A Conversational War of Attrition

Bognar, Meyer-ter-Vehn, and Smith

1. Overview. Debate by individuals in committees is a widespread and important phenomenon. Alas, it is neither (*i*) perfect nor (*ii*) rapid. A notable paper by Li, Rosen, and Suen (2001) (LRS2001) shed light on the imperfection. It showed how cardinal preference differentials between individuals endogenously coarsen communication, in the spirit of Crawford and Sobel (1982); voting exogenously secures this coarsening.

Our original submission instead had the second goal in mind — shedding light on slow communication — why and how it happens, and its efficiency properties. We felt that information misrepresentation with divergent preferences was well-studied, but that costly dynamic communication was not: Anyone who has endured the time-consuming travails of formulating and digesting arguments in meetings should see its importance. If people could just specify their signals, many a debate would end immediately; yet none among us ever just summarizes our total information with a likelihood ratio, even in academic debates (e.g. tenure) — and we are theorists![1]

We thus explored a model of how debate proceeds with "communication frictions". Like LRS2001, we assumed two debaters and captured coarse communication using voting. For this was both a tractable and salient case, and also highly motivated. For parsimony, we focused on informational differences, with no preference conflict. We found that this tandem of coarse yet costly debate induced a war of attrition with curious welfare properties that mimicked typical features of debate. While prior work focused on strategic obfuscation issues, we explored how one might vote against one's preference simply to extend a debate and learn more. For the *option value* of learning guides all choices in our model. Our war of attrition was also new. Normally, an exogenous delay cost forever opposes a strategic incentive to delay. For us, the ex ante conflict eventually resolves itself so that the debater who quits is convinced.

Pondering review comments, *we have now realized that our model insights are enriched by allowing divergent preferences*: As with our original submission, a second order nonlinear difference equation in cutoffs characterizes the solution — since current payoffs depend on your action, and your colleague's last and next choice. Analyzing this has

[1] An impressive literature addresses coarse Bayesian inference (e.g. Andrea Wilson's well-cited finite memory paper, forthcoming in *Econometrica*) or coarse communication (the herding literature assumes only binary actions were observed and communicated). That jurors have trouble implementing Bayes' rule has long been documented. See Robert J. MacCoun, "Experimental Research on Jury Decision-Making", 30 *Jurimetrics J.* 223 (1989).

proved complex, since we have discovered an equilibrium bifurcation. Our original zick-zack equilibrium exemplified *ambivalent debating*, for at each stage, the costs of debate include the possibility of a mistaken decision.[2] This prediction still obtains with small conflicts of interest, or when jurors are informationally biased against their preferences. Alas, we have verified that sufficient conflicts of interest give rise to a more acrimonious equilibrium with *turf guarding* — a pure war of attrition, where at each stage jurors sees no mistaken decision costs at all. These debates no longer end amicably, but when one side "throws in the towel".

Since our model now has three dimensions of explanatory power — the preference bias, the delay cost, and the information precision — there are three instructive limit debate cases when each vanishes. As we converge upon coincident preferences, we return to our original model and its ambivalent arguing. As the delay costs vanish, the voting model of LRS2001 emerges; *however*, a continuity failure arises, since our equilibrium entails turf guarding for all small delay costs, but the main equilibrium finding of LRS2001 is ambivalent debating (see their zig-zag pictures).[3] For absent any waiting costs, their debaters solely trade off the chances of correct and incorrect decisions. Finally, as the information quality vanishes, our model converges upon a standard pure war of attrition that is essentially a shouting match until a concession.

2. Model. Two partially informed jurors, Lones and Moritz, alternately propose verdicts (acquit $\mathcal{A}$ or convict $\mathcal{C}$) until agreement; each round costs $\kappa > 0$. Their ordinal preferences coincide: both want to acquit the innocent and convict the guilty (50-50 prior). The hawk Lones worries more about acquitting the guilty; the dove Moritz is the opposite way. Their decision costs are $u_L = 1, u_M = 1+\beta$ if they convict the innocent, and $u_L = 1+\beta, u_M = 1$ if they acquit the guilty, with *preference bias* $\beta \geq 0$.

Types are log-likelihood ratios in favor of guilt for Lones, and innocence for Moritz. With own type y and other's type x, let $\pi(y,x)$ be the chance of one's *natural verdict* ($\mathcal{A}$ for Moritz and $\mathcal{C}$ for Lones).[4] The ex-post loss is thus $(1-\pi(y,x))$ for one's natural verdict and $\pi(y,x)(1+\beta)$ for the other's verdict. The unconditional density function f is symmetric, unbounded, and log-concave with bounded hazard rate.[5] We can parameterize the density f_λ by a parameter λ, where $f_\lambda(x')/f_\lambda(x)$ decreases in λ with limit zero as $\lambda \to \infty$ for all $x' > x \geq 0$. So information depreciates as λ rises.

[2] We show that this communicative equilibrium is welfare maximizing. It realizes the goals of William Penn, who wrote "In all debates, let truth be thy aim, not victory, or an unjust interest."

