Adversarial decision-making: Choosing between models constructed by interested parties

Froeb, Luke M. and Ganglmair, Bernhard and Tschantz, Steven

Vanderbilt University, University of Texas at Dallas, Vanderbilt University

28 March 2016

Adversarial Decision-Making: Choosing Between Models Constructed by Interested Parties*

Luke M. Froeb† Bernhard Ganglmair‡ Steven Tschantz§

March 28, 2016

Abstract

In this paper, we characterize adversarial decision-making as a choice between competing interpretations of evidence ("models") constructed by interested parties. We show that if a court cannot perfectly determine which party's model is more likely to have generated the evidence, then adversaries face a tradeoff: a model further away from the best (most likely) interpretation has a lower probability of winning, but also a higher payoff following a win. We characterize equilibrium when both adversaries construct optimal models, and use the characterization to compare adversarial decision-making to an inquisitorial benchmark. We find that adversarial decisions are biased, and the bias favors the party with the less-likely, and more extreme, interpretation of the evidence. Court bias disappears when the court is better able to distinguish between the likelihoods of the competing models, or as the amount of evidence grows.

JEL classification: C72; D74; K41

Keywords: adversarial justice; evidence-based decision-making; expert testimony; inquisitorial justice; litigation; persuasion games; science vs. advocacy.

---

*An earlier version of this paper was circulated under the title "A Theory of Competitive Framing."

†Vanderbilt University, Owen Graduate School of Management, 401 21st Avenue South, Nashville, TN 37203, USA. e-mail: luke.froeb@owen.vanderbilt.edu

‡The University of Texas at Dallas, Naveen Jindal School of Management, 800 W. Campbell Rd. (SM31), Richardson, TX 75080, USA. e-mail: ganglmair@utdallas.edu

§Vanderbilt University, Department of Mathematics, Nashville, TN 37240, USA. e-mail: tschantz@math.vanderbilt.edu

1 Introduction

Adversarial justice has two stages: (i) information provision, where information is acquired and reported to the court; and (ii) decision-making, where a court makes a decision to resolve the dispute (Iossa and Palumbo, 2007). Much of the economic literature comparing adversarial to inquisitorial justice focuses on the first stage, where evidence is produced and reported by adversaries, rather than an impartial third party (Milgrom and Roberts, 1986; Shin, 1998; Dewatripont and Tirole, 1999; Froeb and Kobayashi, 2001; Daughety and Reinganum, 2000b; Skaperdas and Vaidya, 2012; Froeb and Kobayashi, 2012; Rantakari, 2016). Although these articles address different aspects of the adversarial system, they all find that competition between the adversaries plays a crucial role in the ability of a court to gather information.

However, the adversarial system also differs from the inquisitorial in the second, decision-making stage: instead of having to choose between competing interpretations of evidence constructed by interested parties ("adversarial"), a court can instead appoint a neutral expert to interpret the evidence for them ("inquisitorial"). A shift to the second option is probably the most commonly called-for reform of the (adversarial) justice system in the United States (Fienberg and Straf, 1991; Froeb and Kobayashi, 2001; Wagner, 2005), especially for scientific or statistical evidence where the court often lacks "knowledge and expertise ... and therefore has to delegate the job to a qualified expert" (Ambrus et al., 2015). For example, antitrust merger trials often involve opposing expert economists who construct oligopoly models to predict post-merger prices (Werden and Froeb, 1994; Tenn et al., 2010). Lay competition tribunals are called on to assess the relative credibility of the two models, even though constructing a model would be beyond the tribunal's capability.

More generally, think about two litigants preparing for trial. Evidence has already been produced and discovered, and the opposing attorneys are devising strategies to win in court. As first-year law students are taught (Tanford, 2009):

> It is your job to sort the information before trial, organize it, simplify it and present it to the jury in a *simple model that explains what happened and why you are entitled to a favorable verdict.*
>
> Remember that there is a lawyer on the other side who will be trying to sell the jury *a story that contradicts yours.* ... If both sides do competent jobs, the *jury will have to choose between two competing versions of events* ....[emphasis added]

In this paper, we characterize the resulting trial as a *persuasion game.*[1] Unlike other persuasion games, where agents have private information and take actions to persuade a principal (i.e., by strategically revealing evidence), in our game, all the evidence is known, and litigants compete by proposing *models* to explain what it means.[2]

[1]Our persuasion game assumes symmetric information across all players. Scenarios in which two agents with private information (and potentially competing interests) try to influence a decision-making principal have been studied by Gilligan and Krehbiel (1989), Glazer and Rubinstein (2001), Krishna and Morgan (2001a), Krishna and Morgan (2001b), Gentzkow and Kamenica (2016), and Rantakari (2016).

[2]In the context of litigation, Mauet (2007:24) describes such a model or *theory of a case* as a "simple story of 'what really happened'." This kind of decision-making could also be motivated by institutional constraints, like the constitutional guarantees of the adversarial process. The *Case and Controversy Clause* of the U.S. Constitution (art. III, sec. 2) limits the court to deciding actual controversies, which are framed by the litigants.

Deciding between the competing interpretations of the evidence is analogous to the statistical problem of "model selection," where a researcher observes data, and then tries to determine which of two models generated it. If we think of the court as a researcher and the evidence as data, then the adversaries' models can be parameterized as probability distributions that describe the evidence-generating process. The court measures the "credibility" of each model by its statistical likelihood and decides in favor of the most likely model.

This decision rule puts adversaries on the horns of a dilemma: a model with a "location" (mean) further from the evidence is less likely to prevail but has a higher payoff if it does. We use distribution families that also have a "spread" (variance) parameter, which makes credibility (likelihood) a choice variable. For example, a player can reduce the likelihood penalty (credibility cost) of choosing a model with a location further from the evidence by also choosing a model with a bigger spread. A bigger spread makes it more likely that the evidence would be located far from the mean, which is equivalent to saying that evidence is not very informative about the mean.

If the court can, through the process of discovery and rebuttal, perfectly rank the likelihoods of the two competing models, then competition between the adversaries forces each party to propose the same, most likely model. This is because any "shading" (i.e., interpretation of the evidence away from the most likely model) can be bested by a more likely model with a higher expected payoff. This leads to an equilibrium in which each party chooses the same, most likely model. In other words, competition between the adversaries leads the court to the best interpretation of the evidence, even though it has to choose between interpretations constructed by interested parties. A similar result comes from final-offer arbitration, where parties make the same utility maximizing offer (Crawford, 1979).

If, instead, the court measures likelihood with error and, as a result, cannot perfectly determine which party's model is more likely to have generated the evidence, then each party shades its model away from the most likely model.

Bias arises if the likelihood is not symmetric, which changes the tradeoff between credibility and payoff for each of the parties. As a result, each party "shades" its model away from the most likely model by differing amounts. In this case, equilibrium decisions are biased, and the bias favors the party with the less likely, and more extreme, interpretation. As such, there is no reason to expect the court's decision to have desirable properties, even when based on evidence. However, bias disappears as the court gets better at distinguishing between the likelihoods of competing models, or as the amount of evidence grows.

We present our main results using a simple litigation game. We introduce the model and notation in Section 2 and illustrate our main results using a parameterized version of the model in Section 3. In the Technical Appendix, we generalize the model and provide more formal statements of our results. We conclude this paper with a brief discussion in Section 4.

2 Litigation as a Persuasion Game

Imagine that a plaintiff (P) sues a defendant (D), and the issue before the court is the level of damages.[3] Before trial, evidence $\bar{z} = (z_1, z_2, ..., z_n)$ is produced and discovered. We parameterize the evidence-generating process so that the level of damages corresponds to the mean $\mu_{_Z} \geq 0$ of an unknown distribution Z belonging to some family $\mathcal{F}$. Such a distribution Z can be interpreted as a *model* that explains the evidence and how it relates to the damages. We assume there is a unique model $Z_{_{ML}} \in \mathcal{F}$ that best explains the evidence (has the highest likelihood given the evidence $\bar{z}$):

$$Z_{_{ML}} \equiv \arg\max_{Z \in \mathcal{F}} \mathscr{L}(Z|\bar{z}) \tag{1}$$

where $\mathscr{L}(Z|\bar{z}) = \prod_{i=1}^{n} f(z_i|Z)$. The level of damages associated with the most likely model is $\mu_{_{ML}}$.

In this litigation game, both parties simultaneously offer competing models, $Z_P \in \mathcal{F}$ and $Z_D \in \mathcal{F}$, to explain the evidence and the respective level of damages. The court (as the decision-maker in this contest) evaluates the model likelihoods $\mathscr{L}_P = \Pr(Z_P|\bar{z})$ and $\mathscr{L}_D = \Pr(Z_D|\bar{z})$ and awards damages of $\mu_{_P} \geq 0$ or $\mu_{_D} \geq 0$. The expected payment from the defendant to the plaintiff is

$$\hat{\mu} = \theta\mu_{_P} + \big(1 - \theta\big)\mu_{_D} \tag{2}$$

[3]The results do not change if we fix the level of damages and let the court draw inference about liability, modeled as a probability of a plaintiff win. In this case, expected damages are the probability of a plaintiff win times the level of damages.

where θ is the probability that the court finds in favor of the plaintiff. We assume that the court's payoffs are directly linked to the accuracy of the decision.[4] In other words, the court's objective is to get it right, and it therefore decides in favor of the party whose model is best supported by the evidence.[5]

Note that multiple models can generate the same level of damages (i.e., different distributions have the same mean). Because the plaintiff's payoffs are strictly increasing in $\hat{\mu}$, equation (2) implies that, for any given level of damages μ_P, the plaintiff chooses the strongest model (i.e., with the highest likelihood) to maximize the value of θ and thus *maximize* $\hat{\mu}$. Likewise, the defendant (whose payoffs are strictly decreasing in $\hat{\mu}$) chooses the strongest model to *minimize* θ and thus minimize $\hat{\mu}$.[6] Optimally chosen models trade off a higher payoff following a win for a lower probability of winning.[7]

[4] This can, for instance, be due to career concerns, as in Farber and Bazerman (1986:1506), who argue that arbitrators "attempt to make awards that maximize the probability they will be hired in subsequent cases" and compromise to maintain "acceptability with both parties." Iossa and Jullien (2012) take an approach similar in spirit. They assume a judge maximizes an external evaluator's posterior belief about the judge's correct (i.e., most credible) decision. Daughety and Reinganum (2000a) model the behavior of a trial court that is constrained by "higher court review." See Choi et al. (2012) for related evidence on federal district judges.

