

MPRA

Munich Personal RePEc Archive

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

Luke M. Froeb and Bernhard Ganglmair and Steven
Tschantz

Vanderbilt University, University of Texas at Dallas, Vanderbilt
University

28 March 2016

Online at <https://mpra.ub.uni-muenchen.de/71501/>
MPRA Paper No. 71501, posted 1 June 2016 04:42 UTC

24 1 Introduction

25 Adversarial justice has two stages: (i) information provision, where informa-
26 tion is acquired and reported to the court; and (ii) decision-making, where
27 a court makes a decision to resolve the dispute (Iossa and Palumbo, 2007).
28 Much of the economic literature comparing adversarial to inquisitorial justice
29 focuses on the first stage, where evidence is produced and reported by adver-
30 saries, rather than an impartial third party (Milgrom and Roberts, 1986; Shin,
31 1998; Dewatripont and Tirole, 1999; Froeb and Kobayashi, 2001; Daughety and
32 Reinganum, 2000b; Skaperdas and Vaidya, 2012; Froeb and Kobayashi, 2012;
33 Rantakari, 2016). Although these articles address different aspects of the ad-
34 versarial system, they all find that competition between the adversaries plays
35 a crucial role in the ability of a court to gather information.

36 However, the adversarial system also differs from the inquisitorial in the
37 second, decision-making stage: instead of having to choose between competing
38 interpretations of evidence constructed by interested parties (“adversarial”), a
39 court can instead appoint a neutral expert to interpret the evidence for them
40 (“inquisitorial”). A shift to the second option is probably the most commonly
41 called-for reform of the (adversarial) justice system in the United States (Fien-
42 berg and Straf, 1991; Froeb and Kobayashi, 2001; Wagner, 2005), especially
43 for scientific or statistical evidence where the court often lacks “knowledge
44 and expertise . . . and therefore has to delegate the job to a qualified expert”
45 (Ambrus et al., 2015). For example, antitrust merger trials often involve oppos-
46 ing expert economists who construct oligopoly models to predict post-merger
47 prices (Werden and Froeb, 1994; Tenn et al., 2010). Lay competition tribunals

48 are called on to assess the relative credibility of the two models, even though
49 constructing a model would be beyond the tribunal’s capability.

50 More generally, think about two litigants preparing for trial. Evidence has
51 already been produced and discovered, and the opposing attorneys are devising
52 strategies to win in court. As first-year law students are taught (Tanford,
53 2009):

54 It is your job to sort the information before trial, organize it, sim-
55 plify it and present it to the jury in a *simple model that explains*
56 *what happened and why you are entitled to a favorable verdict.*

57 Remember that there is a lawyer on the other side who will be
58 trying to sell the jury *a story that contradicts yours*. . . . If both
59 sides do competent jobs, the *jury will have to choose between two*
60 *competing versions of events* . . . [emphasis added]

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

¹Our 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).

²In 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.

66 Deciding between the competing interpretations of the evidence is analo-
67 gous to the statistical problem of “model selection,” where a researcher ob-
68 serves data, and then tries to determine which of two models generated it. If
69 we think of the court as a researcher and the evidence as data, then the adver-
70 saries’ models can be parameterized as probability distributions that describe
71 the evidence-generating process. The court measures the “credibility” of each
72 model by its statistical likelihood and decides in favor of the most likely model.

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

82 If the court can, through the process of discovery and rebuttal, perfectly
83 rank the likelihoods of the two competing models, then competition between
84 the adversaries forces each party to propose the same, most likely model. This
85 is because any “shading” (i.e., interpretation of the evidence away from the
86 most likely model) can be bested by a more likely model with a higher expected
87 payoff. This leads to an equilibrium in which each party chooses the same,
88 most likely model. In other words, competition between the adversaries leads
89 the court to the best interpretation of the evidence, even though it has to

90 choose between interpretations constructed by interested parties. A similar
91 result comes from final-offer arbitration, where parties make the same utility
92 maximizing offer (Crawford, 1979).

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

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

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

110 2 Litigation as a Persuasion Game

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

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

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

121 In this litigation game, both parties simultaneously offer competing models,
 122 $Z_P \in \mathcal{F}$ and $Z_D \in \mathcal{F}$, to explain the evidence and the respective level of
 123 damages. The court (as the decision-maker in this contest) evaluates the
 124 model likelihoods $\mathcal{L}_P = \Pr(Z_P|\bar{z})$ and $\mathcal{L}_D = \Pr(Z_D|\bar{z})$ and awards damages
 125 of $\mu_P \geq 0$ or $\mu_D \geq 0$. The expected payment from the defendant to the plaintiff
 126 is

$$\hat{\mu} = \theta \mu_P + (1 - \theta) \mu_D \quad (2)$$

³The 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.

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

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

⁴This 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.

⁵We 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.

⁶Viewed 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 assume in most cases the parties have put forward the explanation that best helps their case.”

⁷The 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

139 **2.1 A Perfect Court as an Inquisitorial Benchmark**

140 For our benchmark, we suppose the court can perfectly (i.e., without noise)
 141 rank-order the likelihoods of the parties' proposed models. The probability of
 142 a plaintiff win in this case⁸ is

$$\theta = \begin{cases} 1 & \text{for } \mathcal{L}_P > \mathcal{L}_D \\ 1/2 & \text{for } \mathcal{L}_P = \mathcal{L}_D \\ 0 & \text{for } \mathcal{L}_P < \mathcal{L}_D, \end{cases} \quad (3)$$

143 and the expected payment from the defendant to the plaintiff is

$$\hat{\mu} = \begin{cases} \mu_P & \text{for } \mathcal{L}_P > \mathcal{L}_D \\ 1/2 \mu_P + 1/2 \mu_D & \text{for } \mathcal{L}_P = \mathcal{L}_D \\ \mu_D & \text{for } \mathcal{L}_P < \mathcal{L}_D. \end{cases} \quad (4)$$

144 With these payoffs, in equilibrium, the adversaries both choose the same,
 145 most likely model: $Z_P = Z_D = Z_{ML}$. To see this, note that any model
 146 proposed by the plaintiff with damages higher (and a likelihood lower) than
 147 the most likely model Z_{ML} gives the defendant an opportunity to choose a more
 148 likely model (i.e., "closer" to Z_{ML}) with lower damages and thus a higher payoff
 149 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.