[3] This rare failure of upper hemicontinuity owes to the lack of discounting.

[4] Precisely, Bayes rule yields $\pi(y,x) = e^y/(e^x+e^y) = e^{y-x}/(1+e^{y-x})$. So $\pi(x,x) \equiv 1/2$ for all x.

[5] These assumptions are often met — e.g., in a well-studied case where jurors' types arise from uniform probabilities on $[0,1]$, for then their log-likelihood ratio types have density $f(x) = e^x/(1+e^x)^2$.

3. Equilibrium Analysis. Moritz' initial vote fixes the debate roles for the rest of the game. If he initially proposes $\mathcal{A}$, then they enter the *natural subgame* where each juror argues for his natural verdict until conceding; otherwise, they enter the *Nixon in China* (NiC) subgame where jurors argue against their natural verdicts until conceding.

In either subgame, best responses are monotone. Agreeable, sincere best response profiles[6] are thus characterized by cutoff vectors $(x_t)_{t\in\mathbb{Z}}$. In the strategic dynamic programming exercise, there are (a) *communicative* equilibria, where one perseveres until *convinced*, i.e. confident that conceding is better than prolonging the debate two periods, and (b) *deferential equilibria*, where one quits earlier, anticipating permanent intransigence. But with our preference bias, we now must carefully distinguish between play in the two subgames. A cutoff vector (x_t) is *τ-deferential in the natural subgame* if $x_t < \infty$ for all $t < \tau$ and $x_\tau = \infty$, or *communicative in the natural subgame* if $x_t < \infty$ for all $t \geq 0$. Likewise, it is *τ-deferential in the NiC subgame* if $x_{-t} > -\infty$ for all $t < \tau$ and $x_{-\tau} = -\infty$, and *communicative in the NiC subgame* if $x_{-t} > -\infty$ for all $t \geq 0$.

Our analysis bipasses Bellman values, and jumps to the optimality conditions. In the natural subgame, if one's colleague's types $x \leq \underline{x}$ have already conceded and types $x \in [\underline{x}, \bar{x}]$ will next concede, then type y's *propensity* (net payoff) to hold out equals:[7]

$$P(\underline{x}, y, \bar{x}) = \int_{\underline{x}}^{\bar{x}} ((2+\beta)\pi(y,x) - 1 - \kappa) f(x|y)dx - \int_{\bar{x}}^{\infty} 2\kappa f(x|y)dx. \tag{1}$$

The first integrand captures y's *net* payoff $(2+\beta)\pi(y,x) - 1$ of his natural verdict when colleague types $x \in [\underline{x}, \bar{x}]$ concede, minus a one period waiting cost. This naturally falls in the colleague's type x and crosses zero at $y + b$, where $b \equiv \log((1+\beta-\kappa)/(1+\kappa))$. So $b > 0$ iff the jurors' bias is sufficiently large relative to the waiting costs, $\beta \geq 2\kappa$. The second integral reflects the two period waiting cost when types $x > \bar{x}$ hold out. Figure 1 plots this integrand in x, and indicates two roots of the propensity function.

While (1) refers to type y's propensity to hold out, we study the behavior of P a function of the third argument $\bar{x}$. Its derivative $((2+\beta)\pi(y,\bar{x}) - 1 + \kappa) f(\bar{x}, y)$ crosses zero once and from above at $\bar{x} = y + B$, where $B := \log((1+\kappa+\beta)/(1-\kappa))$. P is thus hump-shaped in $\bar{x}$, as seen in Figure 2. Since (1) is clearly initially negative for $\bar{x}$ near $\underline{x}$, it generally admits no roots, or two roots, an *inner* root $\bar{x} = \iota(\underline{x}, y) < y + B$ and an *outer* root $\bar{x} = \omega(\underline{x}, y) > y + B$, as illustrated in Figures 1 and 2.

In equilibrium, any finite cutoff type x_t must be indifferent between holding out and

[6] So we ignore contrarian equilibria, where Moritz insincerely votes opposite to his preference.

[7] The propensity functions P^0 and $P^{\mathcal{N}}$ for the initial period and NiC subgame are in the appendix.

conceding

$$P(x_{t-1}, x_t, x_{t+1}) = 0. \tag{2}$$

Type x_t's indifference condition (2) inversely yields a recursion for his colleague's next cutoff — namely, $(x_{t-1}, x_t) \mapsto x_{t+1}$. The inner root $x_{t+1} = \iota(x_{t-1}, x_t)$ balances the benefit from convincing weaker colleagues' types (the yellow area in Figure 1, top panel) with the waiting costs (the pink area). Type x_t wants to prevail over all (yellow) conceding types $x \in [x_{t-1}, x_{t+1}]$ and, for the same reason, if $x_t = \iota(x_{t-2}, x_{t-1})$, all these types conversely wanted to prevail over x_t in the previous period. This is the essence of *turf guarding*. By contrast, for the outer root analysis, $x_{t+1} = \omega(x_{t-1}, x_t)$ lies inside the pink region (bottom panel of Figure 1). Thus, there is an additional decision cost when holding out, since x_t is *ambivalent* about convincing $x \in [x_{t-1}, x_{t+1}]$; this trade-off is familiar from the cheap talk and pivotal voting literature, e.g. LRS2001.