[5] We interpret the expression in equation (2) as the expectation of the court's decision, wherein the court chooses one of the two proposed models by the litigants. Outside of litigation, this approach is akin to final-offer arbitration (Wittman, 1986) or the "closed rule" in legislation (Austen-Smith, 1993). Our results do not change if, instead, we assume the court's damages assessment is the weighted average of the litigants' claims, where θ is the weight of the plaintiff's claim μ_P. Outside of litigation, this is akin to conventional arbitration, or an "open rule" in legislation, in which the legislative body can amend proposals submitted.

[6] Viewed differently, among the models with the same likelihood, the plaintiff chooses the model that is associated with the highest damages to maximize $\hat{\mu}$, whereas the defendant chooses the model that is associated with the lowest damages to minimize $\hat{\mu}$. Pardo and Allen (2008:234) cite evidence suggesting that "juries asume in most cases the parties have put forward the explanation that best helps their case."

[7] The approach taken by Kartik et al. (2007), Kartik (2009), or Emons and Fluet (2009) is related but differs in one key feature. In their models, players try to sway a decision-maker by tampering with the evidence at a direct cost, whereas in our setting, the facts are given

2.1 A Perfect Court as an Inquisitorial Benchmark

For our benchmark, we suppose the court can perfectly (i.e., without noise) rank-order the likelihoods of the parties' proposed models. The probability of a plaintiff win in this case[8] is

$$\theta = \begin{cases} 1 & \text{for } \mathscr{L}_P > \mathscr{L}_D \\ 1/2 & \text{for } \mathscr{L}_P = \mathscr{L}_D \\ 0 & \text{for } \mathscr{L}_P < \mathscr{L}_D, \end{cases} \tag{3}$$

and the expected payment from the defendant to the plaintiff is

$$\hat{\mu} = \begin{cases} \mu_P & \text{for } \mathscr{L}_P > \mathscr{L}_D \\ 1/2\, \mu_P + 1/2\, \mu_D & \text{for } \mathscr{L}_P = \mathscr{L}_D \\ \mu_D & \text{for } \mathscr{L}_P < \mathscr{L}_D. \end{cases} \tag{4}$$

With these payoffs, in equilibrium, the adversaries both choose the same, most likely model: $Z_P = Z_D = Z_{ML}$. To see this, note that any model proposed by the plaintiff with damages higher (and a likelihood lower) than the most likely model Z_{ML} gives the defendant an opportunity to choose a more likely model (i.e., "closer" to Z_{ML}) with lower damages and thus a higher payoff for the defendant. A similar argument applies to the plaintiff.[9] In equilibrium,

---

and litigants try to influence the decision in their favor by *tampering* with the interpretation of those facts.

[8] We assume the court has no ex ante bias favoring one of the two parties. If the likelihoods are the same, the court tosses a fair coin.

[9] The defendant's best response to a plaintiff's model Z_P with $\mu_P > \mu_{ML}$ so that $\mathscr{L}_P < \mathscr{L}_{ML}$ and $\theta = 1$ is a model Z_D with a mean μ_D such that $\mu_{ML} \geq \mu_D$ and $\mathscr{L}_P < \mathscr{L}_D \leq \mathscr{L}_{ML}$

both parties choose the same, most likely model, and the court awards μ_{ML} with probability one. Heuristically, the opposing interests of the parties allow the court to reach the best (most likely) interpretation of the evidence.

This outcome of adversarial litigation with a perfect court is also the maximum likelihood estimator whose well-known optimality properties (DeGroot, 1970) make it an obvious choice for an *inquisitorial* court with the ability to interpret the evidence itself. The maximum likelihood estimator serves as the benchmark against which bias is measured. We use the term *bias* to measure deviations from the best (most likely) interpretation or model of the evidence, $\Delta\mu = \hat{\mu} - \mu_{ML}$. In our benchmark case of a perfect court, it does not matter whether the court uses inquisitorial or adversarial decision-making because both result in the same, best outcome.

2.2 Noisy Courts

The characterization of adversarial justice with a perfect court is simple and results in an optimal decision, but it cannot explain the salient feature of the adversarial system: the competing interpretations of evidence put forward by the parties. Adding noise to the court's decision-making means that the court sometimes makes errors by selecting the less likely alternative. Such errors make it optimal for the parties to shade their interpretations away from the most likely interpretation because there is some probability that a less likely claim will be chosen. We motivate the addition of noise with the limited ability

so that $\theta = 0$. In equilibrium, both parties' choose the most likely model. A deviation Z'_s (so that $\mu'_s \neq \mu_{ML}$) renders the resulting likelihood $\mathscr{L}'_s < \mathscr{L}_{ML}$ and the probability of winning equal zero.

of humans (and courts) to "detect signals,"[10] as, for instance, in Mueller and Weidemann (2008).

The *perceived* likelihoods of the proposed models are the product of signal and noise:

$$\widetilde{\mathscr{L}_P} = \mathscr{L}_P \exp \xi_P \qquad \text{and} \qquad \widetilde{\mathscr{L}_D} = \mathscr{L}_D \exp \xi_D \tag{5}$$

where ξ_P and ξ_D are independently extreme value distributed *noise*, with mean 0 and scale $1/\lambda$. The probability of a plaintiff win (i.e., the logit choice probability for model P)[11] is

$$\begin{aligned} \tilde{\theta} = \Pr(\widetilde{\mathscr{L}_P} > \widetilde{\mathscr{L}_D}) &= \frac{\exp(\lambda \log \mathscr{L}_P)}{\exp(\lambda \log \mathscr{L}_P) + \exp(\lambda \log \mathscr{L}_D)} \\ &= \frac{\mathscr{L}_P^\lambda}{\mathscr{L}_P^\lambda + \mathscr{L}_D^\lambda}. \end{aligned} \tag{6}$$

This is similar to the approach taken by McKelvey and Palfrey (1995) for modeling optimal strategies when players make errors in non-cooperative games,[12] and to that taken by McFadden (1974) for modeling consumer choice among discrete alternatives.

The probability in equation (6) is also equivalent to Tullock's general for-

[10] For example, the statistician Irving J. Good, who worked with Alan Turing to break the Enigma code, thought that a change in an odds ratio from evens to about 5:4 is about as finely as humans can reasonably perceive their degree of belief in a hypothesis in everyday use (Good, 1979).

[11] See Train (2009:36-37,74-75) for the formal steps of deriving expression (6) for a scale parameter of $\lambda = 1$. Our distributional assumption for the random variable ξ_s is more restrictive than necessary. For a more general characterization, it suffices to assume that the random variable ξ_s (i.e., noise) belongs to the *inverse exponential distribution* (Jia, 2008).

[12] The notion of a *perfect* court without errors, $\lambda \to \infty$, corresponds to a *perfectly rational* decision-maker, whereas a noisy court with finite λ is said to be *boundedly rational*.

mula for the *contest success function* in rent-seeking contests (Tullock, 1980), with the likelihoods $\mathscr{L}_s$ playing the role of "effort." In our game, a party exerts higher effort by choosing a model Z_s with a higher likelihood, which results in a higher chance of winning.[13] This increase in the success probability, however, comes at the cost of a lower payoff following a win, because a higher likelihood $\mathscr{L}_s$ implies a location parameter μ_s closer to the maximum-likelihood location μ_{ML}. Unlike the standard Tullock contest, payoffs (damages) are not exogenous but rather are chosen optimally by the parties.[14]

In Table 1, we summarize three special cases of the contest success function in equation (6), for different values of the noise parameter λ.[15] For $\lambda = 0$, adversarial decision-making is uninformative, and the court's decision is equivalent to a toss of a fair coin where the probability of a plaintiff win is $\tilde{\theta} = 1/2$. In this case, the court's decision is independent of both the players' proposed models as well as the evidence.[16] For $\lambda = 1$, equation (6) becomes the lottery version of the general Tullock contest success function. In addition, it is equivalent to Bayesian hypothesis testing for a court that assigns equal

[13] Others have modeled effort in litigation as the number of arguments presented in court (Katz, 1988), litigation expenditure (Hirshleifer and Osborne, 2001), or quality of the case (Baye et al., 2005).

[14] For contests with endogenous payoffs that depend on parties' efforts, see, for instance, Chung (1996), Skaperdas (1996), Konrad and Schlesinger (1997), Kaplan et al. (2002), or Ambrus et al. (2015). Chowdhury and Sheremeta (2011) provide a generalized version of the Tullock contest that nests a range of contest models, including one in which, as in our paper, a party i's effort affects only party i's but not party j's payoff.

[15] Alternative interpretations of the noise parameter λ are found in Skaperdas and Vaidya (2012:473) (measure of the "sensitivity of the court to the evidence") and in Hirshleifer and Osborne (2001) (measure of effectiveness of "legal effort" relative to the (subjective) "power of truth").

[16] This case corresponds to broad jury instructions in Hirshleifer and Osborne (2001:174). In this scenario, the litigants' "power of advocacy" (i.e., persuasion) is diminished, and the outcome of litigation is based entirely on an objective fault parameter, independent of the litigants' proposed models.

Table 1: Litigation Contest Success Function With Noisy Court

$\lambda = 0$	Coin toss; uninformative decision-making
$\lambda = 1$	Tullock lottery; Bayesian hypothesis testing
$\lambda \to \infty$	*Perfect Court* benchmark; all-pay auction

prior weight to the competing models of the parties (Saks and Neufeld, 2011; Cheng, 2013). After seeing the evidence, the court updates its beliefs about which party is correct using *Bayes' rule*. The posterior odds are equal to the prior odds times the likelihood ratio,

$$\frac{\tilde{\theta}}{1-\tilde{\theta}} = \frac{\gamma}{1-\gamma} \cdot \frac{\mathscr{L}_P}{\mathscr{L}_D}$$

where γ is the prior weight the court places on the plaintiff's model (before seeing the evidence). Setting the prior weight $\gamma = 1/2$ (so that the court is unbiased from an ex ante point of view) and rearranging, we get:

$$\tilde{\theta} = \frac{\mathscr{L}_P}{\mathscr{L}_P + \mathscr{L}_D} \tag{7}$$

which is equivalent to equation (6) for $\lambda = 1$.