⁸We 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.

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

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

153 This outcome of adversarial litigation with a perfect court is also the max-
154 imum likelihood estimator whose well-known optimality properties (DeGroot,
155 1970) make it an obvious choice for an *inquisitorial* court with the ability to
156 interpret the evidence itself. The maximum likelihood estimator serves as the
157 benchmark against which bias is measured. We use the term *bias* to measure
158 deviations from the best (most likely) interpretation or model of the evidence,
159 $\Delta\mu = \hat{\mu} - \mu_{ML}$. In our benchmark case of a perfect court, it does not mat-
160 ter whether the court uses inquisitorial or adversarial decision-making because
161 both result in the same, best outcome.

162 2.2 Noisy Courts

163 The characterization of adversarial justice with a perfect court is simple and
164 results in an optimal decision, but it cannot explain the salient feature of the
165 adversarial system: the competing interpretations of evidence put forward by
166 the parties. Adding noise to the court's decision-making means that the court
167 sometimes makes errors by selecting the less likely alternative. Such errors
168 make it optimal for the parties to shade their interpretations away from the
169 most likely interpretation because there is some probability that a less likely
170 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 $\mathcal{L}'_s < \mathcal{L}_{ML}$ and the probability of winning equal zero.

171 of humans (and courts) to “detect signals,”¹⁰ as, for instance, in [Mueller and](#)
 172 [Weidemann \(2008\)](#).

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

$$\widetilde{\mathcal{L}}_P = \mathcal{L}_P \exp \xi_P \quad \text{and} \quad \widetilde{\mathcal{L}}_D = \mathcal{L}_D \exp \xi_D \quad (5)$$

175 where ξ_P and ξ_D are independently extreme value distributed *noise*, with mean
 176 0 and scale $1/\lambda$. The probability of a plaintiff win (i.e., the logit choice proba-
 177 bility for model P)¹¹ is

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

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

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

¹⁰For 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](#)).

¹¹See [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](#)).

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

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

191 In Table 1, we summarize three special cases of the contest success function
 192 in equation (6), for different values of the noise parameter λ .¹⁵ For $\lambda =$
 193 0, adversarial decision-making is uninformative, and the court’s decision is
 194 equivalent to a toss of a fair coin where the probability of a plaintiff win is
 195 $\tilde{\theta} = 1/2$. In this case, the court’s decision is independent of both the players’
 196 proposed models as well as the evidence.¹⁶ For $\lambda = 1$, equation (6) becomes
 197 the lottery version of the general Tullock contest success function. In addition,
 198 it is equivalent to Bayesian hypothesis testing for a court that assigns equal

¹³Others 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).

¹⁴For 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.

¹⁵Alternative 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”).

¹⁶This 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
$\lambda \rightarrow \infty$	<i>Perfect Court</i> benchmark; all-pay auction

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

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

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

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

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

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

211 The court's decision is an estimator of the mean, a likelihood-weighted

212 average of the means of the parties' models,

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

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

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

222 The court's decision is simply a binomial random variate with probability
223 $p = \tilde{\theta}$ and variance

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

224 The further apart the locations of the adversaries' models are, and the closer
225 $\tilde{\theta}$ is to $1/2$, the larger the variance of the court's decision is.¹⁷

¹⁷This 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, $\tilde{\theta}\mu_P + (1 - \tilde{\theta})\mu_D$. The variance of the expected value is, of course, zero.

226 **3 Equilibrium Results**

227 We relegate the formal results of the persuasion game to the Technical Ap-
 228 pendix. In this Section, we illustrate our main results for a parameterized
 229 version of the model. Evidence \bar{z} is a vector of n independent draws $z_i \in (0, 1)$
 230 with sample mean $\bar{\mu}$ from the same Beta(α, β) distribution with mean $\mu =$
 231 $E(z_i) = \alpha/(\alpha + \beta)$ and variance $\sigma^2 = \text{Var}(z_i) = \alpha\beta/[(\alpha + \beta)^2(1 + \alpha + \beta)]$.
 232 Each party $s = P, D$ chooses a model $Z_s = (\alpha_s, \beta_s)$ (i.e., the parameters
 233 for the Beta distribution) to explain the evidence vector \bar{z} , with any desired
 234 mean $0 < \mu_s < 1$ and any variance $\sigma_s^2 = \mu_s(1 - \mu_s)/(1 + \alpha_s + \beta_s)$ with
 235 $0 < \sigma_s^2 < \mu_s(1 - \mu_s)$. A Nash equilibrium in pure strategies of this zero-sum
 236 game is a strategy profile (Z_P^*, Z_D^*) such that each party chooses an optimal
 237 model, given the model chosen by its rival:

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

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

246 In Figure 1, we assume that the variance of the court’s decision noise is
 247 $\lambda = 1$, which makes the court’s decision equivalent to Bayesian hypothesis
 248 testing, where the court updates its prior belief (about relative merits of the
 249 two sides) with the likelihood to form a posterior. This assumption allows
 250 us to graph the court’s posterior belief as a weighted average of the parties’
 251 models, where the weights are proportional to the likelihoods of each model.
 252 Note that we are *not* saying the court uses Bayesian inference, only that for
 253 $\lambda = 1$, the final assessment of the court $\hat{\mu}$ is also the mean of a posterior belief,
 254 as if formed by Bayesian inference.

