_L}^x \frac{\sigma^2 c(y)}{2(y(1-y))^2} dy dx . \quad (\star)$$

Henceforth, we focus on the static problem, in which L chooses a Bayes-plausible ([Kamenica and Gentzkow \(2011\)](#)) distribution over posteriors, F ,⁴ subject to a posterior-separable cost $C(F) = \int_0^1 \phi(x) dF(x)$, where ϕ is strictly convex. For tractability we restrict attention to cost functions ϕ that are twice continuously differentiable on $(0, 1)$ and that are uniformly posterior separable ([Caplin et al. \(2021\)](#)).

Definition 2.1. A cost function ϕ is **Uniformly Posterior Separable** if it is additively separable in the prior and the posterior.

This class of cost functions includes the three most common cost functions used in the literature. The entropy-based cost function that has dominated the Rational Inattention

⁴Bayes-plausibility merely requires that F has support on a subset of $[0, 1]$ and that $\mathbb{E}_F[X] = \mu_L$.

literature (see e.g. Sims (1998), Sims (2003), and Matějka and McKay (2015)),

$$\phi_{Entropy}(x) = x \log x + (1-x) \log(1-x) - \mu_L \log \mu_L - (1-\mu_L) \log(1-\mu_L) ,$$

is one such additively separable function. As Morris and Strack (2017) show, this cost function corresponds to the dynamic problem in which the flow cost is proportional to the posterior's variance. Another cost function in this class is the log-likelihood cost function that corresponds to the dynamic problem with a constant flow cost:

$$\phi_{LL}(x) = x \log \frac{x}{1-x} + (1-x) \log \frac{1-x}{x} - \mu_L \log \frac{\mu_L}{1-\mu_L} - (1-\mu_L) \log \frac{1-\mu_L}{\mu_L} .$$

Yet another is the Tsallis-entropy-based cost function:

$$\phi_T(x) = \frac{\kappa}{q-1} \left(x^q + (1-x)^q - \mu_L^q - (1-\mu_L)^q \right) ,$$

where $q > 0$ is a scaling parameter and $\kappa > 0$ is a constant.⁵ $\phi_{Entropy}$ is a special case of the Tsallis cost (as $q \rightarrow 1$) and the commonly used (posterior-)variance cost function,

$$\phi_{Var}(x) = \kappa (x - \mu_L)^2 ,$$

is the Tsallis cost when $q = 2$. This cost function corresponds to the dynamic problem in which the flow cost is proportional to the square of the posterior's variance. For some of this paper, we will further assume this particular cost function, which allows us to get a clean closed-form for L 's optimal learning, thereby simplifying the analysis.

3 Bias and its Effects

The true probability that the defendant, D , is guilty is $\mu \in (0, a)$. We are interested in the ramifications of a misspecified prior held by either law enforcement, L , or the decision-maker, the DM . We begin with the case in which L is mistaken: $\mu_L \geq \mu = \mu_{DM}$.

⁵This form of entropy was introduced in Tsallis (1988). Bloedel and Segal (2020) and Caplin et al. (2021) discuss this cost function in economic contexts.

3.1 When Law Enforcement is Biased

We suppose that the decision maker's prior belief about the defendant's likelihood of guilt is *correct*; and moreover, L is aware of their difference in opinions (or beliefs). Consequently, L understands that given a body of evidence that induces belief x from her point of view, the corresponding belief held by the DM , $x_{DM}(x)$ is

$$x_{DM}(x) = \frac{\mu(1 - \mu_L)}{\mu_L(1 - \mu) - x(\mu_L - \mu)}x .$$

Obviously, $x_{DM}(x) \leq x$. L obtains a conviction if and only if the DM 's posterior $x_{DM} \geq a$. Equivalently, she obtains a conviction if and only if L 's posterior satisfies

$$x \geq a_L := \frac{(1 - \mu)\mu_L}{(\mu_L - \mu)a + \mu(1 - \mu_L)}a .$$

It is easy to see that the effective conviction threshold encountered by L , a_L , is strictly increasing in μ_L , which naturally implies that $a_L \geq a$, strictly so if $\mu_L > \mu$. This is the portion of the problem in which the belief heterogeneity enters L 's objective directly: it scales the target belief it needs to acquire to obtain a conviction.

In her *perceived* static problem, L chooses any distribution over posteriors supported on $[0, 1]$ with mean μ_L to maximize

$$V(x) := \begin{cases} -\phi(x), & 0 \leq x < a_L \\ v - \phi(x), & a_L \leq x \leq 1 \end{cases} .$$

Because of the strict convexity of ϕ , there is a uniquely optimal distribution over posteriors F . It is either the degenerate distribution, μ_L with probability 1; or it is a binary distribution with support on a_L and some b_L (with $0 \leq b_L \leq \mu_L$).⁶ Solving for the optimal distribution may be done via the concavification technique of [Kamenica and Gentzkow \(2011\)](#) but it is perhaps easier to do it via basic calculus. Indeed, L solves

$$\max_{b_L, p_L} \{-(1 - p_L)\phi(b_L) + p_L(v - \phi(a_L))\} ,$$

⁶We assume that v is finite and that ϕ is such that certainty is prohibitively costly. The entropy and the log-likelihood cost functions both satisfy this assumption. The posterior variance cost also satisfies this assumption provided κ is sufficiently large, which we assume (see Condition 3.2, below)

subject to

$$0 \leq b_L \leq \mu_L, \quad 0 \leq p_L \leq 1, \quad \text{and} \quad p_L = \frac{\mu_L - b_L}{a_L - b_L}.$$

Substituting in for p_L , an interior solution is pinned down by the first order condition

$$0 = \phi(b_L) + (a_L - b_L) \phi'(b_L) + v - \phi(a_L) =: h(b_L, \mu_L). \quad (1)$$

Another way of thinking about L 's problem is as choosing conditional probabilities $\gamma := \mathbb{P}(g|1)$ and $\lambda := \mathbb{P}(g|-1)$. Using Bayes' law, we can write these in terms of a_L and p_L :

$$\gamma = \frac{a_L}{\mu_L} p_L, \quad \text{and} \quad \lambda = \frac{1 - a_L}{1 - \mu_L} p_L. \quad (2)$$

Combining the Equations in 2, we see that

$$\gamma = \lambda \frac{1 - \mu_L}{\mu_L} \frac{a_L}{1 - a_L} = \lambda \frac{1 - \mu}{\mu} \frac{a}{1 - a} > \lambda.$$

That is, γ is a fixed multiple of λ , and is; moreover, independent of μ_L . Since the true unconditional probability that a defendant is convicted is $p := \mu\gamma + (1 - \mu)\lambda$, we have

Remark 3.1. The true unconditional probability of conviction, p , is (strictly) increasing in L 's prior, μ_L ; if and only if the true probability of convicting an innocent D , λ , is (strictly) increasing in μ_L ; if and only if the true probability of convicting a guilty D , γ , is (strictly) increasing in μ_L .