The story changes in the NiC subgame, since it is impossible that Moritz wants to prevail and convict while Lones prefers to prevail and acquit, as turf guarding requires.

Intransigent equilibria generally are unstable, unless individuals are strongly biased:

Theorem 1 (Stability) *Equilibria that are deferential in the NiC subgame violate forward induction. Those deferential in the natural subgame violate forward induction for small $\beta \geq 0$, and satisfy forward induction for large $\beta > 0$.*

Proof Sketch: Assume Lones unexpectedly holds out in period τ from an equilibrium that is τ-deferential in the natural subgame with τ odd and $x_{\tau-1} = x_{\tau+1} = \ldots$. By forward induction, Moritz thinks that Lones' type ℓ is strong enough to support his desire to convict, given that he knows $m \geq x_{\tau-1}$.

For small β, say $\beta \leq 2\kappa$, any types $\ell < x_{\tau-1}$ are unwilling to invest κ to convince any remaining type of Moritz $m \geq x_{\tau-1}$, because $\pi(\ell, m) < 1/2$ and so $(2+\beta)\pi(\ell, m) - 1 < \beta/2 \leq \kappa$. Forced to believe $\ell \geq x_{\tau-1}$, Moritz' weakest remaining type $m = x_{\tau-1}$ is then sufficiently convinced of guilt to relent and concede, undermining the equilibrium.[8]

For large β, type $\ell = x_{\tau-1}$ may be happy to persist, depending on Moritz' future play. For if $x_{\tau+1} = \ell + B$ then (1) implies $P(x_{\tau-1}, \ell, \ell + B) > 0$ for large β, and so large B. Moritz' remaining types $m \geq x_{\tau-1}$ can think that Lones deviated due to his bias β, rather than strong signal ℓ of guilt. So Moritz' can hold out in period $\tau + 1$. □

4. Ambivalent Debate. We first analyze the model for a small *preference bias* $\beta \geq 0$. In this case, turf guarding is impossible if $x_t = \iota(x_{t-2}, x_{t-1}) < x_{t-1} + B$ implies that

[8] Thus, any $\beta \leq 2\kappa$ is "sufficiently small" for τ-deferential equilibria with $x_{\tau-1} = x_{\tau+1} = \ldots$ to violate forward induction. But other τ-deferential equilibria with $x_{\tau-1} < x_{\tau+1}$ may satisfy forward induction when $\beta = 2\kappa$; ruling out all deferential equilibria therefore requires a tighter upper bound on β.