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

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

262 For $\lambda > 0$, the plaintiff engages in *payoff moderation* (e.g., Konrad, 2009),
 263 essentially trading off a higher payoff following a win for a lower probability of
 264 winning.¹⁸ In panel (a) of Figure 2, for example, we plot the likelihood ratio of
 265 the plaintiff’s model relative to that of the defendant. The graph shows that,
 266 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 Beta(α, β) distribution with density function $f(z; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} z^{1-\alpha} (1-z)^{1-\beta}$, where $B(\alpha, \beta)$ 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

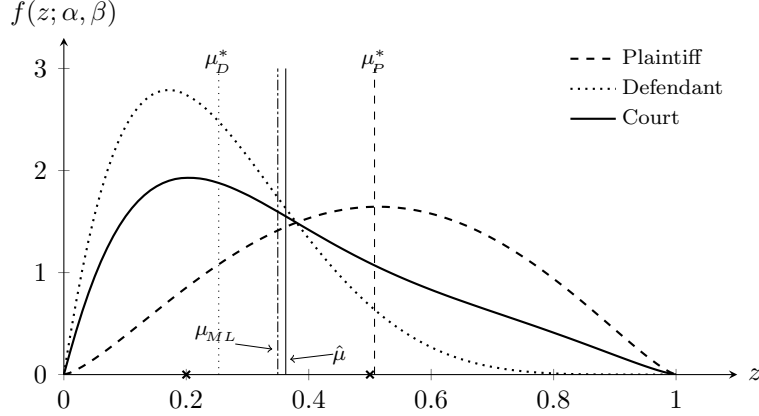
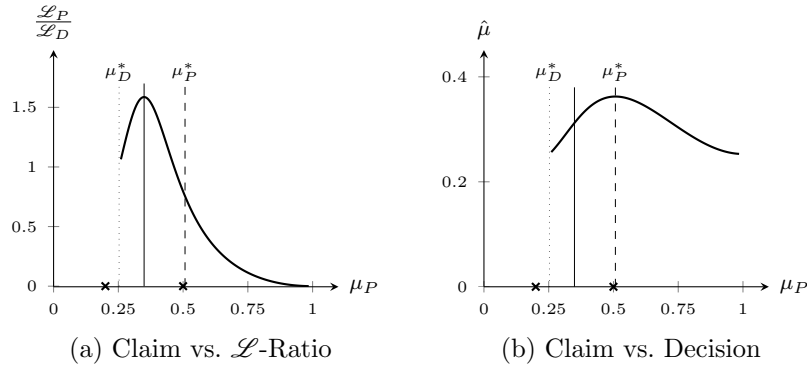


Figure 2: Payoff Moderation as a Best Response

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 $\mathcal{L}_P/\mathcal{L}_D$ 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).



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

272 **Result 2.** *The party with less favorable evidence follows an “obfuscation strat-*
 273 *egy” and chooses a model with (i) a location further away from the most likely*
 274 *model and (ii) with a spread larger than its rival's.*

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

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

282 The intuition for this result is simple. As long as the sample mean $\bar{\mu} \neq 0.5$,
283 the likelihood is asymmetric, which changes the tradeoff between the proba-
284 bility of winning and payoff following a win. For the disfavored party, it lowers
285 the likelihood penalty, or “credibility cost,” of claiming a model with a loca-
286 tion further from the evidence. Anyone who has participated in litigation or
287 has experience in political campaigns will recognize this as analogous to what
288 is known as an “obfuscation strategy.” When the evidence goes against you,
289 your best move is to claim that the evidence is not very informative. The de-
290 fendant’s best response is to choose a model with a mean closer to the evidence
291 and with a smaller spread. This might be called an “elucidation strategy,” as
292 the defendant is essentially claiming that the evidence is informative about
293 the mean.

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

296 We illustrate this result in Figure 1, where the evidence favors the defen-
297 dant, $\bar{\mu} < 1/2$, and the court’s decision is just above the maximum likelihood
298 estimate, $\hat{\mu} > \mu_{ML}$. This “bias” favors the plaintiff who, in equilibrium, offers
299 a more extreme interpretation of the evidence.¹⁹ However, the court’s bias is

¹⁹The 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

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

302 **Result 4.** *As court noise disappears, $\lambda \rightarrow \infty$, (i) the parties’ models con-*
303 *verge to the maximum likelihood estimator, $\mu_s \rightarrow \mu_{ML}$, for $s = P, D$, as does*
304 *the court’s estimator, $\hat{\mu} \rightarrow \mu_{ML}$; and (ii) the probability of a plaintiff win ap-*
305 *proaches 50%, $\tilde{\theta} \rightarrow 1/2$. The court’s estimator converges faster than do the*
306 *models of the parties.*

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

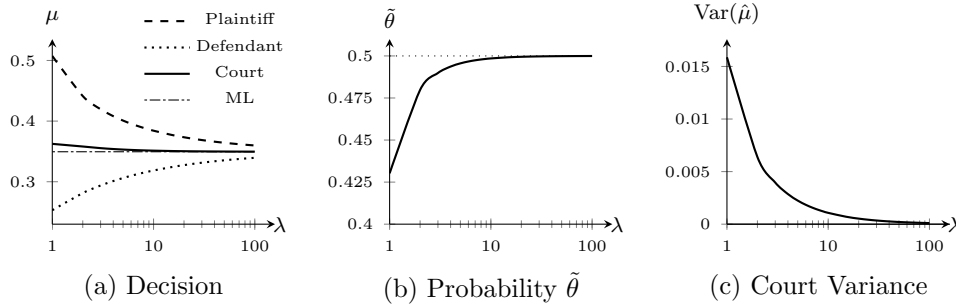
316 We illustrate the convergence in panel (a) of Figure 3, where we plot the
317 court’s equilibrium assessment as a function of court’s decision noise param-
318 eter. While both the competing models of the adversaries and the court’s
319 estimator converge to the maximum likelihood estimate, the court’s estima-
320 tor 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.

²⁰Note, 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, $\text{Var}(\hat{\mu}) = \tilde{\theta}(1 - \tilde{\theta})(\mu_P - \mu_D)^2$ in equation (9) against λ . The horizontal axes are on a logarithmic scale.