Equivalently, the effect of the false prior on the false positive rate must always be in the same direction as the effect of the false prior on the true positive rate. Thus, a bias level that increases the occurrence of type I errors reduces the rate of type II errors and vice-versa.

In order to get more concrete results, we now specify a flow cost function that is proportional to the square of the posterior variance which yields the variance cost, $\phi_{Var} = \kappa(x - \mu_L)^2$, in L 's static problem. We also impose the following condition on the parameters, which guarantees that L never becomes certain that D is innocent.

Condition 3.2. $a_{DM} > \sqrt{v/\kappa} =: d$, where a_{DM} is the cutoff belief L needs to obtain to secure a conviction when the DM is biased.⁷

⁷We derive $a_{DM} = \frac{(1 - \mu_{DM})\mu}{(\mu - \mu_{DM})a + \mu_{DM}(1 - \mu)} a$, below, when we look at the situation with a biased decision-maker.

Furthermore, note that by construction $a_L \geq a \geq a_{DM}$. Intuitively, Condition 3.2 insists that the ratio between reward and information cost cannot be too high.

Substituting the variance cost function into Equation 1, we see that the interior solution is for L to choose a binary distribution with support $\{a_L - d, a_L\}$, which is uniquely optimal provided $\mu_L \in (a_L - d, a_L)$. Phrased in terms of the dynamic problem, in the interior solution, L acquires evidence until she reaches belief $x_h = a_L$, at which point she takes the evidence to DM to obtain a conviction; or until she reaches belief $x_l = a_L - d$, at which point she stops looking for evidence, being too pessimistic about D 's guilt to continue. If $\mu_L \leq a_L - d$, L does not acquire any evidence since she believes the odds are overly stacked against the defendant being guilty.

Substituting in for a_L and rearranging $\mu_L > a_L - d$, we see that L acquires evidence provided

$$d > \frac{(a - \mu)(1 - \mu_L)\mu_L}{(\mu_L - \mu)a + (1 - \mu_L)\mu} =: \underline{d}.$$

Our next result is somewhat counterintuitive: we show that there are parameters for which a biased L does not acquire evidence (which is to D 's benefit since he is therefore convicted with probability 0) whereas an unbiased L does.

Lemma 3.3. *Assume the variance cost, $\phi = \phi_{Var}$. If the prior probability of guilt and the conviction are sufficiently low, $1 > \mu + a$, the minimal reward ratio required for investigation, \underline{d} , is strictly increasing in L 's prior, μ_L , for all μ_L sufficiently close to μ . Otherwise (if $1 < \mu + a$), \underline{d} is strictly decreasing in μ_L .*

Viz., a biased L may actually investigate (and therefore convict) defendants less frequently than a non-biased L . Active information acquisition requires that $d > a_L - \mu_L$ and so \underline{d} is strictly increasing in μ_L if and only if $a_L - \mu_L$ is strictly increasing in μ_L . Moreover, $p_L = (\mu_L + d - a_L)/d$, and so it is easy to see that \underline{d} is strictly increasing in μ_L if and only if p_L is strictly decreasing in μ_L . This is intuitive: crossing the threshold from active information acquisition to no information acquisition requires that L shrink from positive to negative, the latter of which is obviously impossible, requiring a corner solution (no information acquisition).

The key term in the expression for p_L is $\mu_L - a_L$. a_L is strictly increasing in μ_L and so

this second term allows μ_L to exert a negative influence on p_L . The first term, obviously, is where μ_L has a positive influence. It is when the second term (and μ and μ_L) is (are) small that a_L reacts strongly to a change in μ_L , yielding the result. That μ_L cannot be too extreme is a common theme throughout our results: a severe bias is bad for D , regardless of who is biased.

Turning our attention to the case when there is information acquisition, we may use the equations in 2 to derive explicit expressions for γ and λ , which are

$$\gamma = \left(\frac{a_L}{\mu_L} \right) \frac{\mu_L + d - a_L}{d}, \quad \text{and} \quad \lambda = \left(\frac{1 - a_L}{1 - \mu_L} \right) \frac{\mu_L + d - a_L}{d}.$$

In the same spirit as Lemma 3.3, the next theorem details when D benefits from L 's bias.

Theorem 3.4. *Assume the variance cost, $\phi = \phi_{Var}$.*

- i Let the correct prior, reward ratio, and conviction threshold (μ , d , and a , respectively) be such that if she were not biased L would acquire information ($\mu > a - d$). If a is sufficiently low and d is sufficiently high ($2a - d < 1$), then for all μ_L sufficiently close to μ , an innocent D (strictly) prefers for L to be biased rather than unbiased. Otherwise, an innocent D prefers for L to be unbiased rather than biased, for any amount of bias.*
- ii If μ , d , and a are such that if she were not biased L would not acquire information ($\mu \leq a - d$), an innocent D (weakly) prefers for L to have no bias.*

In fact, an even stronger result holds: if a is sufficiently low and d is sufficiently high (and $\mu > a - d$), then for all μ_L sufficiently close to μ , λ , the probability that an innocent D is convicted is (strictly) decreasing in μ_L . Not only is bias better for D than no bias, but D 's welfare is actually increasing in the amount of bias. Outside of this interval, D 's welfare is decreasing in the amount of bias.

The last statement of Theorem 3.4 is trivial: if the parameters are such that L does not investigate D , D can only be (weakly) hurt by any bias. In addition, the statement of this theorem is phrased in terms of the outcome for the innocent D , but recall that bias has the same affect on outcome for the guilty D . Accordingly, D prefers L have a mild bias provided the bar for conviction is not too high and L 's reward for a conviction is not too low.