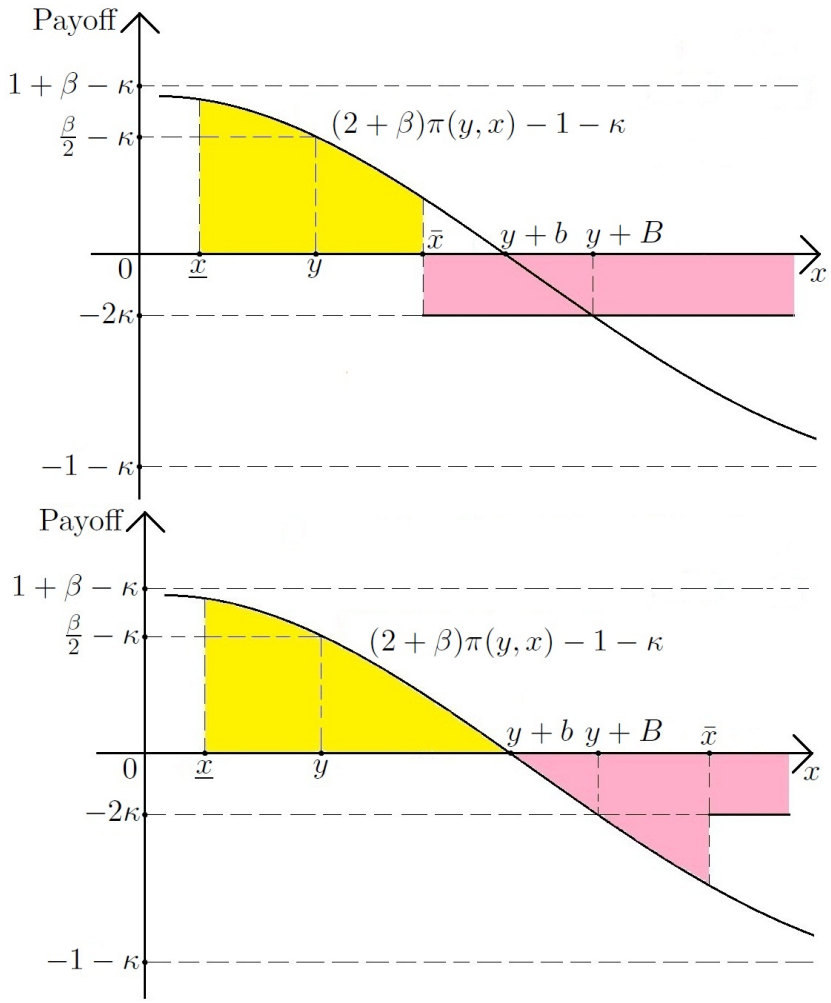

Figure 1: **Inner and Outer Roots: Turf Guarding vs. Ambivalence.** Here, we can see why the propensity function (1) falls in its first argument (yellow area shrinks), rises in its third argument for the inner root (yellow area grows, pink area shrinks) and falls for the outer root (pink area grows).

$P(x_{t-1}, x_t, x_{t+1}) < 0$ for any x_{t+1}, for then type x_t cannot be incentivized to hold out. This propensity falls in its first argument, and so is maximized at $x_{t-1} = x_t - B$ on $[x_t - B, \infty)$; in its third argument, it is maximized at $x_{t+1} = x_t + B$. So to rule out turf guarding, it suffices that this *maximal propensity* be negative for any x_t, namely:

$$P(y - B, y, y + B) < 0 \qquad \text{for all } y. \tag{3}$$

We can show that if condition (3) holds for parameters (β, κ, λ), then it also holds for any parameters $(\beta', \kappa', \lambda')$ with $\beta' \leq \beta$, $\kappa' \geq \kappa$, and $\lambda' \leq \lambda$. Intuitively, the maximal propensity — as seen in Figure 1 for $\underline{x} = y - B$ and $\bar{x} = y + B$ — increases in β (yellow area grows), decreases in κ (yellow area shrinks, pink area grows), and increases in λ

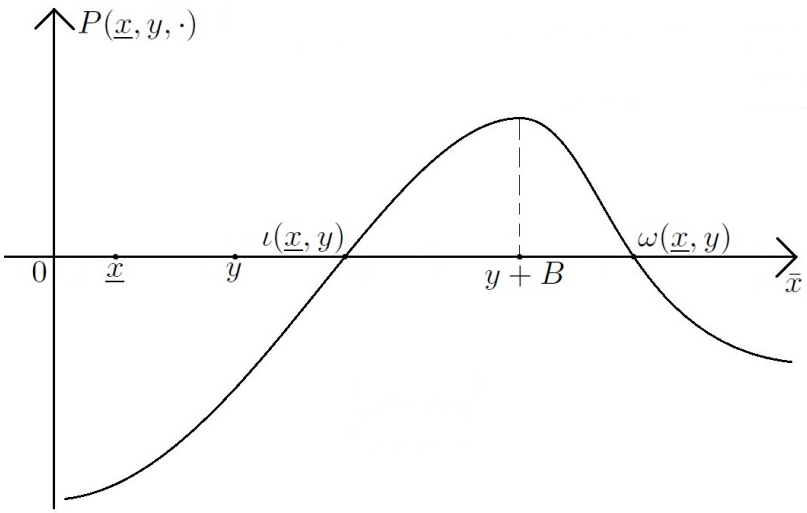

Figure 2: **The Hump-Shaped Propensity Function with Two Roots.** The propensity is rising at ι and falling at λ.

(probability shifts from pink to yellow area, with a higher hazard rate). So we can show:

Lemma 1 (*a*) *A* bifurcation threshold $\beta^*(\kappa, \lambda) > 0$ *exists with* (3) *iff* $\beta \leq \beta^*(\kappa, \lambda)$.
(*b*) $\beta^*(\kappa, \lambda)$ *rises in* κ*, vanishing as* $\kappa \to 0$*, and falls in* λ*, vanishing as* $\lambda \to \infty$.

Theorem 2 (Uniqueness / Existence for a Small Bias) *Assume* $\beta \leq \beta^*$. *There exists a unique communicative equilibrium and, for each stopping profile* (ς, τ)*, a unique deferential equilibrium. All equilibria entail ambivalent debate in all periods.*

By Lemma 1, for any $\beta > 0$, ambivalence also arises for all large $\kappa > 0$ or small $\lambda > 0$. So *if* $\beta > 0$*, when the debate is costly or players are well-informed, debate is ambivalent.*
Proof Sketch: The recursion dynamics (2) are a second-order difference equation, and therefore have two free parameters. Nevertheless, we exploit transversality reasoning, and show that the set of (x_0, x_1) for which (2) has a solution for all $t \geq 0$ is non-empty but degenerate — namely, a curve with a unique *seed* x_1 for every *anchor* x_0. The reason is that the difference equation has a saddle-point property (Appendix B): If another cutoff sequence (x'_t) starts at the same anchor x_0 but has a higher seed x'_1, then $\Delta x_t \equiv x'_t - x_t$ grows exponentially, and thus $x'_{t+1} - x'_t \geq \Delta x_{t+1} - \Delta x_t$ grows exponentially as well. This unbounded "fanning out" is incompatible with the existence of a finite outer root: For then a juror becomes too convinced of his position to think he might change his mind. Geometrically, it would violate the staircase in Figure 3.