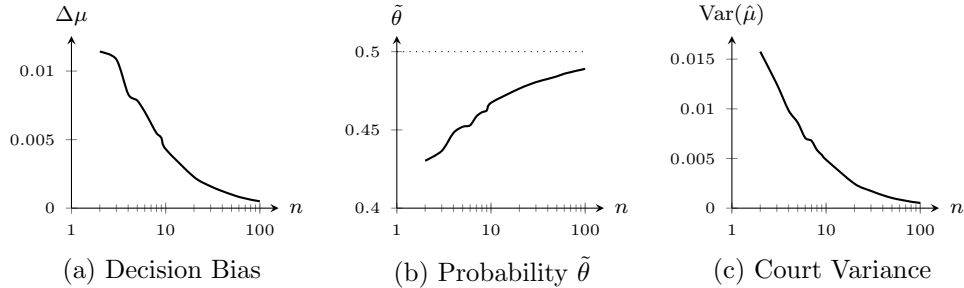
321 court is better able to assess credibility, the parties choose models closer to
 322 the maximum likelihood estimator, but on either side of it. In this sense, the
 323 parties’ models tend to cancel each other out, which has a salutary effect on
 324 the adversarial court’s decision-making.

325 In panel (b) of Figure 3, we plot the probability $\tilde{\theta}$ of a plaintiff win as
 326 $\lambda \rightarrow \infty$. As court noise disappears, the parties choose models that have
 327 the same likelihoods (as their models converge to the maximum likelihood
 328 estimator), so the probability of a win approaches $1/2$. This limit corresponds
 329 to the 50% probability of a trial win found by Priest and Klein (1984), albeit
 330 for a different reason. Their explanation is built around a selection bias story
 331 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 $n \in \{2, \dots, 100\}$, 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.



332 win rate is driven by competition between the parties.²¹

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

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

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

²¹Note, however, that because, for $\lambda \rightarrow \infty$, 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.

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

347 We illustrate Result 5 in Figure 4. In panel (a), we plot the decision
348 bias $\Delta\mu = \hat{\mu} - \mu_{ML}$ against the evidence sample size n . The decision bias in
349 this graph is the mean of the decision bias for 250 random evidence samples
350 for each n , drawn from a Beta(1, 2) distribution. Decision bias disappears as
351 both parties' models converge to the mean for the evidence generating process,
352 $\mu = \alpha/\alpha+\beta = 1/3$.

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

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

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

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

366 4 Discussion

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

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

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

(2012).

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

387 In addition to the implications relating to litigation and arbitration, our
388 model can be applied to the problem of delegating decision rights to subordi-
389 nates who can end up disagreeing with one another or recommending opposing
390 courses of action. Typically, managers higher up in the hierarchy are respon-
391 sible for resolving these disagreements. [Fama and Jensen \(1983\)](#) call this the
392 separation of “decision management” by subordinates from “decision control”
393 by a superior. Our results suggest that, even if superiors resolve disagreements
394 by appealing to evidence, the superior’s decisions are likely to be noisy and
395 potentially biased if the alternatives are strategically chosen by subordinates.
396 This can be thought of as another kind of agency cost.

References

- 398 AMBRUS, A., V. BARANOVSKIY, AND A. KOLB (2015): “A Delegation-Based
399 Theory of Expertise,” ERID Working Paper 193, Duke University, available at
400 <http://ssrn.com/abstract=2643672>. [Cited on pages 2 and 11.]
- 401 AUSTEN-SMITH, D. (1993): “Interested Experts and Policy Advise: Multiple Re-
402 ferrals under Open Rules,” *Games and Economic Behavior*, 5, 3–43. [Cited on
403 page 7.]
- 404 BAYE, M. R., D. KOVENOCK, AND C. G. DE VRIES (1996): “The All-Pay Auction
405 with Complete Information,” *Economic Theory*, 8, 291–305. [Cited on page 12.]
- 406 ——— (2005): “Comparative Analysis of Litigation System: An Auction-Theoretic
407 Approach,” *Economic Journal*, 115, 583–601. [Cited on page 11.]
- 408 BEBCHUK, L. A. AND A. KLEMENT (2012): “Negative-Expected-Value Suits,” in
409 *Procedural Law and Economics*, ed. by C. W. Sanchirico, Edward Elgar, vol. 8
410 of *Encyclopedia of Law and Economics*, chap. 12, 2nd ed. [Cited on pages 20
411 and 22.]
- 412 CHENG, E. K. (2013): “Reconceptualizing the Burden of Proof,” *Yale Law Journal*,
413 122, 1254–1279. [Cited on page 12.]
- 414 CHOI, S., M. GULATI, AND E. A. POSNER (2012): “What Do Federal District
415 Judges Want? An Analysis of Publications, Citations, and Reversals,” *Journal*
416 *of Law, Economics, and Organization*, 28, 518–549. [Cited on page 7.]
- 417 CHOWDHURY, S. M. AND R. M. SHEREMETA (2011): “A Generalized Tullock
418 Contest,” *Public Choice*, 147, 413–420. [Cited on page 11.]

- 419 CHUNG, T.-Y. (1996): “Rent-Seeking Contest When the Prize Increases with Ag-
420 gregate Efforts,” *Public Choice*, 87, 55–66. [Cited on page 11.]
- 421 CRAWFORD, V. P. (1979): “Compulsory Arbitration Schemes,” *Journal of Political*
422 *Economy*, 87, 131–160. [Cited on page 5.]
- 423 DAUGHETY, A. F. AND J. F. REINGANUM (2000a): “Appealing Judgments,”
424 *RAND Journal of Economics*, 31, 502–525. [Cited on pages 7 and 22.]
- 425 ——— (2000b): “On the Economics of Trials: Adversarial Process, Evidence, and
426 Equilibrium Bias,” *Journal of Law, Economics, and Organization*, 16, 365–394.
427 [Cited on pages 2 and 23.]
- 428 DEGROOT, M. (1970): *Optimal Statistical Decisions*, New York, N.Y.: McGraw-
429 Hill. [Cited on page 9.]
- 430 DEWATRIPONT, M. AND J. TIROLE (1999): “Advocates,” *Journal of Political Econ-*
431 *omy*, 107, 1–39. [Cited on page 2.]
- 432 EMONS, W. AND C. FLUET (2009): “Accuracy Versus Falsification Costs: The
433 Optimal Amount of Evidence under Different Procedures,” *Journal of Law, Eco-*
434 *nomics, and Organization*, 25, 134–156. [Cited on page 7.]
- 435 FAMA, E. F. AND M. C. JENSEN (1983): “Separation of Ownership and Control,”
436 *Journal of Law and Economics*, 26, 301–325. [Cited on page 23.]
- 437 FARBER, H. S. AND M. H. BAZERMAN (1986): “The General Basis of Arbitrator
438 Behavior: An Empirical Analysis of Conventional and Final-Offer Arbitration,”
439 *Econometrica*, 54, 1503–1528. [Cited on pages 7 and 13.]