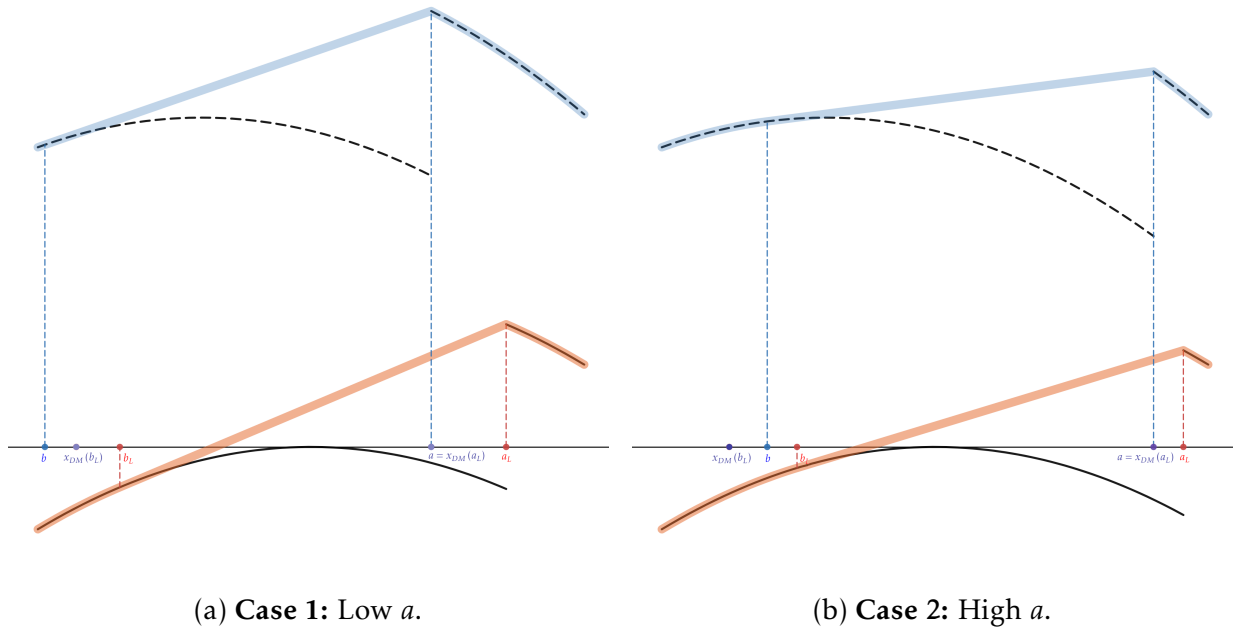


Figure 1: L 's perceived value function, $V(x)$, (solid black); its concavification (red); L 's actual value function (dashed black); and its concavification (blue).

Figure 1 illustrates two specifications of the parameters. The first, Case 1, is an instance in which bias benefits D . The bottom function is L 's perceived value function, given her incorrect prior. She acquires evidence until her posterior is a_L or until her posterior is b_L . Importantly, $x_{DM}(b_L)$ is the correct belief that L should have after acquiring the evidence that yields her belief b_L . It is always less than b_L , but (in general) it may or may not be less than b , the low posterior in the counterfactual problem in which L is not biased. It is precisely when $x_{DM}(b_L) > b$ that bias strictly benefits D : L mistakenly stops her information acquisition too early, earlier than she would have, had she the correct prior. Accordingly, Case 2, is an instance in which bias strictly hurts D . There, $x_{DM}(b_L) < b$: L over-investigates, making it more likely that she encounters the evidence she needs to convict D , to D 's detriment.

3.2 When the Decision-Maker is Biased

We now stipulate that it is the decision-maker who is biased against the defendant and has a prior belief $\mu_{DM} > \mu = \mu_L$ that D is guilty. If L induces x , the corresponding belief

held by the DM is

$$x_{DM}(x) = \frac{\mu_{DM}(1-\mu)}{\mu(1-\mu_{DM}) - x(\mu - \mu_{DM})} x.$$

Thus, L secures a conviction if and only if it induces a belief

$$x \geq a_{DM} := \frac{(1-\mu_{DM})\mu}{(\mu - \mu_{DM})a + \mu_{DM}(1-\mu)} a.$$

Using Bayes' law, we can write out the conditional probabilities $\hat{\gamma} := \mathbb{P}(g|1)$ and $\hat{\lambda} := \mathbb{P}(g|-1)$:

$$\hat{\gamma} = \frac{a_{DM}}{\mu} p_{DM}, \quad \text{and} \quad \hat{\lambda} = \frac{1-a_{DM}}{1-\mu} p_{DM},$$

where p_{DM} is pinned down by the first order condition

$$0 = \phi(b_{DM}) + (a_{DM} - b_{DM})\phi'(b_{DM}) + v - \phi(a_{DM}) =: f(b_{DM}, a_{DM}), \quad (2)$$

and $p_{DM} = (\mu - b_{DM}) / (a_{DM} - b_{DM})$.⁸

The conviction threshold a_{DM} faced by L is strictly decreasing in the DM 's prior, μ_{DM} . If this threshold, a_{DM} , is weakly less than μ , (equivalently, if $a \leq \mu_{DM}$) law enforcement does not obtain any information and secures a conviction with probability 1. Moreover, as we establish in proof of Proposition 3.5, the lower bound for L 's learning, b_{DM} , is also strictly decreasing in μ_{DM} . Consequently, any increase in the DM 's bias that either renders information acquisition optimal or sub-optimal can only hurt the innocent D . As it turns out, we are able to prove a general result concerning the effect of prejudice on an innocent D 's welfare.

Proposition 3.5. *The probability that an innocent D is convicted, $\hat{\lambda}$, is increasing in the DM 's bias, strictly so if L would acquire information in the problem without bias.*

This result is intuitive, once we make the observation that an increase in the DM 's bias is identical from L 's (and D 's) perspective to a decrease in the threshold needed to secure conviction. A lower conviction threshold obviously hurts D overall, and this proposition establishes that it hurts an innocent D to boot.

⁸Keep in mind that b_{DM} is L 's (correct) low belief in her optimal learning, where the subscript " DM " indicates that this is the case with the biased DM . Likewise, p_{DM} is the (correct) probability of conviction.