For the third scenario in Table 1, as $\lambda \to \infty$, the court can perfectly assess the relative likelihoods of the parties' models, and so awards the item to the more likely model with probability one. This is our benchmark case of a *perfect court* and is essentially a first-price all-pay auction (Baye et al., 1996).

The court's decision is an estimator of the mean, a likelihood-weighted average of the means of the parties' models,

$$\hat{\mu}(Z_P, Z_D) = \tilde{\theta}\mu_P + \big(1 - \tilde{\theta}\big)\mu_D = \frac{\mathscr{L}_P^{\lambda} \cdot \mu_P + \mathscr{L}_D^{\lambda} \cdot \mu_D}{\mathscr{L}_P^{\lambda} + \mathscr{L}_D^{\lambda}}. \tag{8}$$

The estimator depends on the evidence and on λ, the variance of the noise in the court's assessment of likelihood.

This award $\hat{\mu} = \hat{\mu}(Z_P, Z_D)$ also minimizes a quadratic loss function, that is, the weighted sum of the squared deviations of decision $\hat{\mu}$ from the parties' proposed damages μ_s: $w_P\left(\mu_P + \hat{\mu}\right)^2 - w_D\left(\mu_D - \hat{\mu}\right)^2$. This latter approach is similar to the one used by Farber and Bazerman (1986) to determine the "appropriate" or "ideal" award in conventional arbitration. Unlike their approach, however, our weights are endogenous and driven by the tradeoff of a higher payoff following a win for a lower probability of winning.

The court's decision is simply a binomial random variate with probability $p = \tilde{\theta}$ and variance

$$\mathrm{Var}(\hat{\mu}) = \tilde{\theta}\big(1 - \tilde{\theta}\big)\big(\mu_P - \mu_D\big)^2. \tag{9}$$

The further apart the locations of the adversaries' models are, and the closer $\tilde{\theta}$ is to $1/2$, the larger the variance of the court's decision is.[17]

[17] This assumes that the court chooses either μ_P or μ_D. None of our results on bias change if the court *splits the baby* by awarding the expected value of the estimate instead, $\tilde{\theta}\mu_P + \big(1 - \tilde{\theta}\big)\mu_D$. The variance of the expected value is, of course, zero.

3 Equilibrium Results

We relegate the formal results of the persuasion game to the Technical Appendix. In this Section, we illustrate our main results for a parameterized version of the model. Evidence $\bar{z}$ is a vector of n independent draws $z_i \in (0,1)$ with sample mean $\bar{\mu}$ from the same Beta(α, β) distribution with mean $\mu = \mathrm{E}(z_i) = \alpha/(\alpha+\beta)$ and variance $\sigma^2 = \mathrm{Var}(z_i) = \alpha\beta/[(\alpha+\beta)^2(1+\alpha+\beta)]$. Each party $s = P, D$ chooses a model $Z_s = (\alpha_s, \beta_s)$ (i.e., the parameters for the Beta distribution) to explain the evidence vector $\bar{z}$, with any desired mean $0 < \mu_s < 1$ and any variance $\sigma_s^2 = \mu_s(1-\mu_s)/(1+\alpha_s+\beta_s)$ with $0 < \sigma_s^2 < \mu_s(1-\mu_s)$. A Nash equilibrium in pure strategies of this zero-sum game is a strategy profile (Z_P^*, Z_D^*) such that each party chooses an optimal model, given the model chosen by its rival:

$$\left.\begin{array}{lll} \text{Plaintiff:} & \hat{\mu}(Z_P^*, Z_D^*) \geq \hat{\mu}(Z_P, Z_D^*) & \forall Z_P = (\alpha_P, \beta_P) \in \mathbb{R}_+^2 \\ \text{Defendant:} & \hat{\mu}(Z_P^*, Z_D^*) \leq \hat{\mu}(Z_P^*, Z_D) & \forall Z_D = (\alpha_D, \beta_D) \in \mathbb{R}_+^2 \end{array}\right\} \quad (10)$$

To illustrate our results, we show three different types of experiments: first, we fix an evidence vector $\bar{z} = (1/5, 1/2)$. We characterize the parties' equilibrium models and show how they affect the court's decision. Second, we show how quickly the court's decision-making improves as the "decision noise" shrinks to zero. Third, we show how quickly the court's decision-making improves as the amount of evidence grows, or alternatively, as "sampling noise" shrinks to zero. In statistics, the third experiment determines the "consistency" of an estimator.

In Figure 1, we assume that the variance of the court's decision noise is $\lambda = 1$, which makes the court's decision equivalent to Bayesian hypothesis testing, where the court updates its prior belief (about relative merits of the two sides) with the likelihood to form a posterior. This assumption allows us to graph the court's posterior belief as a weighted average of the parties' models, where the weights are proportional to the likelihoods of each model. Note that we are *not* saying the court uses Bayesian inference, only that for $\lambda = 1$, the final assessment of the court $\hat{\mu}$ is also the mean of a posterior belief, as if formed by Bayesian inference.

Result 1. *Both parties "shade" the evidence in their favor so that* $0 < \mu_D^* < \mu_{ML} < \mu_P^* < 1$.

In Figure 1, we plot the parties' equilibrium models, represented by the density functions $f(z; \alpha_s, \beta_s)$, and the respective means μ_s when the evidence sample consists of two draws, $\bar{z} = (1/5, 1/2)$. The plaintiff chooses a model with a mean that is above the maximum likelihood estimate, while the defendant chooses one with a mean that is below it.

For $\lambda > 0$, the plaintiff engages in *payoff moderation* (e.g., Konrad, 2009), essentially trading off a higher payoff following a win for a lower probability of winning.[18] In panel (a) of Figure 2, for example, we plot the likelihood ratio of the plaintiff's model relative to that of the defendant. The graph shows that, in equilibrium, the plaintiff is willing to accept a lower likelihood of winning,

[18] *Payoff moderation* means that neither player chooses a model with extreme means so that $\mu_s \in (0, 1)$. Without the likelihood penalty of shading when $\lambda = 0$ so that the court sides with the plaintiff 50% of the time, irrespective of the parties' claims, the parties construct models with the most extreme claims possible so that $\mu_s \in \{0, 1\}$.

Figure 1: Parties' Equilibrium Claims in the Litigation Game

This figure illustrates the equilibrium of the litigation game for evidence $\bar{z} = (1/5, 1/2)$, indicated by two cross marks on the horizontal axis. Pieces of evidence $z_i \in (0, 1)$ are random draws from a $\mathrm{Beta}(\alpha, \beta)$ distribution with density function $f(z; \alpha, \beta) = \frac{1}{\mathrm{B}(\alpha,\beta)} z^{1-\alpha}(1-z)^{1-\beta}$, where $\mathrm{B}(\alpha, \beta)$ is the Beta function. The dashed curve represents the density $f(z; \alpha_P, \beta_P)$ of the plaintiff's equilibrium model; the dotted curve represents the density $f(z; \alpha_D, \beta_D)$ of the defendant's equilibrium model. The means of the two models are depicted by the dotted and dashed vertical lines. The mean of the maximum likelihood μ_{ML} is represented by the vertical dotted-dashed line. The court's equilibrium decision $\hat{\mu}$ is depicted by the solid vertical line. The equilibrium is characterized by the following parameters:

α_P	β_P	μ_P	σ_P^2	α_D	β_D	μ_D	σ_D^2	$\tilde{\theta}$	$\hat{\mu}$	μ_{ML}
2.397	2.326	0.507	0.044	2.026	5.981	0.253	0.021	0.430	0.362	0.349

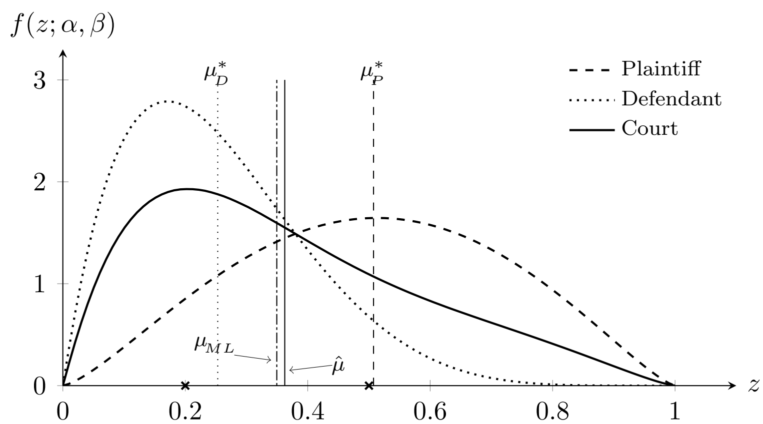
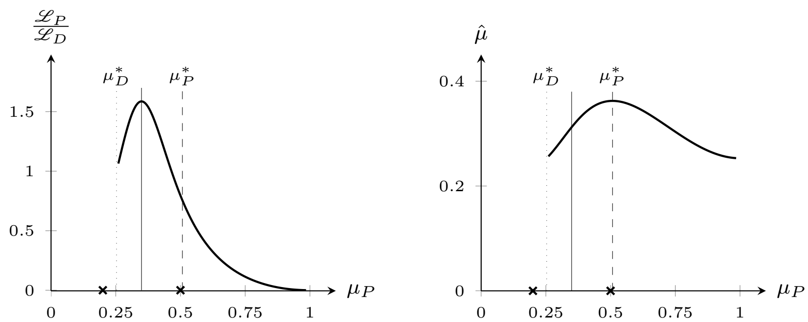


Figure 2: Payoff Moderation as a Best Response

The figure illustrates the plaintiff's best response in the litigation game for evidence $\bar{z} = (1/5, 1/2)$, indicated by two cross marks on the horizontal axis. In panel (a), we fix the defendant's equilibrium model Z_D^* and plot the likelihood ratio $\mathscr{L}_P/\mathscr{L}_D$ for given values of μ_P. In panel (b), we fix the defendant's equilibrium model Z_D^* and derive the plaintiff's optimal response. We plot the court's decision $\hat{\mu}$ for given values of mean μ_P of the plaintiff's model. In both panels, the vertical lines depict the mean μ_D^* of the defendant's equilibrium model (dotted) and the mean μ_P^* of the plaintiff's equilibrium model (dashed), as well as the mean of the maximum likelihood (solid).