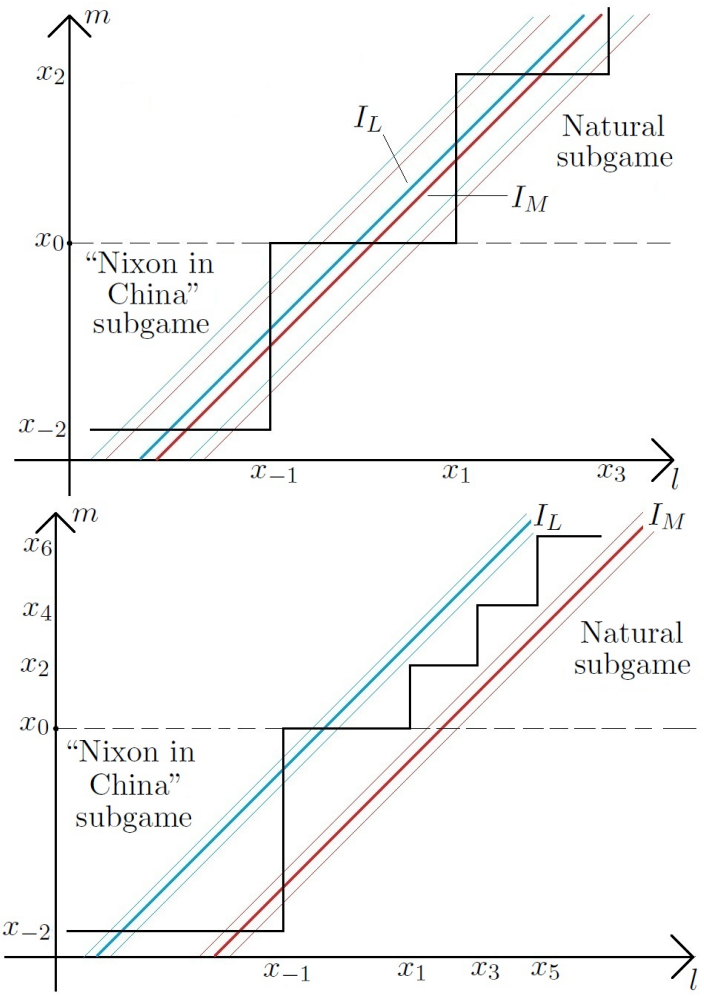

Figure 3: **Equilibrium cutoffs for outer and inner roots.** The three red diagonals correspond to type pairs where Moritz' static net gain to acquitting respectively equals κ, zero, or $-\kappa$ — i.e., the lines $\ell = m + b, \ell = m + \log(1+\beta)$ or $\ell = m + B$. At left, are the analogous blue lines for Lones, namely, $\ell = m - B, \ell = m - \log(1+\beta)$ or $\ell = m - b$. At top, waiting costs dominate the jurors' bias $2\kappa > \beta$, and cutoff vectors zig-zag outside the "disagreement corridor" in the natural subgame (ambivalence). At bottom, jurors' bias dominates the waiting costs $\beta > 2\kappa$, and cutoff vectors zig-zag inside the disagreement corridor in the natural subgame (turf guarding). Both are ambivalent in the NiC subgame.

Likewise for the NiC subgame, the set of (x_{-1}, x_0) such that its recursion, defined by (10), has a solution is a curve. Finally, these curves uniquely connect via the initial indifference condition in the initial period, defined by (9). □

The war of attrition typically entails a great loss in efficiency. By contrast, our conversational war of attrition is welfare better the longer it lasts.

Conjecture 1 (Welfare) *For small $\beta \geq 0$, longer equilibria are Pareto better.*

Indeed, we have proven this result for common interests, i.e. $\beta = 0$. The proof follows by continuity of the unique communicative equilibrium cutoff vector (x_t) in β.

5. Turf Guarding Debate. We now assume a preference bias $\beta > \beta^*$, so that turf guarding is not precluded. In fact, turf guarding must arise in this case.

Conjecture 2 (Existence/Uniqueness for Large Bias) *Assume $\beta > \beta^*$. There exists a unique communicative equilibrium and a unique (ς, τ)-deferential equilibrium, for every (ς, τ). Any communicative equilibrium eventually switches to turf guarding.*

Proof sketch: Since $\beta > \beta^*$, maximal propensity $P(y - B, y, y + B) \geq 0$ for some y. For any anchor $x_{t^*} > y$, we can show that there exists a unique seed $x_{t^*+1} < x_{t^*} + B$ that defines an infinite cut-off vector (x_t) via $x_{t+1} = \iota(x_{t-1}, x_t)$ for all $t > t^*$. As in the case $\beta \leq \beta^*$, the uniqueness of x_{t^*+1} owes to saddle-point stability. Appendix B shows that with a larger seed x'_{t^*+1}, the odd increments Δx_{t^*+2n+1} increase exponentially, but the even increments Δx_{t^*+2n} decrease. □