3.3 Whose Bias?

So far, we have shown that both an innocent and a guilty D may benefit from a biased L , but that an innocent D is always worse off when the DM is biased. Suppose we fix the level of bias. Does an innocent D always prefer that L is biased rather than the DM ? For convenience, we set $\mu_B \geq \mu$ to be the biased belief—we shall now compare the case in which $\mu_B = \mu_{DM}$ with the case in which $\mu_B = \mu_L$.

It is easy to come up with conditions that ensure an innocent D would rather L have the bias. For instance, if the conditions in Theorem 3.4 hold—that guarantee that L 's bias benefits D —then a biased L is better for D than a biased DM (since we show in Proposition 3.5 that an innocent D 's welfare is decreasing in the DM 's bias). Furthermore, if the bias is sufficiently large ($\mu_B \geq a$), then a biased DM requires no convincing and will convict based on the prior alone. Given this extreme prejudice, L will not acquire any information, and an innocent D will be convicted with probability 1. Conversely, a biased L never secures a conviction with probability 1, no matter her bias. These observations, plus the continuity of $\hat{\lambda}$ and λ in μ_B , allow us to conclude the following remark.

Remark 3.6. For any biased prior, μ_B , that is sufficiently large, an innocent D 's welfare when L is biased is strictly higher than an innocent D 's welfare when the DM is biased.

Obviously, for all μ_B sufficiently large, a guilty D is also better off when L , not the DM , is prejudiced.

These observations notwithstanding, an innocent D may still prefer that the biased party be the DM and not L . We begin with this result for the variance cost functional, $C(F) = \int \phi_{Var} dF$.

Proposition 3.7. *Assume the variance cost: $\phi = \phi_{Var}$. Let the correct prior, reward ratio, and conviction threshold (μ , d , and a , respectively) be such that if she were not biased L would acquire information ($\mu > a - d$). If a is sufficiently high and if d is sufficiently small, there exists a threshold (false) prior μ_B^\diamond , such that an innocent D prefers the DM to be biased to L if and only if $\mu_B \in (\mu, \mu_B^\diamond)$. Otherwise, an innocent D prefers L to be biased for all $\mu_B \geq \mu$.*

It is illuminating to contrast this result to Theorem 3.4, in which we establish that a large reward (high d) and a low conviction threshold (a) are required in order for bias

to benefit an innocent D . Conversely, Proposition 3.7 states that a *low* reward and a *high* conviction threshold are essential parts of an environment in which D prefers the DM 's prejudice to L 's. Regardless of who is biased, D is hurt by bias if and only if the bias causes L to over-investigate (relative to the situation when there is no bias).⁹ When the DM is prejudiced, this excess of information acquisition is guaranteed and is due to the effective decrease in the conviction threshold. The cause of L 's over-investigation due to her own bias is more subtle: it occurs when the (false) conviction probability L is increasing sufficiently quickly in μ_B . An even greater rate of increase, thus, is needed for L 's partisanship to be worse for D than the DM 's: it is then that the over-investigation L is "tricked" into by her false prior exceeds the over-investigation engendered by the conviction threshold's decrease.

Can the findings of Proposition 3.7 be generalized? Note that the conditions in this proposition ensure that the bias is not too large and that the likelihood of conviction is small; *viz.*, both the perceived likelihood, p_L , from the scenario in which L is biased; and the true likelihood, p_{DM} , from the scenario in which the DM is biased—which are close to one another when the bias is small—are small. Accordingly, the effect of a small amount of bias is approximated well by rate of change of the likelihood of conviction (again, p_{DM} versus p_L) in the size of the bias.

Theorem 3.8. *Suppose one of the following three conditions holds:*

- i The cost is the log-likelihood cost, $\phi = \phi_{LL}$, and the correct prior is low, $\mu < 1/2$;*
- ii The cost is the Tsallis cost, $\phi = \phi_T$ (which includes $\phi = \phi_{Entropy}$); or*
- iii The cost, ϕ , has bounded slope on $[0, 1]$.*

Let the conviction threshold $a = 1 - \epsilon$, $\epsilon > 0$, and define v_ϵ to be the lowest reward ratio, v , such that there is an interior solution when there is zero bias, $\mu_B = \mu$. Then, for all ϵ sufficiently

⁹As an aside, one might think that an innocent defendant should prefer more investigation to less, so that the truth is more likely to be discovered. However, this is only true if the current body of evidence is such that the defendant would be convicted. An adversarial investigator would never do such a thing, rendering an increase in scrutiny always bad in our scenario.

small, there exists an interval (v_ϵ, \bar{v}) such that an innocent D 's welfare is strictly higher when the DM is biased instead of L for all μ_B sufficiently close to μ .

In short, we find if the conviction threshold is sufficiently high, and L is very close to moving on from the suspect (abandoning her investigation of D), and bias is not too large, the DM 's bias is better for an innocent defendant than L 's. The over-investigation that arises due to L 's bias is worse for the innocent defendant than the lower conviction threshold induced by the DM 's bias.

4 What Kind of Bias?

So far, we have modeled bias as an incorrect prior. That is, we characterize bias in terms of agents' beliefs. However, bias may also manifest through agents' preferences. We can model this preference-based bias as follows: suppose now that law enforcement's payoff from conviction is of the form

$$\eta u(\alpha, \theta) + (1 - \eta)v \mathbb{1}_{\alpha=convict},$$

where $\alpha \in \{acquit, convict\}$ is the decision-maker's action and where

$$u(\alpha, \theta) = \begin{cases} 0, & \text{if } (\alpha, \theta) = (convict, 1) \text{ or } (\alpha, \theta) = (acquit, -1) \\ -1, & \text{if } (\alpha, \theta) = (acquit, 1) \\ -\rho, & \text{if } (\alpha, \theta) = (convict, -1) \end{cases},$$

where $\rho \geq 0$. We impose that the DM 's conviction threshold $a \geq \frac{\rho\eta - (1-\eta)v}{\eta(1+\rho)}$, so that the DM has a higher conviction threshold than L . When $\eta = 0$, we are in the scenario of the earlier portion of the paper: in terms of preferences, L is maximally biased in our main specification. In addition, both DM and L share the same (correct) prior μ that D is guilty.