(a) Claim vs. $\mathscr{L}$-Ratio (b) Claim vs. Decision

relative to the most likely model in exchange for a higher payoff following a win. Panel (b) of Figure 2 plots the profit function of the plaintiff's location, for a fixed value of Z_D^*, the equilibrium model for the defendant. As in panel (a), we see that in equilibrium, the plaintiff optimally chooses a mean μ_P above maximum likelihood mean, but strictly less than unity, $\mu_{ML} < \mu_P^* < 1$.

Result 2. *The party with less favorable evidence follows an "obfuscation strategy" and chooses a model with (i) a location further away from the most likely model and (ii) with a spread larger than its rival's.*

For example, the evidence vector $\bar{z} = (1/5, 1/2)$ in Figure 1 favors the defendant. In this case, the plaintiff optimally chooses a model with a location further from the most likely model $|\mu_P^* - \mu_{ML}| = 0.158 > 0.096 = |\mu_D^* - \mu_{ML}|$, and with a higher variance, $\sigma_P^* = 0.044 > 0.021 = \sigma_D^2$. We easily see this in

Figure 1 by comparing the spread of the density of the plaintiff's model (large) to that of the defendant's (small). A model with a larger spread is more likely to generate evidence further away from its mean.

The intuition for this result is simple. As long as the sample mean $\bar{\mu} \neq 0.5$, the likelihood is asymmetric, which changes the tradeoff between the probability of winning and payoff following a win. For the disfavored party, it lowers the likelihood penalty, or "credibility cost," of claiming a model with a location further from the evidence. Anyone who has participated in litigation or has experience in political campaigns will recognize this as analogous to what is known as an "obfuscation strategy." When the evidence goes against you, your best move is to claim that the evidence is not very informative. The defendant's best response is to choose a model with a mean closer to the evidence and with a smaller spread. This might be called an "elucidation strategy," as the defendant is essentially claiming that the evidence is informative about the mean.

Result 3. *The court's assessment of liability is biased in favor of the party with, on average, less favorable evidence.*

We illustrate this result in Figure 1, where the evidence favors the defendant, $\bar{\mu} < 1/2$, and the court's decision is just above the maximum likelihood estimate, $\hat{\mu} > \mu_{ML}$. This "bias" favors the plaintiff who, in equilibrium, offers a more extreme interpretation of the evidence.[19] However, the court's bias is

[19]The general characterization of this result in Theorem A3 in the Technical Appendix does not explicitly refer to less or more favorable evidence but to the "credibility cost" of shading the model location away from the most likely model location. For Result 3, if the evidence is closer to the lower range of the Beta(α, β) distribution (so that the evidence is

small compared to the widely disparate models of the parties, which give the court's estimator a big variance, as computed in equation (9).

Result 4. *As court noise disappears, $\lambda \to \infty$, (i) the parties' models converge to the maximum likelihood estimator, $\mu_s \to \mu_{ML}$, for $s = P, D$, as does the court's estimator, $\hat{\mu} \to \mu_{ML}$; and (ii) the probability of a plaintiff win approaches 50%, $\tilde{\theta} \to 1/2$. The court's estimator converges faster than do the models of the parties.*

For Results 1, 2, and 3, we have kept the court noise parameter constant. For Result 4, we vary λ to show how an improvement in the court's assessment of credibility affects the parties' strategies and the court's decision. First, note that, for $\lambda = 0$, adversarial decision-making is uninformative, as the court sides with the plaintiff half the time (i.e., $\tilde{\theta} = 1/2$) regardless of the parties' claims. This eliminates the credibility cost of an extreme claim, so the parties construct models with the most extreme claims possible: $\mu_D = 0$ and $\mu_P = 1$. As λ increases, the credibility cost of shading their models away from the most likely model increases, so the parties shade less.

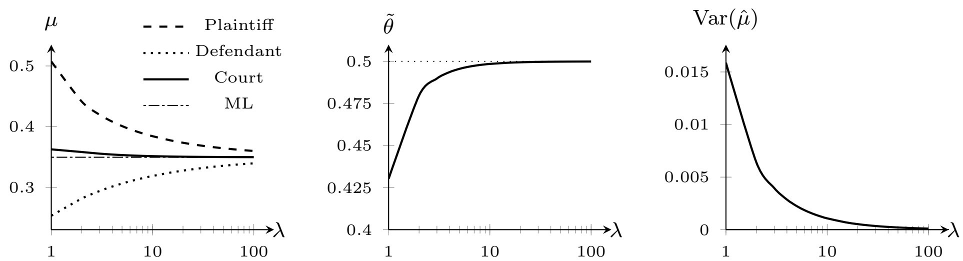
We illustrate the convergence in panel (a) of Figure 3, where we plot the court's equilibrium assessment as a function of court's decision noise parameter. While both the competing models of the adversaries and the court's estimator converge to the maximum likelihood estimate, the court's estimator converges faster than do the competing models of the parties.[20] As a

less favorable for the plaintiff), then there is more "room" to explain the evidence with a larger μ than with a smaller μ. This translates into lower credibility costs for the plaintiff, resulting in a bias in favor of the plaintiff.

[20] Note, however, that this convergence is not the usual convergence in probability, which

Figure 3: Effect of Court Noise

This figure illustrates the relationship between decision-making noise and the equilibrium in the litigation game with evidence $\bar{z} = (1/5, 1/2)$. In panel (a), we plot the means of the models for the plaintiff (dashed curve) and the defendant (dotted curve), the court's equilibrium decision (solid curve), and the mean of the maximum likelihood model (dotted-dashed curve) against the noise parameter λ. In panel (b), we plot the probability of a plaintiff win, $\tilde{\theta}$ against λ. In panel (c), we plot the variance of the court's estimator, $\mathrm{Var}(\hat{\mu}) = \tilde{\theta}(1-\tilde{\theta})\,(\mu_P - \mu_D)^2$ in equation (9) against λ. The horizontal axes are on a logarithmic scale.



(a) Decision (b) Probability $\tilde{\theta}$ (c) Court Variance

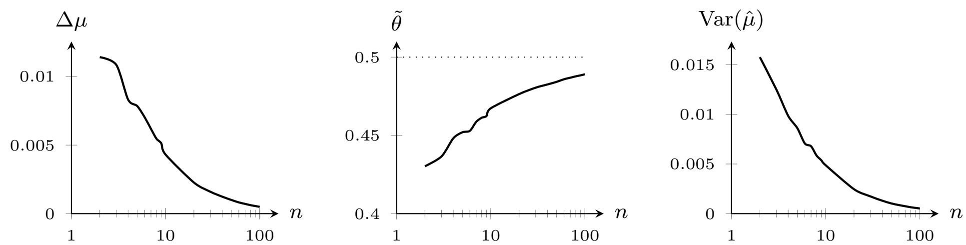
court is better able to assess credibility, the parties choose models closer to the maximum likelihood estimator, but on either side of it. In this sense, the parties' models tend to cancel each other out, which has a salutary effect on the adversarial court's decision-making.

In panel (b) of Figure 3, we plot the probability $\tilde{\theta}$ of a plaintiff win as $\lambda \to \infty$. As court noise disappears, the parties choose models that have the same likelihoods (as their models converge to the maximum likelihood estimator), so the probability of a win approaches $1/2$. This limit corresponds to the 50% probability of a trial win found by Priest and Klein (1984), albeit for a different reason. Their explanation is built around a selection bias story driven by overconfidence (Nalebuff, 1987). In our model, the equilibrium 50%

tells us whether or not an estimator is "consistent." Rather, it is convergence to the best (i.e., most likely) interpretation of the evidence. There is still sampling error because the maximum likelihood estimator still has variance, but the adversarial court will reach the same, most likely explanation as an inquisitorial court.

Figure 4: Effect of More Evidence

This figure illustrates the results for varying evidence sample sizes n. Evidence $z_i \in (0, 1)$ are independent draws from a Beta(1, 2) with $\mu = 1/3$ and $\sigma^2 = 1/18$. For evidence sample sizes $n \in \{2, \ldots, 100\}$, we draw 250 random evidence samples. In panel (a), we plot the sample mean for the bias $\Delta\mu = \hat{\mu} - \mu_{ML}$ against n. In panel (b), we plot the sample mean for plaintiff probability to win $\tilde{\theta}$ against n. In panel (c), we plot the sample mean for variance of the court's decision, $\text{Var}(\hat{\mu}) = \tilde{\theta}(1 - \tilde{\theta})\,(\mu_P - \mu_D)^2$ in equation (9), against n. The horizontal axes are on a logarithmic scale.



(a) Decision Bias (b) Probability $\tilde{\theta}$ (c) Court Variance

win rate is driven by competition between the parties.[21]

In panel (c) of Figure 3, we plot the variance $\text{Var}(\hat{\mu})$ of the court's decision as $\lambda \to \infty$. The first part of the expression for the variance in equation (9) approaches $1/4$ because $\tilde{\theta} \to 1/2$. The second part of the expression approaches zero because both $\mu_P \to \mu_{ML}$ and $\mu_D \to \mu_{ML}$. The reduction in variance will benefit risk averse parties, and potentially reduce the option value of suits (Bebchuk and Klement, 2012).

Result 5. *As the amount of evidence increases, $n \to \infty$, the court's estimator converges in probability to the true $\mu = \alpha/\alpha+\beta$, as do the models of the parties, $\mu_s \to \alpha/\alpha+\beta$ for $s = P, D$.*

As the evidence sample size n increases, the maximum likelihood estimator converges in probability to the true mean $\mu = \alpha/\alpha+\beta$ of the Beta(α, β) process

[21] Note, however, that because, for $\lambda \to \infty$, the parties' claims are the same, convergence of the plaintiff's probability to win when court noise disappears is without practical consequence in our model.

by the law of large numbers. Because the likelihood collapses onto ${}^{\alpha}/_{\alpha+\beta}$, the likelihood penalty (credibility cost) of deviating from ${}^{\alpha}/_{\alpha+\beta}$ increases and the parties shade their models less, which implies that $\mu_s \to {}^{\alpha}/_{\alpha+\beta}$ for $s = P, D$.