By Conjecture 2, *turf guarding arises in the communicative equilibrium for all small delay costs $\kappa > 0$.* For the bifurcation threshold $\beta^*(\kappa, \lambda)$ vanishes as $\kappa \to 0$ by Lemma 1(b). Yet the equilibrium disappears when $\kappa = 0$, since that is a dynamic version of LRS2001. They show that any equilibrium entails ex-post inefficient decisions, corresponding to our ambivalent outer roots analysis.[9] There is no sense that zero communication costs approximates small communication costs.[10]

In the same way, the bifurcation threshold $\beta^*(\kappa, \lambda)$ vanishes as $\lambda \to \infty$. So by Conjecture 2, *turf guarding arises in the communicative equilibrium whenever jurors are ill-informed.* More strongly, the hazard rate of the type distribution explodes as $\lambda \to \infty$, and therefore the inner root $\iota(\underline{x}, y)$ converges to $\underline{x}$. Intuitively, a juror's posterior barely budges when his colleague holds out since jurors have little information to share. We converge upon a "shouting match" war of attrition.

6. The Option Value of Debate. Let us now compare our dynamic strategic conversation to a natural embedded decision problem. Consider the *dictator's problem*: how a juror would vote if he were suddenly afforded dictatorial power and asked to choose the verdict unilaterally. Since the dictator pays no deliberation cost, one might think him less willing to concede than our debators. In fact, for small waiting costs and small preference bias, the opposite is true: One may act as a devil's advocate, arguing for a verdict despite

[9] Proposition 4 in LRS2001 asserts that any equilibrium crosses the disagreement zone, the region between the indifference lines. The distinction between static and dynamic voting doesn't matter because the inference from the fact that one's vote is pivotal is the same in the two models.

[10] The closed graph property fails because our game violates continuity at infinity (no discounting).

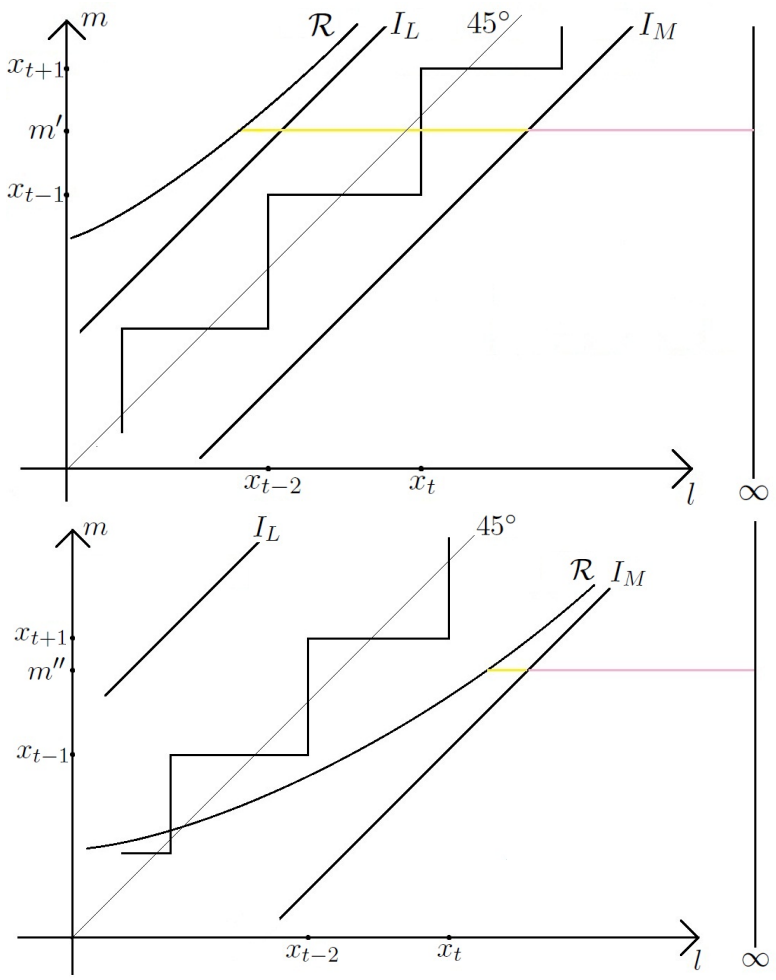

Figure 4: **The Dictator Problem.** Moritz' types m' (m'' in the bottom panel) prefers acquittal when Lones' type ℓ is in the yellow interval and conviction if ℓ is in the pink interval. Conditioning on the event $\ell \geq \rho(m')$, he is indifferent.

no longer believing it. For conceding ends the game, while holding out retains the option value of conceding later based on new information about the other juror's type.