- 440 FIENBERG, S. E. AND M. L. STRAF (1991): “Statistical Evidence in the US Courts:
441 An Appraisal,” *Journal of the Royal Statistical Society. Series A (Statistics in*
442 *Society)*, 154, 49–59. [Cited on page 2.]
- 443 FROEB, L. M. AND B. KOBAYASHI (1996): “Naïve, Biased yet Bayesian: Can
444 Juries Interpret Selectively Produced Evidence?” *Journal of Law, Economics,*
445 *and Organization*, 12, 257–276. [Cited on page 23.]
- 446 ——— (2001): “Evidence Production in Adversarial vs. Inquisitorial Regimes,”
447 *Economics Letters*, 70, 267–272. [Cited on pages 2 and 23.]
- 448 ——— (2012): “Adversarial versus Inquisitorial Justice,” in *Procedural Law and*
449 *Economics*, ed. by C. W. Sanchirico, Edward Elgar, vol. 8 of *Encyclopedia of Law*
450 *and Economics*, chap. 1, 2nd ed. [Cited on pages 2 and 23.]
- 451 GENTZKOW, M. AND E. KAMENICA (2016): “Competition in Persuasion,” Unpub-
452 lished manuscript, Stanford University. [Cited on page 3.]
- 453 GILLIGAN, T. W. AND K. KREHBIEL (1989): “Asymmetric Information and Leg-
454 islative Rules with a Heterogeneous Committee,” *American Journal of Political*
455 *Science*, 53, 459–490. [Cited on page 3.]
- 456 ——— (1997): “Specialization Decisions Within Committee,” *Journal of Law, Eco-*
457 *nomics, and Organization*, 13, 366–386. [Cited on page 23.]
- 458 GLAZER, J. AND A. RUBINSTEIN (2001): “Debates and Decisions: On a Rationale
459 of Argumentation Rules,” *Games and Economic Behavior*, 36, 158–173. [Cited
460 on page 3.]
- 461 GOOD, I. J. (1979): “Studies in the History of Probability and Statistics. XXXVII

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

464 HIRSHLEIFER, J. AND E. OSBORNE (2001): “Truth, Effort, and the Legal Battle,”
465 *Public Choice*, 108, 169–195. [Cited on page 11.]

466 IOSSA, E. AND B. JULLIEN (2012): “The Market for Lawyers and Quality Layers
467 in Legal Services,” *RAND Journal of Economics*, 43, 677–704. [Cited on page 7.]

468 IOSSA, E. AND G. PALUMBO (2007): “Information Provision and Monitoring of the
469 Decision-Maker in the Presence of an Appeal Process,” *Journal of Institutional
470 and Theoretical Economics*, 163, 657–682. [Cited on page 2.]

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

473 KAPLAN, T., I. LUSKI, A. SELA, AND D. WETTSTEIN (2002): “All-Pay Auctions
474 with Variable Rewards,” *Journal of Industrial Economics*, 50, 417–430. [Cited on
475 page 11.]

476 KARTIK, N. (2009): “Strategic Communication with Lying Costs,” *Review of Eco-
477 nomic Studies*, 76, 1359–1395. [Cited on page 7.]

478 KARTIK, N., M. OTTAVIANI, AND F. SQUINTANI (2007): “Credulity, Lies, and
479 Costly Talk,” *Journal of Economic Theory*, 134, 93–116. [Cited on page 7.]

480 KATZ, A. W. (1988): “Judicial Decisionmaking and Litigation Expenditure,” *In-
481 ternational Review of Law and Economics*, 8, 127–143. [Cited on page 11.]

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

- 484 KONRAD, K. A. AND H. SCHLESINGER (1997): “The Risk-Aversion in Rent-Seeking
485 and Rent-Augmenting Games,” *Economic Journal*, 107, 1671–1683. [Cited on
486 page 11.]
- 487 KRISHNA, V. AND J. MORGAN (2001a): “Asymmetric Information and Legislative
488 Rules: Some Amendments,” *American Political Science Review*, 95, 435–452.
489 [Cited on page 3.]
- 490 ——— (2001b): “A Model of Expertise,” *Quarterly Journal of Economics*, 116,
491 747–775. [Cited on page 3.]
- 492 LEONARD, T. AND J. S. J. HSU (1999): *Bayesian Methods*, Cambridge, UK: Cam-
493 bridge University Press. [Cited on page 38.]
- 494 MAUET, T. A. (2007): *Trial Techniques*, New York, N.Y.: Wolters Kluwer Law &
495 Business, 7th ed. [Cited on page 3.]
- 496 MCFADDEN, D. (1974): “Conditional Logit Analysis of Qualitative Choice Behav-
497 ior,” in *Frontiers in Econometrics*, ed. by P. Zarembka, N.Y.: Academic Press,
498 105–142. [Cited on page 10.]
- 499 MCKELVEY, R. AND T. PALFREY (1995): “Quantal Response Equilibria for Normal
500 Form Games,” *Games and Economic Behavior*, 10, 6–38. [Cited on page 10.]
- 501 MILGROM, P. AND J. ROBERTS (1986): “Relying on the Information of Interested
502 Parties,” *RAND Journal of Economics*, 17, 18–32. [Cited on pages 2 and 23.]
- 503 MUELLER, S. T. AND C. T. WEIDEMANN (2008): “Decision Noise: An Explanation
504 for Observed Violations of Signal Detection Theory,” *Psychonomic Bulletin &
505 Review*, 15, 465–494. [Cited on page 10.]