We illustrate Result 5 in Figure 4. In panel (a), we plot the decision bias $\Delta\mu = \hat{\mu} - \mu_{ML}$ against the evidence sample size n. The decision bias in this graph is the mean of the decision bias for 250 random evidence samples for each n, drawn from a Beta$(1, 2)$ distribution. Decision bias disappears as both parties' models converge to the mean for the evidence generating process, $\mu = {}^{\alpha}/_{\alpha+\beta} = {}^{1}/_{3}$.

In panel (b) of Figure 4, we plot the probability of a plaintiff win as $n \to \infty$. In this case, the limit corresponds to 50% probability (i.e., $\tilde{\theta} \to {}^{1}/_{2}$). As the likelihood collapses, it becomes symmetric in a neighborhood around the true mean, so the parties choose locations equidistant from, and on either side of the true mean. Symmetry gives these equidistant locations the same likelihood, implying $\tilde{\theta} \to {}^{1}/_{2}$.[22]

In panel (c) of Figure 4, we plot the variance Var$(\hat{\mu})$ of the court's decision as $n \to \infty$. The first part of the expression for the variance in equation (9) approaches ${}^{1}/_{4}$ because $\tilde{\theta} \to {}^{1}/_{2}$ as $n \to \infty$. The second part of the expression for the variance Var$(\hat{\mu})$ approaches zero because both $\mu_s \to {}^{\alpha}/_{\alpha+\beta}$ for $s = P, D$. The reduction in variance associated with better court decisions benefits risk-averse parties and can reduce the number of suits with a negative expected value.[23]

[22]As with the convergence result for the noise parameter λ, because, for $n \to \infty$, the parties' claims are the same, convergence of the plaintiff's probability to win is without practical consequence in our model.

[23]For a thorough discussion of negative-expected-value suits see Bebchuk and Klement

4 Discussion

In this paper, we model adversarial decision-making by turning scientific inquiry upside down. Instead of objective truth seekers who formulate hypotheses and then gather evidence to test them, we study self-interested parties who strategically choose models to influence a decision-maker—after the evidence has already been produced and discovered. Nevertheless, we show that, under certain conditions, the decision-maker (e.g., a court) can still reach the best (most likely) interpretation of the evidence.

We model court decision-making using the metaphor of statistical model selection, where models are proposed by interested parties. This metaphor is rich in that it allows us to identify conditions under which decision-making is likely to be biased away from the best explanation—even when decisions are based on evidence, as for instance in Pfeffer and Sutton (2006). The metaphor also suggests ways to mitigate bias, for example, by reducing court noise, or by increasing the amount of information available.

Our work can be viewed as opening up the "black box" of court decision-making, as in Daughety and Reinganum (2000a), after all the evidence has been produced and discovered. As such, the model captures the trial subgame that can be appended onto games of evidence production or revelation.[24] Whether and how this kind of strategic framing of the evidence would affect the outcome of the larger game is a question for future research.

(2012).

[24] See Gilligan and Krehbiel (1997), Froeb and Kobayashi (1996, 2001, 2012), Daughety and Reinganum (2000b), or Yilankaya (2002) and Milgrom and Roberts (1986), or Shin (1994).

In addition to the implications relating to litigation and arbitration, our model can be applied to the problem of delegating decision rights to subordinates who can end up disagreeing with one another or recommending opposing courses of action. Typically, managers higher up in the hierarchy are responsible for resolving these disagreements. Fama and Jensen (1983) call this the separation of "decision management" by subordinates from "decision control" by a superior. Our results suggest that, even if superiors resolve disagreements by appealing to evidence, the superior's decisions are likely to be noisy and potentially biased if the alternatives are strategically chosen by subordinates. This can be thought of as another kind of agency cost.

References

Ambrus, A., V. Baranovskyi, and A. Kolb (2015): "A Delegation-Based Theory of Expertise," ERID Working Paper 193, Duke University, available at http://ssrn.com/abstract=2643672. [Cited on pages 2 and 11.]

Austen-Smith, D. (1993): "Interested Experts and Policy Advise: Multiple Referrals under Open Rules," *Games and Economic Behavior*, 5, 3–43. [Cited on page 7.]

Baye, M. R., D. Kovenock, and C. G. de Vries (1996): "The All-Pay Auction with Complete Information," *Economic Theory*, 8, 291–305. [Cited on page 12.]

——— (2005): "Comparative Analysis of Litigation System: An Auction-Theoretic Approach," *Economic Journal*, 115, 583–601. [Cited on page 11.]

Bebchuk, L. A. and A. Klement (2012): "Negative-Expected-Value Suits," in *Procedural Law and Economics*, ed. by C. W. Sanchirico, Edward Elgar, vol. 8 of *Encyclopedia of Law and Economics*, chap. 12, 2nd ed. [Cited on pages 20 and 22.]

Cheng, E. K. (2013): "Reconceptualizing the Burden of Proof," *Yale Law Journal*, 122, 1254–1279. [Cited on page 12.]

Choi, S., M. Gulati, and E. A. Posner (2012): "What Do Federal District Judges Want? An Analysis of Publications, Citations, and Reversals," *Journal of Law, Economics, and Organization*, 28, 518–549. [Cited on page 7.]

Chowdhury, S. M. and R. M. Sheremeta (2011): "A Generalized Tullock Contest," *Public Choice*, 147, 413–420. [Cited on page 11.]

Chung, T.-Y. (1996): "Rent-Seeking Contest When the Prize Increases with Aggregate Efforts," *Public Choice*, 87, 55–66. [Cited on page 11.]

Crawford, V. P. (1979): "Compulsory Arbitration Schemes," *Journal of Political Economy*, 87, 131–160. [Cited on page 5.]

Daughety, A. F. and J. F. Reinganum (2000a): "Appealing Judgments," *RAND Journal of Economics*, 31, 502–525. [Cited on pages 7 and 22.]

——— (2000b): "On the Economics of Trials: Adversarial Process, Evidence, and Equilibrium Bias," *Journal of Law, Economics, and Organization*, 16, 365–394. [Cited on pages 2 and 23.]

DeGroot, M. (1970): *Optimal Statistical Decisions*, New York, N.Y.: McGraw-Hill. [Cited on page 9.]

Dewatripont, M. and J. Tirole (1999): "Advocates," *Journal of Political Economy*, 107, 1–39. [Cited on page 2.]

Emons, W. and C. Fluet (2009): "Accuracy Versus Falsification Costs: The Optimal Amount of Evidence under Different Procedures," *Journal of Law, Economics, and Organization*, 25, 134–156. [Cited on page 7.]

Fama, E. F. and M. C. Jensen (1983): "Separation of Ownership and Control," *Journal of Law and Economics*, 26, 301–325. [Cited on page 23.]

Farber, H. S. and M. H. Bazerman (1986): "The General Basis of Arbitrator Behavior: An Empirical Analysis of Conventional and Final-Offer Arbitration," *Econometrica*, 54, 1503–1528. [Cited on pages 7 and 13.]

Fienberg, S. E. and M. L. Straf (1991): "Statistical Evidence in the US Courts: An Appraisal," *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 154, 49–59. [Cited on page 2.]

Froeb, L. M. and B. Kobayashi (1996): "Naïve, Biased yet Bayesian: Can Juries Interpret Selectively Produced Evidence?" *Journal of Law, Economics, and Organization*, 12, 257–276. [Cited on page 23.]

——— (2001): "Evidence Production in Adversarial vs. Inquisitorial Regimes," *Economics Letters*, 70, 267–272. [Cited on pages 2 and 23.]

——— (2012): "Adversarial versus Inquisitorial Justice," in *Procedural Law and Economics*, ed. by C. W. Sanchirico, Edward Elgar, vol. 8 of *Encyclopedia of Law and Economics*, chap. 1, 2nd ed. [Cited on pages 2 and 23.]

Gentzkow, M. and E. Kamenica (2016): "Competition in Persuasion," Unpublished manuscript, Stanford University. [Cited on page 3.]

Gilligan, T. W. and K. Krehbiel (1989): "Asymmetric Information and Legislative Rules with a Heterogeneous Committee," *American Journal of Political Science*, 53, 459–490. [Cited on page 3.]

——— (1997): "Specialization Decisions Within Committee," *Journal of Law, Economics, and Organization*, 13, 366–386. [Cited on page 23.]

Glazer, J. and A. Rubinstein (2001): "Debates and Decisions: On a Rationale of Argumentation Rules," *Games and Economic Behavior*, 36, 158–173. [Cited on page 3.]

Good, I. J. (1979): "Studies in the History of Probability and Statistics. XXXVII

A. M. Turing's statistical work in World War II," *Biometrika*, 66, 393–396. [Cited on page 10.]

Hirshleifer, J. and E. Osborne (2001): "Truth, Effort, and the Legal Battle," *Public Choice*, 108, 169–195. [Cited on page 11.]

Iossa, E. and B. Jullien (2012): "The Market for Lawyers and Quality Layers in Legal Services," *RAND Journal of Economics*, 43, 677–704. [Cited on page 7.]

Iossa, E. and G. Palumbo (2007): "Information Provision and Monitoring of the Decision-Maker in the Presence of an Appeal Process," *Journal of Institutional and Theoretical Economics*, 163, 657–682. [Cited on page 2.]

Jia, H. (2008): "A Stochastic Derivation of the Ratio Form of Contest Success Functions," *Public Choice*, 135, 125–130. [Cited on pages 10 and 32.]

Kaplan, T., I. Luski, A. Sela, and D. Wettstein (2002): "All-Pay Auctions with Variable Rewards," *Journal of Industrial Economics*, 50, 417–430. [Cited on page 11.]

Kartik, N. (2009): "Strategic Communication with Lying Costs," *Review of Economic Studies*, 76, 1359–1395. [Cited on page 7.]

Kartik, N., M. Ottaviani, and F. Squintani (2007): "Credulity, Lies, and Costly Talk," *Journal of Economic Theory*, 134, 93–116. [Cited on page 7.]

Katz, A. W. (1988): "Judicial Decisionmaking and Litigation Expenditure," *International Review of Law and Economics*, 8, 127–143. [Cited on page 11.]

Konrad, K. A. (2009): *Strategy and Dynamics in Contests*, Oxford, UK: Oxford University Press. [Cited on pages 15 and 40.]

