

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

ONLINE APPENDIX

Luke M. Froeb* Bernhard Ganglmair[†] Steven Tschantz[‡]

September 2016

Abstract

This is the Online Appendix for “Adversarial Decision-Making: Choosing Between Models Constructed by Interested Parties,” *Journal of Law and Economics*, Volume 59.

*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

The General Persuasion Game

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.

1. 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.

2. 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 $\tilde{\chi}_i = \chi_i \exp \xi_i$ for $i = L, R$, where ξ_i are independently extreme value distributed with location 0 and scale $1/\lambda$.¹ The principal therefore follows agent R 's advice if $\tilde{\chi}_R > \tilde{\chi}_L$ and agent L 's otherwise. If $\tilde{\chi}_L = \tilde{\chi}_R$, 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

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

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

$$x_R = \frac{1}{\chi_R^\lambda} \quad (\text{A2})$$

and of agent L 's advice as

$$x_L = -\frac{1}{\chi_L^\lambda}. \quad (\text{A3})$$

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

¹The structure in Jia (2008) is less restrictive, requiring the random variable ξ_i to belong to the *inverse exponential distribution*.

agent L and $x_R \in \mathbb{R}^+$ 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 (that is, 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

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

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 assessment² and decision is thus

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

²This 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 .

It is the credibility-weighted sum of the agent's location advice. We can further rewrite the expression in equation (A5) as

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

where

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

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 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 (for example, 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.

3. 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

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

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 $m(a_L^*, a_R^*)$.

4. 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.

4.1. 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,” $x = 1/\mathcal{L}^\lambda$. 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 $(-1/\mathcal{L}_D^\lambda, \mu_D)$ and A_R is the set $(1/\mathcal{L}_P^\lambda, \mu_P)$, both defined over all possible Beta(α, β) distribution functions. This means, there are multiple parameterizations to obtain a fixed $\mu = \alpha / (\alpha + \beta)$ and varying $\sigma^2 = \mu(1 - \mu) / (1 + \alpha + \beta)$. 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 $1/\mathcal{L}^\lambda$ as a function of the type μ plotted on the y -axis. The dashed lines represent the reciprocal likelihoods for varying proposed type y , holding the variance σ^2 fixed. This gives a family of overlapping curves, the envelope of which is

also drawn (solid curve), 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, a_i^* , for the two sides. The 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 (that is, minimum incredibility) value \bar{y} , marked by a horizontal line between the “peaks” of the two sets, A_L and A_R .

[Figure 1 about here.]

4.2. *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 $\hat{x}_R > 0$ (that is, 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).

4.3. 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.

4.4. Convergence as $n \rightarrow \infty$

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*

$$\lim_{n \rightarrow \infty} \hat{y}_n^* = \bar{y}.$$

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, $\lim_{n \rightarrow \infty} \hat{y}_n^* = \bar{y}$. 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 .

References

- Jia, Hao. 2008. "A Stochastic Derivation of the Ratio Form of Contest Success Functions." *Public Choice* 135:125–130.
- Konrad, Kai A. 2009. *Strategy and Dynamics in Contests*. Oxford, UK: Oxford University Press.
- Leonard, Thomas and John S. J. Hsu. 1999. *Bayesian Methods: An Analysis for Statisticians and Interdisciplinary Researchers*. Cambridge, UK: Cambridge University Press.

List of Figures

A1	Advice in the Simple Litigation Game	16
----	--	----

Figure A1: Advice in the Simple Litigation Game