- 506 NALEBUFF, B. (1987): “Credible Pretrial Negotiation,” *RAND Journal of Eco-*
507 *nomics*, 18, 198–210. [Cited on page 20.]
- 508 PARDO, M. S. AND R. J. ALLEN (2008): “Juridical Proof and the Best Explana-
509 tion,” *Law and Philosophy*, 27, 223–268. [Cited on page 7.]
- 510 PFEFFER, J. AND R. I. SUTTON (2006): “Evidence-Based Management,” *Harvard*
511 *Business Review*, 84, 62–74. [Cited on page 22.]
- 512 PRIEST, G. L. AND B. KLEIN (1984): “The Selection of Disputes for Litigation,”
513 *Journal of Legal Studies*, 13, 1–56. [Cited on page 20.]
- 514 RANTAKARI, H. (2016): “Soliciting Advice: Active versus Passive Principals,” *Jour-*
515 *nal of Law, Economics, and Organization*. [Cited on pages 2 and 3.]
- 516 SAKS, M. J. AND S. L. NEUFELD (2011): “Convergent Evolution in Law and
517 Science: The Structure of Decision-Making Under Uncertainty,” *Law, Probability*
518 *and Risk*, 10, 133–148. [Cited on page 12.]
- 519 SHIN, H. S. (1994): “The Burden of Proof in a Game of Persuasion,” *Journal of*
520 *Economic Theory*, 64, 253–264. [Cited on page 23.]
- 521 ——— (1998): “Adversarial and Inquisitorial Procedures in Arbitration,” *RAND*
522 *Journal of Economics*, 29, 378–405. [Cited on page 2.]
- 523 SKAPERDAS, S. (1996): “Contest Success Functions,” *Economic Theory*, 7, 283–290.
524 [Cited on page 11.]
- 525 SKAPERDAS, S. AND S. VAIDYA (2012): “Persuasion as a Contest,” *Economic The-*
526 *ory*, 51, 465–486. [Cited on pages 2 and 11.]

- 527 TANFORD, J. A. (2009): *The Trial Process: Law, Tactics and Ethics*, Newark, N.J.:
528 LexisNexis, 4th ed. [Cited on page 3.]
- 529 TENN, S., L. M. FROEB, AND S. TSCHANTZ (2010): “Merger Effects When Firms
530 Compete by Choosing Both Price and Promotion,” *International Journal of In-*
531 *dustrial Organization*, 28, 695–707. [Cited on page 2.]
- 532 TRAIN, K. (2009): *Discrete Choice Methods with Simulation*, Cambridge, UK: Cam-
533 bridge University Press, 2nd ed. [Cited on page 10.]
- 534 TULLOCK, G. (1980): “Efficient Rent Seeking,” in *Toward a Theory of the Rent-*
535 *Seeking Society*, ed. by J. M. Buchanan, R. D. Tollison, and G. Tullock, College
536 Station, TX: Texas A&M University Press, 97–112. [Cited on page 11.]
- 537 WAGNER, W. (2005): “The Perils of Relying on Interested Parties to Evaluate
538 Scientific Quality,” *American Journal of Public Health*, 95, s99–s106. [Cited on
539 page 2.]
- 540 WERDEN, G. J. AND L. M. FROEB (1994): “The Effects of Mergers in Differen-
541 tiated Products Industries: Logit Demand and Merger Policy,” *Journal of Law,*
542 *Economics, and Organization*, 10, 407–426. [Cited on page 2.]
- 543 WITTMAN, D. (1986): “Final-Offer Arbitration,” *Management Science*, 32, 1551–
544 1561. [Cited on page 7.]
- 545 YILANKAYA, O. (2002): “A Model of Evidence Production and Optimal Standard
546 of Proof and Penalty in Criminal Trials,” *Canadian Journal of Economics/Revue*
547 *canadienne d’économique*, 35, 385–409. [Cited on page 23.]

Technical Appendix

549 **A The General Persuasion Game**

550 **A.1 Introduction**

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

555 **A.2 Problem**

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

566 **A.3 Notation**

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

575 The principal solicits advice from two agents, $i = L, R$. Each agent’s advice
 576 is modeled as an interpretation that characterizes the sample as coming from
 577 a process of type y_i with credibility $\chi_i \geq 0$. The principal assesses the agents’
 578 advice and chooses the most credible of the two. We assume this assessment
 579 of credibility is noisy and refer to it as $\tilde{\chi}_i = \chi_i \exp \xi_i$ for $i = L, R$, where ξ_i are
 580 independently extreme value (or Gumbel) distributed with mean 0 and scale
 581 $1/\lambda$.²⁵ The principal therefore follows agent R ’s advice if $\tilde{\chi}_R > \tilde{\chi}_L$ and agent
 582 L ’s otherwise. If $\tilde{\chi}_L = \tilde{\chi}_R$, then the principal flips a fair coin. This is akin to
 583 the structure of the logit choice model. The principal sides with agent R with
 584 probability

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

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

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

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

586 and of agent L 's advice as

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

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

594 We further limit the agent's strategy space to be a compact and convex
595 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
596 of feasible strategies is characterized by a type-credibility tradeoff. In other
597 words, the further advice y_i is from the objectively best assessment \bar{y} , the
598 less credible this advice will be with a value of χ_i , or, alternatively, the more
599 incredible the advice will be with a higher value of $|x_i|$. Extreme advice with
600 very high (or low) type y_i and low incredibility $|x_i|$ is therefore not feasible,
601 and the strategy space is convex.