Konrad, K. A. and H. Schlesinger (1997): "The Risk-Aversion in Rent-Seeking and Rent-Augmenting Games," *Economic Journal*, 107, 1671–1683. [Cited on page 11.]

Krishna, V. and J. Morgan (2001a): "Asymmetric Information and Legislative Rules: Some Amendments," *American Political Science Review*, 95, 435–452. [Cited on page 3.]

——— (2001b): "A Model of Expertise," *Quarterly Journal of Economics*, 116, 747–775. [Cited on page 3.]

Leonard, T. and J. S. J. Hsu (1999): *Bayesian Methods*, Cambridge, UK: Cambridge University Press. [Cited on page 38.]

Mauet, T. A. (2007): *Trial Techniques*, New York, N.Y.: Wolters Kluwer Law & Business, 7th ed. [Cited on page 3.]

McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka, N.Y.: Academic Press, 105–142. [Cited on page 10.]

McKelvey, R. and T. Palfrey (1995): "Quantal Response Equilibria for Normal Form Games," *Games and Economic Behavior*, 10, 6–38. [Cited on page 10.]

Milgrom, P. and J. Roberts (1986): "Relying on the Information of Interested Parties," *RAND Journal of Economics*, 17, 18–32. [Cited on pages 2 and 23.]

Mueller, S. T. and C. T. Weidemann (2008): "Decision Noise: An Explanation for Observed Violations of Signal Detection Theory," *Psychonomic Bulletin & Review*, 15, 465–494. [Cited on page 10.]

Nalebuff, B. (1987): "Credible Pretrial Negotiation," *RAND Journal of Economics*, 18, 198–210. [Cited on page 20.]

Pardo, M. S. and R. J. Allen (2008): "Juridical Proof and the Best Explanation," *Law and Philosophy*, 27, 223–268. [Cited on page 7.]

Pfeffer, J. and R. I. Sutton (2006): "Evidence-Based Management," *Harvard Business Review*, 84, 62–74. [Cited on page 22.]

Priest, G. L. and B. Klein (1984): "The Selection of Disputes for Litigation," *Journal of Legal Studies*, 13, 1–56. [Cited on page 20.]

Rantakari, H. (2016): "Soliciting Advice: Active versus Passive Principals," *Journal of Law, Economics, and Organization*. [Cited on pages 2 and 3.]

Saks, M. J. and S. L. Neufeld (2011): "Convergent Evolution in Law and Science: The Structure of Decision-Making Under Uncertainty," *Law, Probability and Risk*, 10, 133–148. [Cited on page 12.]

Shin, H. S. (1994): "The Burden of Proof in a Game of Persuasion," *Journal of Economic Theory*, 64, 253–264. [Cited on page 23.]

——— (1998): "Adversarial and Inquisitorial Procedures in Arbitration," *RAND Journal of Economics*, 29, 378–405. [Cited on page 2.]

Skaperdas, S. (1996): "Contest Success Functions," *Economic Theory*, 7, 283–290. [Cited on page 11.]

Skaperdas, S. and S. Vaidya (2012): "Persuasion as a Contest," *Economic Theory*, 51, 465–486. [Cited on pages 2 and 11.]

Tanford, J. A. (2009): *The Trial Process: Law, Tactics and Ethics*, Newark, N.J.: LexisNexis, 4th ed. [Cited on page 3.]

Tenn, S., L. M. Froeb, and S. Tschantz (2010): "Merger Effects When Firms Compete by Choosing Both Price and Promotion," *International Journal of Industrial Organization*, 28, 695–707. [Cited on page 2.]

Train, K. (2009): *Discrete Choice Methods with Simulation*, Cambridge, UK: Cambridge University Press, 2nd ed. [Cited on page 10.]

Tullock, G. (1980): "Efficient Rent Seeking," in *Toward a Theory of the Rent-Seeking Society*, ed. by J. M. Buchanan, R. D. Tollison, and G. Tullock, College Station, TX: Texas A&M University Press, 97–112. [Cited on page 11.]

Wagner, W. (2005): "The Perils of Relying on Interested Parties to Evaluate Scientific Quality," *American Journal of Public Health*, 95, s99–s106. [Cited on page 2.]

Werden, G. J. and L. M. Froeb (1994): "The Effects of Mergers in Differentiated Products Industries: Logit Demand and Merger Policy," *Journal of Law, Economics, and Organization*, 10, 407–426. [Cited on page 2.]

Wittman, D. (1986): "Final-Offer Arbitration," *Management Science*, 32, 1551–1561. [Cited on page 7.]

Yilankaya, O. (2002): "A Model of Evidence Production and Optimal Standard of Proof and Penalty in Criminal Trials," *Canadian Journal of Economics/Revue canadienne d'économique*, 35, 385–409. [Cited on page 23.]

Technical Appendix

A The General Persuasion Game

A.1 Introduction

The results presented in the main text obviously depend on the specific distribution chosen. In this appendix, we generalize the game to any arbitrary distribution. We use the generalized game to identify the properties of a distribution (locations and likelihood) that give rise to our results.

A.2 Problem

We consider an unobservable evidence-generating process that is characterized by its theoretical mean. A principal is charged with making an assessment about the type of this unknown process. We assume that the principal does not have the capability or capacity to make her own assessment of the type. Instead, she solicits advice from agents with vested and opposing interests. The principal's objective is to make the best possible assessment of the type of the process. She therefore follows the advice of the agent who is most credible, given a publicly observable sample drawn from the unknown process. We assume that the principal's assessment of an agent's advice is noisy so that her decision comes with error.

A.3 Notation

We refer to the unknown process by its theoretical mean as type $y \in \mathbb{R}$. A principal is charged with making an assessment $\hat{y} \in \mathbb{R}$ of the unknown type of the process. We denote by $\hat{y}$ the principal's decision in this game of persuasion. The principal's objective is to make the best assessment given an available (and publicly observable) sample of evidence drawn according to the unknown process. We refer to the objectively best assessment as $\bar{y}$. By assumption, the principal does not have access to this assessment but rather solicits advice from outside experts.

The principal solicits advice from two agents, $i = L, R$. Each agent's advice is modeled as an interpretation that characterizes the sample as coming from a process of type y_i with credibility $\chi_i \geq 0$. The principal assesses the agents' advice and chooses the most credible of the two. We assume this assessment of credibility is noisy and refer to it as $\widetilde{\chi}_i = \chi_i \exp \xi_i$ for $i = L, R$, where ξ_i are independently extreme value (or Gumbel) distributed with mean 0 and scale $1/\lambda$.[25] The principal therefore follows agent R's advice if $\widetilde{\chi}_R > \widetilde{\chi}_L$ and agent L's otherwise. If $\widetilde{\chi}_L = \widetilde{\chi}_R$, then the principal flips a fair coin. This is akin to the structure of the logit choice model. The principal sides with agent R with probability

$$\tilde{\theta} = \Pr(\widetilde{\chi}_R > \widetilde{\chi}_L) = \frac{\exp(\lambda \log \chi_R)}{\exp(\lambda \log \chi_R) + \exp(\lambda \log \chi_L)} = \frac{\chi_R^\lambda}{\chi_R^\lambda + \chi_L^\lambda}. \quad \text{(A1)}$$

[25] The structure in Jia (2008) is less restrictive, requiring the random variable ξ_i to belong to the *inverse exponential distribution*.

We define the *incredibility* of agent R's advice as

$$x_R = \frac{1}{\chi_R^\lambda} \tag{A2}$$

and of agent L's advice as

$$x_L = -\frac{1}{\chi_L^\lambda}. \tag{A3}$$

An agent's advice strategy can thus be represented by a pair $a_i = (x_i, y_i) \in A_i$ with a proposed type $y_i \in \mathbb{R}$ and an incredibility of that advice of $x_L \in \mathbb{R}^-$ for agent L and $x_L \in \mathbb{R}^+$ for agent R. Because we measure agent L's incredibility with a negative number, in (x, y)-space, agent L's strategy space A_L is to the left of the y-axis, whereas agent R's strategy space A_R is to the right of the y-axis. Advice located further from the y-axis is less credible (i.e., more incredible).

We further limit the agent's strategy space to be a compact and convex subset of $\mathbb{R}^2$ so that $A_L \subset \mathbb{R}^- \times \mathbb{R}$ and $A_R \subset \mathbb{R}^+ \times \mathbb{R}$. We assume the set of feasible strategies is characterized by a type-credibility tradeoff. In other words, the further advice y_i is from the objectively best assessment $\bar{y}$, the less credible this advice will be with a value of χ_i, or, alternatively, the more incredible the advice will be with a higher value of $|x_i|$. Extreme advice with very high (or low) type y_i and low incredibility $|x_i|$ is therefore not feasible, and the strategy space is convex.

Using the expressions for agent's incredibility, the probability that the principal follows agent R's advice in equation (A1) can be rewritten as

$$\tilde{\theta}(x_L, x_R) = \frac{x_L}{x_R - x_L}. \tag{A4}$$

The principal's assessment of the process type is y_R when she follows agent R's advice and y_L when she follows L's advice. In expectations, the principal's assessment[26] and decision is thus

$$\begin{aligned} \hat{y}(a_L, a_R) &= \tilde{\theta}(x_L, x_R) y_R + \big(1 - \tilde{\theta}(x_L, x_R)\big) y_L \\ &= \frac{x_R y_L - x_L y_R}{x_R - x_L}. \end{aligned} \tag{A5}$$

It is the credibility-weighted sum of the agent's location advice.

We can further rewrite the expression in equation (A5) as

$$\hat{y}(a_L, a_R) = y_L - m(a_L, a_R) x_L = y_R - m(a_L, a_R) x_R \tag{A6}$$

where

$$m(a_L, a_R) = \frac{y_R - y_L}{x_R - x_L} \tag{A7}$$

is the slope of the line connecting the two points $a_L = (x_L, y_L)$ and $a_R = (x_R, y_R)$ in (x, y)-space.