If type y dictator Moritz knows that Lones' type exceeds $\underline{x}$, his net payoff gain to acquitting is:

$$Q(\underline{x}, y) = \int_{\underline{x}}^{\infty} ((2+\beta)\pi(y,x) - 1) f(y|x) dx. \tag{4}$$

In our conversation, jurors retain the option to drop out in the future. Hence, the right signal tail is the option value, namely, learning that one's colleague's type is so strong that he holds out. A larger right tail leads to a larger loss in (1). By contrast, the dictator's payoff is an expected value across all remaining colleague's types.

The integrand in (6) is positive for x less than $y + \log(1+\beta)$, and then negative. This

is illustrated by the yellow and pink intervals in Figure 4. Since $Q(\underline{x}, y)$ strictly falls in $\underline{x}$, it admits at most one root $\rho(y)$, where $Q(\rho(y), y) = 0$. Namely, Moritz is indifferent knowing that Lones' type exceeds $\rho(y)$.[11] The locus $\mathcal{R}$ of such pairs $(\rho(y), y)$ is left of Moritz' indifference line I_M, since the posterior given $\ell \in (\rho(y), \infty)$ exceeds his belief given $\rho(y)$ (see Figure 4). Hence, $\rho(y) < y + \log(1+\beta)$.

Given equilibrium cutoffs (x_t), type $y \in [x_{t-1}, x_{t+1}]$ is a *devil's advocate* if $\rho(y) \leq x_{t-2}$. For then, y sticks to his own verdict in period $t-1$ even though he would implement his colleague's verdict if he had dictatorial power — for he knows that his colleague just held out, and thus $x \geq x_{t-2} \geq \rho(y)$. In the top panel of Figure 4, type m' is a devil's advocate. In the bottom panel, type m'' is not; he concedes in period $t+1$, even though as a dictator he would still acquit conditional on $\ell > x_t$.

Altogether, we see that a devil's advocate can only arise if $\mathcal{R}$ is sufficiently left of the equilibrium cutoff vector (x_t). For small $\kappa > 0$, this zig-zag tightly straddles the main diagonal, as seen in the bottom panel of Figure 3. We therefore compare $\mathcal{R}$ to the diagonal. We can show that $\mathcal{R}$ lies left of the main diagonal iff $\beta \leq \beta^\dagger$, for some threshold $\beta^\dagger > 0$.[12] For any preference bias $\beta \leq \beta^\dagger$, type y would agree with his colleague's verdict when he learns that he holds superior information, $x \geq y$. For any larger preference bias $\beta > \beta^\dagger$, type y would not concede in this event.

Theorem 3 (Devil's Advocacy) *For all $\beta \leq \beta^\dagger$ and small $\kappa > 0$, all strong enough types are devil's advocates in the communicative equilibrium. For all $\beta > \beta^\dagger$ and small $\kappa > 0$, strong enough types are not devil's advocates in the communicative equilibrium.*

7. How Long Does Debate Last? In LRS2001, as in Crawford and Sobel (1982), greater conflicts of interest resulted in less refined communication. In our dynamic model with communication costs, preference bias arguably has the opposite effect — prolonging debates and refining communication. Fixing a cutoff vector (x_t), jurors' propensity to hold out increases in β. But then the jurors best respond by lowering their cutoff types, conceding more slowly. Conflict forces individuals to better articulate their positions. The resulting communication is in fact more refined, unlike the coarsened communication in LRS2001. At the opposite extreme, if jurors share the same preferences they reach agreement sooner.

[11] For $\underline{y} \equiv -\log(1+\beta)$ we have $\rho(y) = -\infty$: With interim belief $\underline{y}$ and no information about Lones' type, Moritz is indifferent between acquittal and conviction as $(2+\beta)e^{\underline{y}}/(1+e^{\underline{y}}) = 1$. This implies that the root $\rho(y)$ exists for all $y \geq \underline{y}$.

[12] We can show that $\rho(y) - y$ increases continuously but is bounded above by $\log(1+\beta)$, thus admitting a finite right limit $\lim_{y\to\infty}(\rho(y) - y)$. We can show that this limit increases in β, crossing zero once, at $\beta^\dagger$.

Appendix

A. Propensity Functions. In the NiC subgame, if colleague types $x > \bar{x}$ have conceded, and types $x \in [\underline{x}, \bar{x}]$ will next concede, then y's propensity to hold out is

$$P^{\mathcal{N}}(\underline{x}, y, \bar{x}) = \int_{\underline{x}}^{\bar{x}} (1 - (2+\beta)\pi(y,x) - \kappa) f(x|y) dx - \int_{-\infty}^{\underline{x}} 2\kappa f(x|y) dx. \tag{5}$$

In the initial period, given Lones' period one cutoffs x_{-1} and x_1, Moritz' propensity to vote $\mathcal{A}$ equals

$$P^0(x_{-1}, y, x_1) = \int_{-\infty}^{x_{-1}} \kappa f(x|y) dx + \int_{x_{-1}}^{x_1} ((2+\beta)\pi(y,x) - 1) f(x|y) dx - \int_{x_1}^{\infty} \kappa f(x|y) dx \tag{6}$$