L chooses any Bayes-plausible distribution to maximize

$$V(x) := \begin{cases} -\eta x - \phi(x), & 0 \leq x \leq a \\ -\eta\rho(1-x) + (1-\eta)v - \phi(x), & a \leq x \leq 1 \end{cases}.$$

It is easy to see that there are two varieties of interior solution: either L obtains beliefs a and b ($< \mu$) (the DM 's preferences bind) or L obtains beliefs $x_H > a$ and b ($< \mu$) (the DM 's preferences do not bind). If the DM 's preferences bind, L 's behavior (at an interior solution) is pinned down by

$$0 = \phi(b) + \eta a + (a - b)\phi'(b) - \eta\rho(1 - a) + (1 - \eta)v - \phi(a) =: s(b, a) . \quad (\Delta)$$

Henceforth, we make the following assumption, which is necessary and sufficient for L 's solution to be interior, in which the DM 's preferences bind.

Assumption 4.1. $b > 0$, $\mu \in (b, a)$, and $-\eta - \phi'(b) \geq \eta\rho - \phi'(a)$, where b is the unique root of s defined in Expression Δ .

The first result of this section states that the DM 's bias has the same effect on an innocent D 's welfare as in the main specification. Importantly, preference-based and belief-based bias for the DM have the same effect on outcomes: both lower the DM 's conviction threshold.

Proposition 4.2. *The probability that an innocent D is convicted is increasing in the DM 's bias.*

There are a number of ways of modeling L 's preference-based bias. An increase in this bias could correspond to an increase in v , or a decrease in η or ρ . Given this multiplicity, we establish the following result concerning the effect of L 's preference-based bias.

Proposition 4.3. *When an increase in bias corresponds to an increase in L 's payoff from conviction, v , or a decrease in penalty from convicting an innocent defendant, ρ , the probability that an innocent D is convicted is increasing in L 's bias.*

When an increase in bias corresponds to a decrease in the weight the DM places on getting the decision correct, η , the probability that an innocent D is convicted is increasing in L 's bias if and only if v and ρ are sufficiently large and the conviction threshold, a , is sufficiently small ($v \geq a - \rho(1 - a)$).

It is interesting to contrast this proposition with Theorem 3.4. Here, interpreting a smaller η as an increase in L 's (preference-based) bias, it is only when a is sufficiently high

and v is sufficiently low that an innocent D benefits from bias. Moreover, an innocent D never benefits from bias, when a larger v or a smaller ρ corresponds to a more biased L . But we should not be surprised that preference-based bias has such a different effect to belief-based bias. With preference-based bias, it does not matter so much who (L or the DM) is the biased party since in both cases bias works to shift the incentives of L . In contrast, if L suffers from belief-based bias, there is no change in her incentives, since she is aware of the DM 's lack of bias and tailors the high stopping threshold accordingly. Instead, her stopping problem is altered, unbeknownst to her. Accordingly, it is not surprising that under two of the three interpretations of (preference-based) bias, Propositions 4.2 and 4.3 reveal that it does not matter who has the bias: an innocent D is always worse off.

One interesting parallel between Proposition 4.3 and the earlier Proposition 3.1 is that independent of whether bias is preference- or belief-based, the guilty D is affected in the same manner as the innocent D and hence the unconditional effect of bias also has the same sign. That is, if an innocent D is made worse off by bias (regardless of whether it is preference- or belief-based), so is a guilty D . Accordingly, a change in the level of bias effects an exchange of type I errors for type II errors, or vice-versa.

5 Discussion

In this paper, we take an agent's bias as given, and explore its ramifications. Perhaps unexpectedly, our results challenge the notion that bias must necessarily be bad for a suspect: a prejudiced law-enforcement may be to suspect's benefit, due to its effect on the evidence-gathering process. On the other hand, a partisan decision-maker (judge, jury) is always worse for a defendant than a fair one. Given these observations, one might suppose that for a fixed level of bias, a defendant should always prefer that law-enforcement be biased, but our findings contest that result as well.

Under a max-min criterion, this paper has clear policy implications: efforts in eliminating prejudice should be applied to decision-makers (judges and juries), since such efforts are guaranteed to improve equity. Whether a reduction in law-enforcement's bias

does so is more nebulous.¹⁰

Furthermore, although the motivation for this paper was the specific legal environment that we explore, our findings apply to other settings in which there is delegated (and motivated) costly information acquisition. For instance, we can replace law enforcement with a pharmaceutical company who wants its drug approved by a regulator (*DM*), and who runs costly trials to obtain evidence about the drug's effectiveness. Our framework allows us to explore how the approval rate of ineffective (or dangerous drugs)—corresponding to the low state $\theta = -1$ —is affected by the subjective beliefs of the two parties. Similarly, a headhunter in a CEO search can take the role of law enforcement, and the firm looking for a CEO would be the decision-maker. In this scenario, our results shed light on the effects of discrimination both belief- and, thanks to Section 4, taste-based.

One final comment: the perspicacious reader may have picked up on the fact that *the true prior, μ , is irrelevant* in our model. Although we specify that the prior of one of the actors is correct and that the other is biased (has a higher prior) this is just to ease presentation and discussion. The conditional probabilities that innocent and guilty defendants are convicted depend *only* on the subjective beliefs held by the decision-maker and law enforcement. Naturally, the unconditional probability that a defendant is convicted is affected by the true likelihood of guilt.

References

David S Abrams, Marianne Bertrand, and Sendhil Mullainathan. Do judges vary in their treatment of race? *The Journal of Legal Studies*, 41(2):347–383, 2012.

Shamena Anwar, Patrick Bayer, and Randi Hjalmarsson. The impact of jury race in criminal trials. *The Quarterly Journal of Economics*, 127(2):1017–1055, 2012.

¹⁰Of course, there are other benefits from having a less prejudiced law-enforcement that our model does not capture. For instance, police interactions occasionally turn violent (even deadly), and our framework does not address the consequences of bias along this dimension.