The two agents have vested and opposing interests. We assume that the

[26] This expected assessment $\hat{y}$ is also the outcome of a decision-maker who minimizes a quadratic loss function $-w_R\left(y_R - \hat{y}\right)^2 - w_L\left(y_L - \hat{y}\right)^2$, that is, the weighted sum of the squared deviations of assessment $\hat{y}$ from the agent's proposed types y_i.

agents' payoffs are directly affected by the principal's assessment of type. Agent L prefers low values of $\hat{y}$, whereas agent R prefers high values. For given $y_R > y_L$, the expression for the principal's expected decision in equation (A5) implies that both agents will choose the most credible interpretations given their advice types y_i. For agent L, this means the highest possible $x_L \in \mathbb{R}^-$; and for agent R the lowest possible $x_R \in \mathbb{R}^+$. We define these "incredibility frontiers" as

$$\hat{x}_L(y_L, \cdot) = \max\{x : (x, y_L) \in A_L\} \quad \text{(A8)}$$

and

$$\hat{x}_R(y_R, \cdot) = \min\{x : (x, y_R) \in A_R\} \quad \text{(A9)}$$

where $a_L = (\hat{x}_L(y), y)$ dominates any other strategy for agent L with a given y value, and similarly for $a_R = (\hat{x}_R(y), y)$. These incredibility frontiers are the hulls of A_i facing the y-axis in (x, y)-space.

An incredibility frontier $\hat{x}_i(y_i, \cdot)$ depends on the agent's advice type y_i as well as environmental characteristics (e.g., evidence sample, a potential prior bias by the principal, the noise parameter λ, or the expertise of the agent) captured by the properties of the agent's strategy space A_i. This strategy space A_i and thus the agent's incredibility frontier does not depend on the other agent's strategy.

A.4 Equilibrium Concept

A *persuasion game* is a simultaneous-move, non-cooperative game between two agents $i = L, R$ providing strategic advice $a_i \in A_i$ to maximize payoffs $\pi_L = -\hat{y}(a_L, a_R)$ for agent L and $\pi_R = \hat{y}(a_L, a_R)$ for agent R with $\hat{y}(a_L, a_R)$ defined in equation (A6). A Nash equilibrium in this game is a strategy profile (a_L^*, a_R^*) such that

$$\left.\begin{aligned} \hat{y}(a_L^*, a_R^*) \leq \hat{y}(a_L, a_R^*) \qquad \forall a_L \in A_L \text{ for agent } L \\ \hat{y}(a_L^*, a_R^*) \geq \hat{y}(a_L^*, a_R) \qquad \forall a_R \in A_R \text{ for agent } R \end{aligned}\right\}. \tag{A10}$$

From the expression for the principal's decision in equation (A6), we can conclude that, because agent L's incredibility is by definition negative, $x_L < 0$, if $m(a_L', a_R') > m(a_L, a_R)$, then either $m(a_L', a_R) > m(a_L, a_R)$ or $m(a_L, a_R') > m(a_L, a_R)$. In other words, if a strategy profile (a_L, a_R) does not result in a maximum for m, at least one of the agents can unilaterally move to increase the slope.

Lemma A1. *Both agents present advice a_i to maximize the slope $m(a_L, a_R)$.*

An immediate implication of Lemma A1 is that, if it exists, a Nash equilibrium (a_L^*, a_R^*) in this game determines a line of maximum slope $m(a_L^*, a_R^*)$.

A.5 Equilibrium Results

In the sequel, we present our main results from the general persuasion game and relate them back to the model presented in the main text of the paper.

A.5.1 Nash Equilibrium

By Lemma A1, in equilibrium, the advice strategy profile $(a_L^*, a_R^*) \in A_L \times A_R$ will be such that slope $m(a_L, a_R)$ is maximized. As A_L is all on or above the line with slope m connecting a_L and a_R, and A_R is all on or below that line, it follows that there is a unique line with this maximum slope, m^*. Agents L and R can choose any points along this line in A_L and A_R, or any mixed strategies between such points (as a mixture of pure strategies), but the value of the game $\hat{y}^* \equiv \hat{y}(a_L^*, a_R^*)$ is the y-intercept of the line of maximum slope between the choice sets. We summarize these results in Theorem A1.

Theorem A1. *A pure strategy Nash equilibrium of the persuasion game will exist if, and only if, the slope function $m(a_L, a_R)$ has a maximum value on $A_L \times A_R$, that is, when there is a unique common line of support below A_L and above A_R. If this line meets A_L or A_R in more than one point, then there are also mixed strategy equilibria that are mixtures of pure strategies along this line, and result in the same assessment $\hat{y}$ for the game.*

Two properties of this result are worth mentioning. First, if the projections of A_L and A_R onto the y-axis are bounded, then there is a maximum slope line. More generally, if a line of positive slope cuts off a bounded region of A_L below the line and a bounded region of A_R above the line, then there is a maximum slope line. This is true if the incredibility grows faster than linearly for large positive and large negative values.

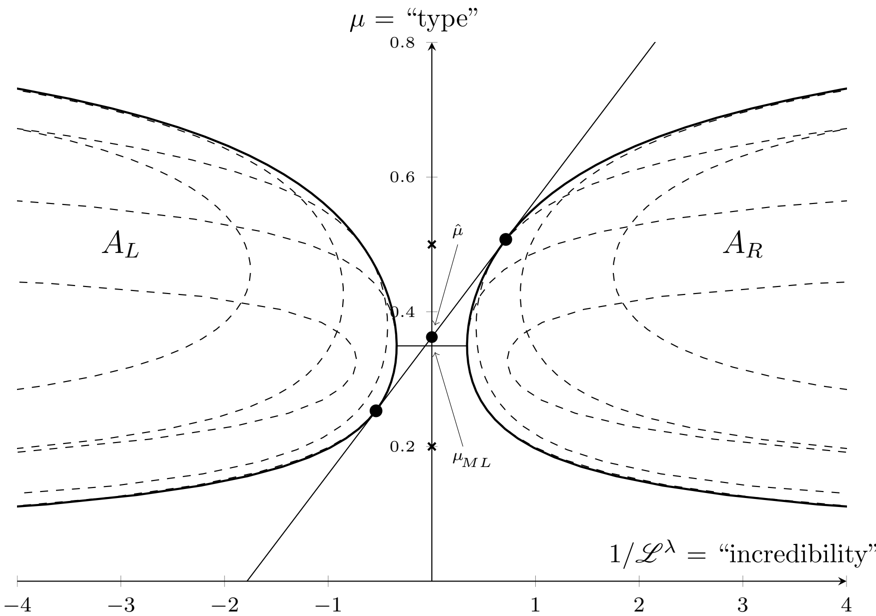
Second, if $\hat{x}_L$ and $\hat{x}_R$ are strictly concave, differentiable functions, defined on a convex subset of the real line, with unbounded derivatives, then these functions have unique maxima and minima, respectively, and define the relevant frontiers of the strategy sets. These assumptions also guarantee the existence of a unique Nash equilibrium solution $a_L^* = (\hat{x}_L(y_L^*), y_L^*)$ and $a_R^* = (\hat{x}_R(y_R^*), y_R^*)$, and the line through these points is simultaneously tangent to both the L and R curves.

In Figure A1, we relate the general game to our litigation game in the main test by using the specific parameterization of the game in the main text. The unobservable type is the theoretical mean of the Beta(α, β) distribution, $y = \mu$, and the inverse credibility is the reciprocal likelihood or "incredibility," $x = 1/\mathscr{L}^\lambda$. Agent L is the defendant D (preferring low-valued outcomes) and agent R is the plaintiff (preferring high-valued outcomes), where A_L is the set $\left(-1/\mathscr{L}_D^\lambda, \mu_D\right)$ and A_R is the set $\left(1/\mathscr{L}_P^\lambda, \mu_P\right)$, both defined over all possible Beta(α, β) distribution functions. This means, there are multiple parameterizations to obtain a fixed $\mu = \alpha/\left(\alpha + \beta\right)$ and varying $\sigma^2 = \mu\left(1 - \mu\right)/\left(1 + \alpha + \beta\right)$. Alternatively, there are multiple parameterizations (and thus likelihoods) to obtain a fixed σ^2 and varying μ (Leonard and Hsu, 1999).

With this set up, the x-axis in Figure A1 measures incredibility $1/\mathscr{L}^\lambda$ as a function of the type μ plotted on the y-axis. The dashed lines represent the reciprocal likelihoods of various types, for various fixed values of variance σ^2. This gives a family of overlapping curves, the envelope of which is also drawn, and whose union defines the A_R set to the right, and which is mirrored in the A_L set to the left. The line of maximum slope is drawn between the points in these sets, defining the optimal advice strategies for the two sides. The

Figure A1: Advice in the Simple Litigation Game

In Figure A1, we illustrate the geometrical characterization of the agents' equilibrium strategies. Choice set A_L for agent L is to the left of the vertical axis; choice set A_R for agent R is to the right of the axis. Each dashed line represents the reciprocal likelihood for varying proposed type y, holding the variance fixed. The solid curve represents the envelope of the family of these overlapping curves. The bullet point on the vertical axis represents the equilibrium decision $\hat{y}$. The bullet points on the envelopes of A_L and A_R represent the agents' advice a_i^*. The most credible type $\bar{y}$ is marked by the horizontal line between the peaks of the envelopes of A_L and A_R.



y-intercept of the line is denoted by a dot on the vertical axis. It represents the equilibrium assessment $\hat{y}^*$ of the game. This assessment is slightly above the maximum likelihood (i.e., minimum incredibility) value $\bar{y}$, marked by a horizontal line between the "peaks" of the two sets, A_L and A_R.

A.5.2 Payoff Shading

We have denoted the objectively best assessment of the type as $\bar{y}$. Suppose that this type $\bar{y}$ is also the most credible advice the agents can give. That means, the maximum of $\hat{x}_L < 0$ and the minimum of $\hat{x}_R > 0$ (i.e., the points where these come closest to the y-axis) are at the same $\bar{y}$. This then implies that that the strategy $(\hat{x}_L(y_L), y_L)$ for L with $y_L > \bar{y}$ is dominated by $(\hat{x}_L(\bar{y}), \bar{y})$. Similarly, a strategy $(\hat{x}_R(y_R), y_R)$ for R with $y_R < \bar{y}$ is dominated by $(\hat{x}_R(\bar{y}), \bar{y})$. Because the incredibility functions $\hat{x}_i$ cannot be differentiable and have a corner at $\bar{y}$, agents will "shade" their advice, with L offering a type y_L^* less than the most likely $\bar{y}$, and R offering a type y_R^* greater than this $\bar{y}$.