A cutoff vector (x_t) constitutes an agreeable, sincere equilibrium iff it obeys (2), and:

$$P^0(x_{-1}, x_0, x_1) = 0 \qquad \text{if } x_0 \text{ is finite,} \tag{7}$$

$$P^{\mathcal{N}}(x_{-t-1}, x_{-t}, x_{-t+1}) = 0 \qquad \text{for all } t \text{ with } x_{-t} \text{ finite.} \tag{8}$$

B. Shooting Function Stability. Assume below that $\underline{x}, y, \bar{x}$ solve $P(\underline{x}, y, \bar{x}) = 0$, i.e. $\bar{x} = \iota(\underline{x}, y)$ or $\bar{x} = \omega(\underline{x}, y)$. The caption to Figure 1 justifies the signs of the partial derivatives, namely, $\partial_1 P(\underline{x}, y, \bar{x}) < 0 < \partial_2 P(\underline{x}, y, \bar{x})$ at both roots, while $\partial_3 P(\underline{x}, y, \iota(\underline{x}, y)) > 0 > \partial_3 P(\underline{x}, y, \omega(\underline{x}, y))$. We can also sign key directional derivatives of P: Lemma 4(b) in our original paper proved that $(\partial_1 + \partial_2 + \partial_3) P(\underline{x}, y, \bar{x}) > 0$. We can also show that $|\partial_1 P(\underline{x}, y, \bar{x})| > (1+\varepsilon)|\partial_3 P(\underline{x}, y, \bar{x})|$ on the relevant parameter range.[13]

Outer Shooting Function. The Implicit Function Theorem then yields outer shooting function partial derivatives:

$$\lambda_1 = -\frac{\partial_1 P}{\partial_3 P} < -(1+\varepsilon) \quad \text{and} \quad \lambda_1 + \lambda_2 = -\frac{(\partial_1 + \partial_2) P}{\partial_3 P} = -\frac{(\partial_1 + \partial_2 + \partial_3) P}{\partial_3 P} + 1 > 1$$

Assume that there are two cut-off sequence $(x_t), (x'_t)$ obeying $\omega(x_{t-1}, x_t) = x_{t+1}$ and $\omega(x'_{t-1}, x'_t) = x'_{t+1}$ for all $t > 1$. Define the operator $\Delta x_t \equiv x'_t - x_t$ and assume $0 = \Delta x_0 < \Delta x_1$. Then, the Mean Value Theorem yields:

$$\begin{aligned} \Delta x_{t+1} - \Delta x_t &= \lambda_1 \Delta x_{t-1} + \lambda_2 \Delta x_t - \Delta x_t = \lambda_1 (\Delta x_{t-1} - \Delta x_t) + (\lambda_1 + \lambda_2 - 1) \Delta x_t \\ &> (1+\varepsilon)(\Delta x_t - \Delta x_{t-1}). \end{aligned}$$

[13] For the outer root $\bar{x} = \omega(\underline{x}, y)$ compare to Lemma 6(b) in our original paper. For the inner root, $\bar{x} = \iota(\underline{x}, y)$, we can show this for large enough $\underline{x}, y$.

Inductively, $\Delta x_{t+1} - \Delta x_t > (1+\varepsilon)^t \Delta x_1$ explodes, thus implying a unique ambivalent communicative equilibrium. This fanning out reflects strategic complementarity: x_t's propensity to hold out decreases as his colleague concedes faster and x_{t-1}, x_{t+1} rise.

Inner Shooting Function. Similarly for the inner shooting function:

$$\iota_1 = -\frac{\partial_1 P}{\partial_3 P} = 1 - \frac{(\partial_1 + \partial_3)P}{\partial_3 P} > 1 \qquad \text{and} \qquad \iota_2 = -\frac{\partial_2 P}{\partial_3 P} < 0.$$

Then, if $(x_t), (x'_t)$ obey $\iota(x_{t-1}, x_t) = x_{t+1}$ and $\iota(x'_{t-1}, x'_t) = x'_{t+1}$ for all $t > t^*$ with $0 = \Delta x_{t^*} < \Delta x_{t^*+1}$ the dynamical system obeys:

$$\Delta x_{t^*+2} = \iota_1 \Delta x_{t^*} + \iota_2 \Delta x_{t^*+1} < 0 \quad \text{and} \quad \Delta x_{t^*+3} = \iota_1 \Delta x_{t^*+1} + \iota_2 \Delta x_{t^*+2} > (1+\varepsilon)\Delta x_{t^*+1}.$$

Induction yields $\Delta x_{t^*+2n} < 0$ and $\Delta x_{t^*+2n+1} > (1+\varepsilon)^n \Delta x_{t^*+1} \to \infty$, implying uniqueness of the communicative turf guarding equilibrium. This sign alternation reflects that actions are "strategic substitutes": Type x_t's propensity to hold out increases when his colleague concedes faster (i.e. x_{t+1} increases).