- Victor Augias and Daniel Barreto. Wishful thinking: Persuasion and polarization. *arXiv preprint arXiv:2011.13846*, 2020.
- Alexander W Bloedel and Ilya Segal. Persuading a rationally inattentive agent. *Working paper*, 2020.
- Isabelle Brocas and Juan D Carrillo. Influence through ignorance. *The RAND Journal of Economics*, 38(4):931–947, 2007.
- Andrew Caplin, Mark Dean, and John Leahy. Rationally inattentive behavior: Characterizing and generalizing shannon entropy. *Forthcoming, Journal of Political Economy*, 2021.
- Yeon-Koo Che and Navin Kartik. Opinions as incentives. *Journal of Political Economy*, 117(5):815–860, 2009.
- Alma Cohen and Crystal S Yang. Judicial politics and sentencing decisions. *American Economic Journal: Economic Policy*, 11(1):160–91, 2019.
- Tanaya Devi and Roland G Fryer Jr. Policing the police: The impact of “pattern-or-practice” investigations on crime. *Mimeo*, 2020.
- Mike Felgenhauer and Petra Loerke. Bayesian persuasion with private experimentation. *International Economic Review*, 58(3):829–856, 2017.
- Mike Felgenhauer and Elisabeth Schulte. Strategic private experimentation. *American Economic Journal: Microeconomics*, 6(4):74–105, 2014.
- Mike Felgenhauer and Fangya Xu. The face value of arguments with and without manipulation. *International Economic Review*, 62(1):277–293, 2021.
- Marcelo Fernandez and Tatiana Mayskaya. Implications of overconfidence on information investment. *Available at SSRN 3038356*, 2017.
- Roland G Fryer Jr. An empirical analysis of racial differences in police use of force. *Journal of Political Economy*, 127(3):1210–1261, 2019.

Matthew Gentzkow and Emir Kamenica. Costly persuasion. *American Economic Review*, 104(5):457–62, 2014.

Emeric Henry. Strategic disclosure of research results: The cost of proving your honesty. *The Economic Journal*, 119(539):1036–1064, 2009.

Emeric Henry and Marco Ottaviani. Research and the approval process: the organization of persuasion. *American Economic Review*, 109(3):911–55, 2019.

Claudia Herresthal. Hidden testing and selective disclosure of evidence. *Mimeo*, 2017.

Mathijs Janssen. The whole truth? generating and suppressing hard evidence. *Mimeo*, 2018.

Shaofei Jiang. Costly persuasion by a partially informed sender. *Mimeo*, 2021.

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

Filip Matějka and Alisdair McKay. Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–98, 2015.

Stephen Morris and Philipp Strack. The wald problem and the equivalence of sequential sampling and static information costs. *Mimeo*, 2017.

Moses Shayo and Asaf Zussman. Judicial ingroup bias in the shadow of terrorism. *The Quarterly Journal of Economics*, 126(3):1447–1484, 2011.

Christopher A Sims. Stickiness. In *Carnegie-rochester conference series on public policy*, volume 49, pages 317–356, 1998.

Christopher A Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690, 2003.

Constantino Tsallis. Possible generalization of boltzmann-gibbs statistics. *Journal of Statistical Physics*, 1988.

Abraham Wald. Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, 16(2):117–186, 1945.

A Omitted Proofs

A.1 Lemma 3.3 Proof

Proof. A direct calculation reveals that $\underline{d}'(\mu_L)$ has the same sign as

$$(1-a)\mu(1-2\mu_L) - (a-\mu)\mu_L^2 =: t(\mu_L) .$$

Directly,

$$t'(\mu_L) = -2[(1-a)\mu + (a-\mu)\mu_L] < 0 .$$

Thus,

$$t(\mu_L) \leq t(\mu) = (1-\mu)\mu(1-\mu-a) ,$$

strictly if and only if $\mu_L > \mu$. The right-hand side of this is strictly positive if and only if $1-\mu-a > 0$. Moreover,

$$t(\mu_L) > t(1) = -a(1-\mu) < 0 .$$

Accordingly, if $1 \leq \mu + a$, \underline{d} is always decreasing in μ_L . If $1 > \mu + a$ then by the intermediate value theorem there exists some μ_L^\dagger such that \underline{d} is strictly increasing in μ_L if and only if $\mu_L \in (\mu, \mu_L^\dagger)$. ■

A.2 Theorem 3.4 Proof

Proof. A direct calculation reveals that $\lambda'(\mu_L)$ has the same sign as

$$((d-2a+1)\mu - ad + a)\mu_L + (a-1)(d+1)\mu =: f(\mu_L) .$$

Then,

$$f'(\mu_L) = (1-2a)\mu - (a-\mu)d + a =: t(d) ,$$

which is strictly decreasing in d and hence

$$t(d) > t(a) = (\mu+a)(1-a) > 0 .$$

Thus, f is strictly increasing in μ_L . Moreover,

$$f(1) = a(1 - \mu)(1 - d) > 0 .$$

Next,

$$f(\mu) = -\mu(1 - \mu)(1 + d - 2a) ,$$

which is strictly negative provided a is sufficiently small ($2a < 1 + d$). In this case, λ is u-shaped in μ_L . Moreover,

$$\lambda(\mu) = \left(\frac{1 - a}{1 - \mu} \right) \frac{\mu - a + d}{d} , \quad \text{and} \quad \lambda(1) = \frac{\mu}{a} \left(\frac{1 - a}{1 - \mu} \right) ;$$

so, $\lambda(\mu) < \lambda(1)$ if and only if

$$(a - \mu)(a - d) > 0 ,$$

which holds by assumption. Naturally, if $2a \geq 1 + d$, f (and hence λ') are always positive. ■

A.3 Proposition 3.5 Proof

Proof. First, $\hat{\lambda}'(\mu_{DM})$ has the same sign as

$$[(1 - a_{DM})p'_{DM}(a_{DM}) - p_{DM}]a'_{DM}(\mu_{DM}) .$$

Second, it is easy to derive

$$a'_{DM}(\mu_{DM}) = -\frac{(1 - \mu)\mu(1 - a)a}{((\mu - \mu_{DM})a + \mu_{DM}(1 - \mu))^2} < 0 . \quad (A3)$$

Thus, it suffices to show that $p'_{DM}(a_{DM}) \leq 0$.