602 Using the expressions for agent's incredibility, the probability that the prin-

603 cipal follows agent R 's advice in equation (A1) can be rewritten as

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

604 The principal's assessment of the process type is y_R when she follows agent
 605 R 's advice and y_L when she follows L 's advice. In expectations, the principal's
 606 assessment²⁶ and decision is thus

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

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

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

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

609 where

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

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

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

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

613 agents' payoffs are directly affected by the principal's assessment of type.
614 Agent L prefers low values of \hat{y} , whereas agent R prefers high values. For given
615 $y_R > y_L$, the expression for the principal's expected decision in equation (A5)
616 implies that both agents will choose the most credible interpretations given
617 their advice types y_i . For agent L , this means the highest possible $x_L \in \mathbb{R}^-$;
618 and for agent R the lowest possible $x_R \in \mathbb{R}^+$. We define these "incredibility
619 frontiers" as

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

620 and

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

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

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

630 **A.4 Equilibrium Concept**

631 A *persuasion game* is a simultaneous-move, non-cooperative game between
 632 two agents $i = L, R$ providing strategic advice $a_i \in A_i$ to maximize payoffs
 633 $\pi_L = -\hat{y}(a_L, a_R)$ for agent L and $\pi_R = \hat{y}(a_L, a_R)$ for agent R with $\hat{y}(a_L, a_R)$
 634 defined in equation (A6). A Nash equilibrium in this game is a strategy profile
 635 (a_L^*, a_R^*) such that

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

636 From the expression for the principal's decision in equation (A6), we can
 637 conclude that, because agent L 's incredibility is by definition negative, $x_L < 0$,
 638 if $m(a'_L, a'_R) > m(a_L, a_R)$, then either $m(a'_L, a_R) > m(a_L, a_R)$ or $m(a_L, a'_R) >$
 639 $m(a_L, a_R)$. In other words, if a strategy profile (a_L, a_R) does not result in a
 640 maximum for m , at least one of the agents can unilaterally move to increase
 641 the slope.

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

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

645 **A.5 Equilibrium Results**

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

648 **A.5.1 Nash Equilibrium**

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

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

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

669 Second, if \hat{x}_L and \hat{x}_R are strictly concave, differentiable functions, de-
670 fined on a convex subset of the real line, with unbounded derivatives, then

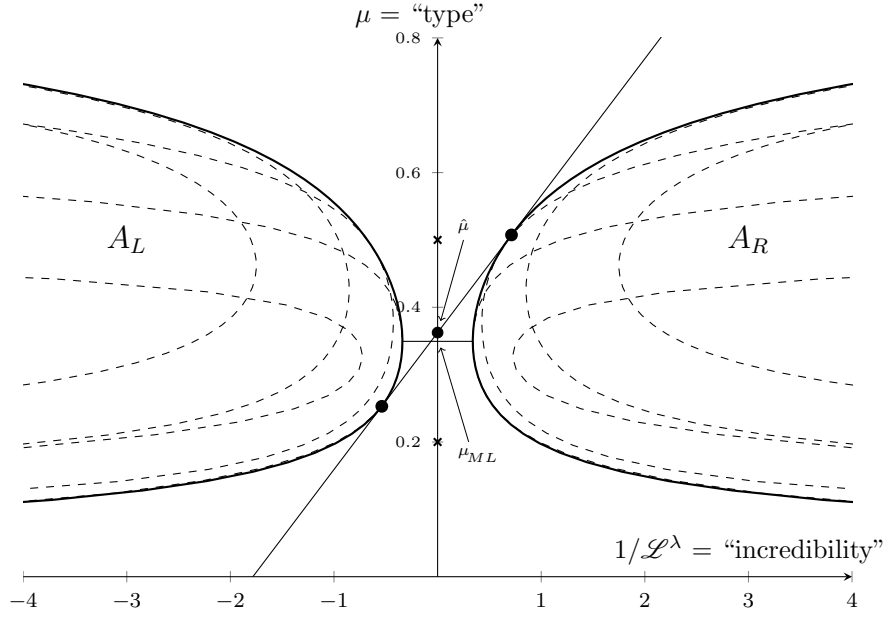
671 these functions have unique maxima and minima, respectively, and define
672 the relevant frontiers of the strategy sets. These assumptions also guaran-
673 tee the existence of a unique Nash equilibrium solution $a_L^* = (\hat{x}_L(y_L^*), y_L^*)$ and
674 $a_R^* = (\hat{x}_R(y_R^*), y_R^*)$, and the line through these points is simultaneously tangent
675 to both the L and R curves.

676 In Figure A1, we relate the general game to our litigation game in the
677 main text by using the specific parameterization of the game in the main
678 text. The unobservable type is the theoretical mean of the Beta(α, β) dis-
679 tribution, $y = \mu$, and the inverse credibility is the reciprocal likelihood or
680 “incredibility,” $x = 1/\mathcal{L}^\lambda$. Agent L is the defendant D (preferring low-
681 valued outcomes) and agent R is the plaintiff (preferring high-valued out-
682 comes), where A_L is the set $(-1/\mathcal{L}_D^\lambda, \mu_D)$ and A_R is the set $(1/\mathcal{L}_P^\lambda, \mu_P)$, both
683 defined over all possible Beta(α, β) distribution functions. This means, there
684 are multiple parameterizations to obtain a fixed $\mu = \alpha / (\alpha + \beta)$ and varying
685 $\sigma^2 = \mu(1 - \mu) / (1 + \alpha + \beta)$. Alternatively, there are multiple parameteriza-
686 tions (and thus likelihoods) to obtain a fixed σ^2 and varying μ (Leonard and
687 Hsu, 1999).

688 With this set up, the x -axis in Figure A1 measures incredibility $1/\mathcal{L}^\lambda$ as
689 a function of the type μ plotted on the y -axis. The dashed lines represent the
690 reciprocal likelihoods of various types, for various fixed values of variance σ^2 .
691 This gives a family of overlapping curves, the envelope of which is also drawn,
692 and whose union defines the A_R set to the right, and which is mirrored in the
693 A_L set to the left. The line of maximum slope is drawn between the points
694 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 .



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

699 A.5.2 Payoff Shading

700 We have denoted the objectively best assessment of the type as \bar{y} . Suppose that
 701 this type \bar{y} is also the most credible advice the agents can give. That means, the
 702 maximum of $\hat{x}_L < 0$ and the minimum of $\hat{x}_R > 0$ (i.e., the points where these

703 come closest to the y -axis) are at the same \bar{y} . This then implies that that the
704 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,
705 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
706 the incredibility functions \hat{x}_i cannot be differentiable and have a corner at \bar{y} ,
707 agents will “shade” their advice, with L offering a type y_L^* less than the most
708 likely \bar{y} , and R offering a type y_R^* greater than this \bar{y} .

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

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

717 A.5.3 Bias

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

726 distribution, then there is more “room” to explain the evidence with a larger
 727 μ than with a smaller μ .