Theorem A2. *In equilibrium, the agents shade and present advice a_i^* with types y_i^* on either side of the most credible type $\bar{y}$. The Nash equilibrium advice strategies with proposed types y_L^* and y_R^* satisfy $y_L^* < \bar{y} < y_R^*$.*

The result in Theorem A2 is analogous to Result 1 in the main text. The agents shade their advice in their favor. Moreover, if the incredibility functions $\hat{x}_i$ are strictly concave with $|\hat{x}_i(y)| > |\hat{x}_i(\bar{y})|$ increasing in $|y - \bar{y}|$, then the equilibrium types presented by the agents are finite, $y_L^* > -\infty$ and $y_R^* < \infty$. The agents therefore engage in *payoff moderation* (Konrad, 2009).

A.5.3 Bias

If the shape of the incredibility function is not symmetric about the most credible $\bar{y}$, but instead favors one side over the other with less incredibility for equal offsets from $\bar{y}$, then the equilibrium assessment will be *biased* from $\bar{y}$ in the direction of that side. In other words, $|\hat{y}^* - \bar{y}| > 0$. We illustrate this in Figure A1 where the likelihood function for the litigation game example decreases more slowly for Beta(α, β) distributions having μ greater than the maximum likelihood estimate ($\bar{y} = \mu_{ML}$) than it does for μ less than this value. Heuristically, if the evidence is closer to the lower range of the Beta(α, β)

distribution, then there is more "room" to explain the evidence with a larger μ than with a smaller μ.

It may be that the principal holds a *biased prior* or that there are differences in the capabilities of the agents such that one side offering the theory with type $\bar{y}$ would be viewed more favorably than the other offering what should amount to the same most credible theory. We set aside this sort of asymmetry between the sides and assume:

$$\hat{x}_L(\bar{y}) = -\hat{x}_R(\bar{y}). \tag{A11}$$

This assumption means that either player can offer up this best theory with the same resulting weight. It implies that the identity of the agent does not matter

Because, by Theorem A2, agent L shades down, $y_L < \bar{y}$, and agent R shades up, $y_R > \bar{y}$, values of $\hat{x}_L$ for $y_L > \bar{y}$ and values of $\hat{x}_R$ for $y_R < \bar{y}$ are observed only off equilibrium. For the properties of the equilibrium decision $\hat{y}^*$ we can therefore ignore these values. This means that we may as well take a single function $\hat{x}$ describing both parties' incredibility functions: $\hat{x}(y) = -\hat{x}_L(y)$ for $y \leq \bar{y}$ and $\hat{x}(y) = \hat{x}_R(y)$ for $y \geq \bar{y}$. The bias of the principal's decision relative to $\bar{y}$ is then determined by how quickly the incredibility increases for $y > \bar{y}$ as compared to $y < \bar{y}$ as a function of the difference from the most credible type $\bar{y}$. In Theorem A3 below, we make use of the following definitions:

Definition A1 (Symmetry)**.** *The incredibility function $\hat{x}(y)$ is symmetric about $y = \bar{y}$ if, for every $\delta > 0$, $\hat{x}(\bar{y} - \delta) = \hat{x}(\bar{y} + \delta)$.*

Definition A2 (Credibility Costs)**.** *Agent L has lower credibility costs in $\hat{x}$ (and agent R has higher credibility costs) if, for every $\delta > 0$, $\hat{x}(\bar{y}-\delta) < \hat{x}(\bar{y}+\delta)$; that is, advice a_L with type shaded down by δ is more credible than advice a_R with type shaded up by an equal amount δ. Analogously for agent R.*

Definition A3 (Monotonic Credibility Costs)**.** *Agent L has monotonically lower credibility costs (and agent R has monotonically higher credibility costs) if $\hat{x}(\bar{y}+\delta) - \hat{x}(\bar{y}-\delta)$ is a strictly increasing function for $\delta > 0$. Analogously for agent R.*

Theorem A3. *For the general persuasion game with equilibrium strategies $a_L^* = (-\hat{x}(y_L^*), y_L^*)$ and $a_R^* = (\hat{x}(y_R^*), y_R^*)$ and equilibrium assessment $\hat{y}^* = \hat{y}(a_L^*, a_R^*)$, the following bias properties hold:*

1. *If $\hat{x}(y)$ is symmetric, then $y_R^* - \bar{y} = \bar{y} - y_L^*$ and $\hat{y}^* = \bar{y}$.*

2. *If agent L has lower credibility costs, then $\hat{y}^* < \bar{y}$, and the equilibrium assessment is biased down. If agent R has lower credibility costs, then $\hat{y}^* > \bar{y}$, and the equilibrium assessment is biased up.*

3. *If agent L has monotonically lower credibility costs, then agent L's advice a_L^* exhibits more shading than agent R's advice, $\bar{y} - y_L^* > y_R^* - \bar{y}$. Analogously for agent R.*

Proof. 1. Suppose $\hat{x}(y)$ is symmetric (Definition A1). If $y_R^* = \bar{y} + \delta$, then for $y'_L = \bar{y} - \delta$ and $a'_L = (-\hat{x}(y'_L), y'_L)$, $\hat{x}(y'_L) = \hat{x}(y_R^*)$ so that $\hat{y}^* \leq \hat{y}(a'_L, a_R^*) = \bar{y}$ since L can do no worse than respond to a_R^* with strategy a'_L. Similarly if $y_L^* = \bar{y} - \delta$, taking $y'_R = \bar{y} + \delta$ shows $\hat{y}^* \geq \bar{y}$. Hence $\hat{y}^* = \bar{y}$, and the same $\delta = y_R^* - \bar{y} = \bar{y} - y_L^*$.

2. With lower credibility costs (Definition A2) for agent L, $\hat{x}(\bar{y}-\delta) < \hat{x}(\bar{y}+\delta)$ for all $\delta > 0$. If $y_R^* = \bar{y}+\delta$, then take $y_L' = \bar{y}-\delta$ and $a_L' = (-\hat{x}(y_L'), y_L')$. Because $\hat{x}(\bar{y}-\delta) < \hat{x}(\bar{y}+\delta)$, $\hat{y}^* \leq \hat{y}(a_L', a_R^*) < \bar{y}$. Analogously for agent R.

3. With monotonically lower credibility costs (Definition A3) for agent L, $\hat{x}(\bar{y}+\delta) - \hat{x}(\bar{y}-\delta)$ is strictly increasing. Then, for $\delta = y_R^* - \bar{y}$, the derivative $-\hat{x}'(\bar{y}-\delta) < \hat{x}'(\bar{y}+\delta) = \hat{x}'(y_R^*) = -\hat{x}'(y_L^*)$ because the maximum slope line is tangent to both incredibility curves at the equilibrium solution. But $\hat{x}'(y)$ is strictly increasing so $y_L^* < \bar{y}-\delta$, that is, $\delta = y_R^* - \bar{y} < \bar{y} - y_L^*$. The analogous arguments hold when R has lower credibility costs. Q.E.D.

A.5.4 Convergence as $n \to \infty$

The illustration in Figure A1 is based on an evidence sample with only two values: $\bar{z} = (1/5, 1/2)$. In other words, there is not a lot of evidence constraining the agents' advice. With more evidence, the likelihood function has a narrower peak, so advice away from the maximum likelihood become much less credible. In general, as the sample size n increases, we expect the credibility function $\hat{x}$ to collapse on $\bar{y}$ for the true process generating the evidence.

More specifically, suppose a family of incredibility functions denoted by $\hat{x}(y|n)$ are parameterized by a variable n denoting the amount of evidence available. Suppose that the most credible $\bar{y}$ is the same for all incredibility functions $\hat{x}(y|n)$. Scaling the incredibility by a constant factor does nothing to change the outcome of the game. We thus assume that these functions are all normalized to one, $\hat{x}(\bar{y}|n) = 1$. The notion of narrowing incredibility functions is then captured formally as a hypothesis of the following consistency result.

Theorem A4. *Let the equilibrium assessment in the persuasion game with incredibility function $\hat{x}(y|n)$ be denoted by $\hat{y}_n^*$. Suppose that for every $\epsilon > 0$, for all sufficiently large n, and any y we have $\hat{x}(y|n) > |y - \bar{y}|/\epsilon$. Then $\lim_{n\to\infty} \hat{y}_n^* = \bar{y}$.*

Proof. Suppose $\epsilon > 0$ is given and take N so for all $n \geq N$ and any y we have $\hat{x}(y|n) > |y - \bar{y}|/\epsilon$. Let $a_L^* = (\hat{x}(y_L^*|n), y_L^*)$ and $a_R^* = (\hat{x}(y_R^*|n), y_R^*)$ be equilibrium strategies for the persuasion game with $\hat{x}(y|n)$. Let $a'_L = (-1, \bar{y})$ be the maximally credible strategy for agent L. Then

$$\hat{y}_n^* = \hat{y}(a_L^*, a_R^*) \leq \hat{y}(a'_L, a_R^*) = \frac{\hat{x}(\hat{y}^*|n)\bar{y} + \hat{y}^*}{\hat{x}(\hat{y}^*|n) + 1} < \bar{y} + \frac{\hat{y}^* - \bar{y}}{\hat{x}(\hat{y}^*|n)} < \bar{y} + \epsilon.$$

On the other hand, taking $a'_R = (1, \bar{y})$ shows $\hat{y}_n^* \geq \hat{y}(a_L^*, a'_R) > \bar{y} - \epsilon$ in similar fashion. Hence, for every $\epsilon > 0$, for all sufficiently large n, $|\hat{y}_n^* - \bar{y}| < \epsilon$, that is, $\lim_{n\to\infty} \hat{y}_n^* = \bar{y}$. Q.E.D.

This result is stronger than what we illustrate with Result 5 in the main text where we show that the bias decreases with more evidence. In Theorem A4, we show that the equilibrium assessment converges to the most credible assessment $\bar{y}$. In other words, any bias in assessments away from the most credible $\bar{y}$ due to the adversarial process disappears with increasing evidence. Advice that deviates from the most credible explanation simply faces an increasing credibility penalty the more evidence there is. The argument gives a bound for the deviation of $\hat{y}_n^*$ from $\bar{y}$, but the argument cannot tell us that this bias decreases monotonically with n without much more detailed assumptions about the dependence of $\hat{x}(y|n)$ on n.