Third, $p'_{DM}(a_{DM})$ has the same sign as

$$-(\mu - b_{DM}) - (a_{DM} - \mu)b'_{DM}(a_{DM}) . \quad (A2)$$

Fourth,

$$f'(b_{DM}) = (a_{DM} - b_{DM})\phi''(b_{DM}) > 0 , \quad \text{and} \quad f'(a_{DM}) = -(\phi'(a_{DM}) - \phi'(b_{DM})) < 0 ,$$

so by the implicit function theorem, $b'_{DM}(a_{DM}) > 0$, and so Expression A2 and $p'_{DM}(a_{DM})$ are negative. ■

A.4 Proposition 3.7 Proof

Proof. The support of the optimal distribution over posteriors when the DM is biased is $\{a_{DM} - d, a_{DM}\}$. A direct calculation reveals that

$$\lambda'(\mu_L)|_{\mu_L=\mu} > \hat{\lambda}'(\mu_{DM})|_{\mu_{DM}=\mu} \Leftrightarrow h(d) := 4a^2 - 3a\mu - 2a + \mu + (\mu - 2a)d > 0.$$

Since $a > \mu$, h is obviously decreasing in d . Thus, $h(d)$ is strictly positive if and only if

$$d < \frac{4a^2 - 3a\mu - 2a + \mu}{2a - \mu} =: \bar{d}.$$

Note that by assumption $d > a - \mu (> 0)$. Accordingly, we need

$$a > \frac{1 + \sqrt{1 - 2\mu + 2\mu^2}}{2} =: \bar{a}.$$

By “ a sufficiently high” and “ d sufficiently small” in the proposition’s statement, we mean $a > \bar{a}$ and $d < \bar{d}$. Next, $\lambda(\mu_B) - \hat{\lambda}(\mu_B)$ has the same sign as

$$\begin{aligned} f(\mu_B) := & \mu_B((\mu - a)\mu_B + (a - 1)\mu)^2 \left((\mu^2 + (d - 1)\mu + (a - 1)d)\mu_B - a\mu^2 + (a - ad)\mu \right) \\ & - \mu((\mu + a - 1)\mu_B - a\mu)^2 \left((\mu - a)\mu_B^2 + (d - 1)(\mu - a)\mu_B + (a - 1)d\mu \right). \end{aligned}$$

Obviously, $f(\mu) = 0$. Moreover,

$$f(a) = (a - 1)a^3(\mu - a) \left(((2a - 1)\mu - a^2)d - (a - 1)^2\mu \right) =: g(d).$$

Next,

$$g'(d) = (1 - a)a^3(a - \mu) \left((2a - 1)\mu - a^2 \right) < 0,$$

since $a > \mu$. Thus,

$$g(d) < g(0) = -(1 - a)^3 a^3 (a - \mu)\mu < 0,$$

so $f(a) < 0$. Directly,

$$\frac{\partial f}{\partial d} = \underbrace{(\mu_B - \mu)}_{\geq 0} \underbrace{((\mu - a)\mu_B + (a - 1)\mu)}_{< 0} \underbrace{((\mu - a)\mu_B - a\mu)}_{< 0} \underbrace{((\mu + a - 1)\mu_B - a\mu)}_{< 0} \leq 0,$$

strictly so when $\mu_B > \mu$. Consequently, when $d \geq \bar{d}$ and $\mu_B > \mu$, $f(\mu_B, d) < f(\mu_B, \bar{d})$, which has the same sign as

$$\begin{aligned} & - \left((\mu - a)(\mu(a(\mu - a + 2) - 1) + 2(a - 1)a(2a - 1))\mu_B^2 + (2a^2 - 4a + 1)\mu^2(\mu - a)\mu_B \right. \\ & \quad \left. - (a - 1)a(3a - 1)\mu^3 + 2(a - 1)a^2(2a - 1)\mu^2 \right) \end{aligned}$$

This expression is strictly decreasing in μ_B and equals 0 when $\mu_B = \mu$. Thus, when $d \geq \bar{d}$, f is strictly negative for all $\mu_B \in (\mu, a)$. On the other hand, when $d < \bar{d}$, we need merely show that $f(\mu_B)$ has at most one real root in the interval (μ, a) . But this is clear since f is a quartic, $f(\mu) = 0$, $f'(\mu_B)|_{\mu=\mu_B} > 0$, $f(a) < 0$, and $f(0) = a^2(1-a)d\mu^2 \geq 0$ (*viz.*, f would have to have three roots in (μ, a) , which is impossible). ■

A.5 Theorem 3.8 Proof

Proof. Let us begin with a sequence of claims. First,

Claim A.1. b_L is strictly increasing in μ_L .

Proof. Recall that b_L is implicitly defined as the root to h given in Expression 1. Then,

$$b'_L(\mu_L) = -\frac{h'(\mu_L)}{h'(b_L)} = \frac{(\phi'(a_L) - \phi'(b_L))a'_L(\mu)}{(a_L - b_L)\phi''(b_L)} > 0, \quad (A4)$$

since $a'_L(\mu_L) > 0$. ■

Second,

Claim A.2. b_L is strictly decreasing in v .

Proof. Directly, $h'(v) = 1$, so we may again appeal to the implicit function theorem to get the result. ■

Third,

Claim A.3. b_{DM} is strictly decreasing in μ_{DM} .

Proof. Recall that b_{DM} is implicitly defined as the root to f given in Expression 2. Then,

$$b'_{DM}(\mu_{DM}) = -\frac{f'(\mu_{DM})}{f'(b_{DM})} = \frac{(\phi'(a_{DM}) - \phi'(b_{DM}))a'_{DM}(\mu_{DM})}{(a_{DM} - b_{DM})\phi''(b_{DM})} < 0, \quad (A5)$$

since $a'_{DM}(\mu_{DM}) < 0$ (see Expression A3). ■

Fourth,

Claim A.4. b_{DM} is strictly decreasing in v .

Proof. Directly, $f'(v) = 1$, so we may again appeal to the implicit function theorem to get the result. ■

Using Claim A.2, it is easy to see that for any μ_L and a , there exists an interior solution if and only if v is sufficiently high. The same is true when the DM is biased (using Claim A.4). We define v_ϵ to be this minimal payoff when $a = 1 - \epsilon$ and $\mu_B = \mu$. Formally,

$$v_\epsilon := \inf\{v: b < \mu, a = 1 - \epsilon, \mu_B = \mu, b \text{ is the solution to } h(b, a, v) = 0\}.$$

Next,

Claim A.5. For all $v > v_\epsilon$, the solutions are interior with $b_L < \mu_L$ and $b_{DM} < \mu$ for all μ_B sufficiently close to μ .