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

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

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

736 Because, by Theorem A2, agent L shades down, $y_L < \bar{y}$, and agent R shades
 737 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
 738 only off equilibrium. For the properties of the equilibrium decision \hat{y}^* we can
 739 therefore ignore these values. This means that we may as well take a single
 740 function \hat{x} describing both parties’ incredibility functions: $\hat{x}(y) = -\hat{x}_L(y)$ for
 741 $y \leq \bar{y}$ and $\hat{x}(y) = \hat{x}_R(y)$ for $y \geq \bar{y}$. The bias of the principal’s decision relative
 742 to \bar{y} is then determined by how quickly the incredibility increases for $y > \bar{y}$ as
 743 compared to $y < \bar{y}$ as a function of the difference from the most credible type
 744 \bar{y} . In Theorem A3 below, we make use of the following definitions:

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

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

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

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

- 758 1. *If $\hat{x}(y)$ is symmetric, then $y_R^* - \bar{y} = \bar{y} - y_L^*$ and $\hat{y}^* = \bar{y}$.*
- 759 2. *If agent L has lower credibility costs, then $\hat{y}^* < \bar{y}$, and the equilibrium*
760 *assessment is biased down. If agent R has lower credibility costs, then*
761 *$\hat{y}^* > \bar{y}$, and the equilibrium assessment is biased up.*
- 762 3. *If agent L has monotonically lower credibility costs, then agent L's ad-*
763 *vice a_L^* exhibits more shading than agent R's advice, $\bar{y} - y_L^* > y_R^* - \bar{y}$.*
764 *Analogously for agent R.*

765 *Proof.* 1. Suppose $\hat{x}(y)$ is symmetric (Definition A1). If $y_R^* = \bar{y} + \delta$, then
766 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 $\hat{y}^* \leq$
767 $\hat{y}(a_L', a_R^*) = \bar{y}$ since L can do no worse than respond to a_R^* with strategy
768 a_L' . Similarly if $y_L^* = \bar{y} - \delta$, taking $y_R' = \bar{y} + \delta$ shows $\hat{y}^* \geq \bar{y}$. Hence
769 $\hat{y}^* = \bar{y}$, and the same $\delta = y_R^* - \bar{y} = \bar{y} - y_L^*$.

770 2. With lower credibility costs (Definition A2) for agent L , $\hat{x}(\bar{y} - \delta) <$
771 $\hat{x}(\bar{y} + \delta)$ for all $\delta > 0$. If $y_R^* = \bar{y} + \delta$, then take $y_L' = \bar{y} - \delta$ and
772 $a_L' = (-\hat{x}(y_L'), y_L')$. Because $\hat{x}(\bar{y} - \delta) < \hat{x}(\bar{y} + \delta)$, $\hat{y}^* \leq \hat{y}(a_L', a_R^*) < \bar{y}$.
773 Analogously for agent R .

774 3. With monotonically lower credibility costs (Definition A3) for agent L ,
775 $\hat{x}(\bar{y} + \delta) - \hat{x}(\bar{y} - \delta)$ is strictly increasing. Then, for $\delta = y_R^* - \bar{y}$, the
776 derivative $-\hat{x}'(\bar{y} - \delta) < \hat{x}'(\bar{y} + \delta) = \hat{x}'(y_R^*) = -\hat{x}'(y_L^*)$ because the max-
777 imum slope line is tangent to both incredibility curves at the equilib-
778 rium solution. But $\hat{x}'(y)$ is strictly increasing so $y_L^* < \bar{y} - \delta$, that is,
779 $\delta = y_R^* - \bar{y} < \bar{y} - y_L^*$. The analogous arguments hold when R has lower
780 credibility costs. Q.E.D.

781 A.5.4 Convergence as $n \rightarrow \infty$

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

788 More specifically, suppose a family of incredibility functions denoted by
789 $\hat{x}(y|n)$ are parameterized by a variable n denoting the amount of evidence
790 available. Suppose that the most credible \bar{y} is the same for all incredibility
791 functions $\hat{x}(y|n)$. Scaling the incredibility by a constant factor does nothing to
792 change the outcome of the game. We thus assume that these functions are all

793 normalized to one, $\hat{x}(\bar{y}|n) = 1$. The notion of narrowing incredibility functions
 794 is then captured formally as a hypothesis of the following consistency result.

795 **Theorem A4.** *Let the equilibrium assessment in the persuasion game with*
 796 *incredibility function $\hat{x}(y|n)$ be denoted by \hat{y}_n^* . Suppose that for every $\epsilon > 0$,*
 797 *for all sufficiently large n , and any y we have $\hat{x}(y|n) > |y - \bar{y}|/\epsilon$. Then*
 798 $\lim_{n \rightarrow \infty} \hat{y}_n^* = \bar{y}$.

799 *Proof.* Suppose $\epsilon > 0$ is given and take N so for all $n \geq N$ and any y we
 800 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
 801 equilibrium strategies for the persuasion game with $\hat{x}(y|n)$. Let $a'_L = (-1, \bar{y})$
 802 be the maximally credible strategy for agent L . Then

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

803 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
 804 fashion. Hence, for every $\epsilon > 0$, for all sufficiently large n , $|\hat{y}_n^* - \bar{y}| < \epsilon$, that
 805 is, $\lim_{n \rightarrow \infty} \hat{y}_n^* = \bar{y}$. Q.E.D.

806 This result is stronger than what we illustrate with Result 5 in the main text
 807 where we show that the bias decreases with more evidence. In Theorem A4,
 808 we show that the equilibrium assessment converges to the most credible assess-
 809 ment \bar{y} . In other words, any bias in assessments away from the most credible
 810 \bar{y} due to the adversarial process disappears with increasing evidence. Advice
 811 that deviates from the most credible explanation simply faces an increasing
 812 credibility penalty the more evidence there is. The argument gives a bound for

813 the deviation of \hat{y}_n^* from \bar{y} , but the argument cannot tell us that this bias de-
814 creases monotonically with n without much more detailed assumptions about
815 the dependence of $\hat{x}(y|n)$ on n .