Proof. Let $v > v_\epsilon$. By the continuity of b_L in μ_L , for any $\eta \in (0, \mu - b)$, there exists a $\delta > 0$ such that if $\mu_L \leq \mu + \delta$, $b_L \leq b + \eta$. By definition $b < \mu$, so let us set $\eta := (\mu - b)/2$. Thus, we get $b_L \leq b + (\mu - b)/2 < \mu < \mu_L$.

Claim A.3 immediately implies the second part of this claim. ■

Next, explicitly,

$$\begin{aligned} \lambda'(\mu_L) &= \left(\frac{1 - a_L - (1 - \mu_L) a'_L(\mu_L)}{(1 - \mu_L)^2} \right) p_L + \frac{1 - a_L}{1 - \mu_L} p'_L(\mu_L) \\ &= \underbrace{-\frac{(1 - a)\mu(a - \mu)}{((\mu_L - \mu)a + \mu(1 - \mu_L))^2} p_L}_{<0} + \frac{(1 - a)\mu}{(\mu_L - \mu)a + \mu(1 - \mu_L)} p'_L(\mu_L). \end{aligned} \quad (\text{A1})$$

By definition, when $v = v_\epsilon$, $b = \mu$. Accordingly, from Expression A1, when $v = v_\epsilon$

$$\lambda'(\mu_L)|_{\mu_L=\mu} = \frac{1 - a}{1 - \mu} p'_L(\mu_L)|_{\mu_L=\mu};$$

that is, the first term vanishes. Analogously,

$$\hat{\lambda}'(\mu_{DM})|_{\mu_{DM}=\mu} = \frac{1 - a}{1 - \mu} p'_{DM}(\mu_{DM})|_{\mu_{DM}=\mu}.$$

Directly,

$$\begin{aligned} p'_L(\mu_L)|_{\mu_L=\mu} &= \frac{a(1 - b'_L(\mu_L)|_{\mu_L=\mu}) - b(1 - a'_L(\mu_L)|_{\mu_L=\mu}) - \mu_L(a'_L(\mu_L)|_{\mu_L=\mu} - b'_L(\mu_L)|_{\mu_L=\mu})}{(a - b)^2}, \\ &= \frac{1 - b'_L(\mu_L)|_{\mu_L=\mu}}{a - \mu} \end{aligned}$$

when $v = v_e$. Similarly,

$$\begin{aligned} p'_{DM}(\mu_{DM})|_{\mu_{DM}=\mu} &= \frac{-ab'_{DM}(\mu_{DM})|_{\mu_{DM}=\mu} + ba'_{DM}(\mu_{DM})|_{\mu_{DM}=\mu} - \mu(a'_{DM}(\mu_{DM})|_{\mu_{DM}=\mu} - b'_{DM}(\mu_{DM})|_{\mu_{DM}=\mu})}{(a-b)^2} \\ &= -\frac{b'_{DM}(\mu_{DM})|_{\mu_{DM}=\mu}}{a-\mu} \end{aligned}$$

when $v = v_e$. Since $a < 1$, $\lambda'(\mu_L)|_{\mu_L=\mu} - \hat{\lambda}'(\mu_{DM})|_{\mu_{DM}=\mu}$ has the same sign as $p'_L(\mu_L)|_{\mu_L=\mu} - p'_{DM}(\mu_{DM})|_{\mu_{DM}=\mu}$, which has the same sign as

$$1 - b'_L(\mu_L)|_{\mu_L=\mu} + b'_{DM}(\mu_{DM})|_{\mu_{DM}=\mu},$$

which equals, using Expression A3 and Equations A4 and A5,

$$1 - 2(1-a)a \underbrace{\frac{\phi'(a) - \phi'(\mu)}{(a-\mu)\mu(1-\mu)\phi''(\mu)}}_{q(a)}. \quad (A7)$$

If ϕ' is bounded, then so is ϕ'' . Thus, as $a \nearrow 1$, $q(a) \searrow 0$. On the other hand, if $\phi = \phi_{LL}$ Expression A7 becomes

$$1 - 2(1-a)a\mu(1-\mu) \underbrace{\frac{\phi'_{LL}(a) - \phi'_{LL}(\mu)}{(a-\mu)}}_{q_{LL}(a)}.$$

As $a \nearrow 1$, $q_{LL} \searrow 2\mu$. Finally, if $\phi = \phi_T$, Expression A7 becomes

$$1 - 2(1-a)a \underbrace{\frac{\phi'_T(a) - \phi'_T(\mu)}{(a-\mu)}}_{q_T(a)}.$$

As $a \nearrow 1$, $q_T \searrow 0$.

The result then follows from the continuity of the solution in the parameters. ■

A.6 Proposition 4.2 Proof

Proof. Here, the conditional probabilities are

$$\gamma = \frac{a}{\mu}p, \quad \text{and} \quad \lambda = \frac{1-a}{1-\mu}p, \quad \text{where} \quad p = \frac{\mu-b}{a-b}.$$

Directly, $\lambda'(a)$ has the same sign as

$$-p + (1 - a)p'(a) .$$

Consequently, we need to show that $p'(a) \leq 0$. As in the proof of Proposition 3.5, it suffices to show that $b'(a) \geq 0$. To that end, we compute

$$s'(b) = (a - b)\phi''(b) > 0 ,$$

and

$$s'(a) = \eta(\rho + 1) - (\phi'(a) - \phi'(b)) \leq 0 ,$$

by Assumption 4.1. Finally, by the implicit function theorem, $b'(a) \geq 0$. ■

A.7 Proposition 4.3

Proof. $\lambda'(v)$ has the same sign as $p'(v)$, which has the same sign as $-b'(v)$. Analogously, $\lambda'(\rho)$ has the same sign as $-b'(\rho)$ and $\lambda'(\eta)$ has the same sign as $-b'(\eta)$. From the proof of Proposition 4.2, $s'(b) > 0$. Directly, $s'(v) = 1 - \eta \geq 0$ and $s'(\rho) = -\eta(1 - a) \leq 0$. Accordingly, the implicit function theorem yields that $b'(v) \leq 0$ and $b'(\rho) \geq 0$.

Directly,

$$s'(\eta) = a - \rho(1 - a) - v ,$$

and appealing to the implicit function theorem, $s'(\eta)$ has the same sign as $\lambda'(\eta)$. ■