B Appendix: Extensions

B.1 Contrarian Contradictory Debate

Our analysis has focused on sincere agreeable equilibria. Now assume that Lones is not sincere, but *contrarian*: he initially proposes to convict if his signal indicates innocence, say $\ell < x_0 = 0$, and acquit if his signal indicates guilt. To see that contrarianism can arise in equilibrium, suppose further that Lones defers to Moritz in period two.[1] Thus, Moritz calls the verdict in period one and must be careful to second Lones only if his type strongly supports this verdict; for after all, a proposal to acquit by Lones indicates guilt. Formally, if κ_M is small, Moritz' first period cutoffs must satisfy $x_1 < x_{-1}$, as illustrated in Figure 1(a). Since types $m \in [x_1, x_{-1}]$ contradict either one of Lones' initial proposals, Moritz is not agreeable but *contradictory*. Anticipating this response by Moritz, Lones' contrarian strategy is indeed optimal.

A discordance between the literal semantics of a proposal like "Acquit" and its equilibrium interpretation is common in the cheap-talk literature Crawford and Sobel (1982). In our game, this reversal can only arise in period zero, since that is the unique period when the debate is not dispositive. But even in period zero, Lones' equilibrium choice of arguments is not arbitrary: a sincere strategy allows Moritz to agree and reach a verdict in period one in the likely case that the signals indicate the same verdict; the contrarian strategy forces Moritz to contradict Lones in period one and only reach the verdict in period two. Thus contrarianism increases delay costs.

B.2 Intransigence as the Continuous Time Limit of Small Delay Costs

In §5.2, we decomposed delay costs $\kappa_i = k_i\eta$ into juror i's flow waiting cost $k_i > 0$ and the real time period length $\eta > 0$ and studied debate for small η. We now take this idea to its intuitive limit by directly considering a continuous time version of our model. In line with our insights in §5.2, we show that continuous time rules out ambivalence and Nixon-China debates, and that the equilibrium predictions for intransigent debate carry over.

Best-response strategies are monotone by Lemma 1,[2] and can thus be characterized by (weakly) increasing cutoff functions (ℓ_t) and (m_t) that represent the strongest types holding out until time t in natural debate. To characterize these cutoff functions, let

$$\tilde{\Pi}_i(x, y, \dot{x}) \equiv \Delta(x - y, \beta_i)\dot{x} f(x|y)/(1 - F(x|y)) \tag{1}$$

be the propensity to hold out for type y of juror i when his colleague concedes at rate $\dot{x}$, and types below x have already conceded. As in §3.3, we call (ℓ_t, m_t) *communicative* if $\ell_t, m_t < \infty$

[1] There may exist other deferential contrarian equilibria where conversation continues past $t = 2$. Lones' behavior is contrarian only in period zero, but later is monotone, by Lemma 1.

[2] This lemma extends readily to continuous time, as do most other preliminary results in §3.

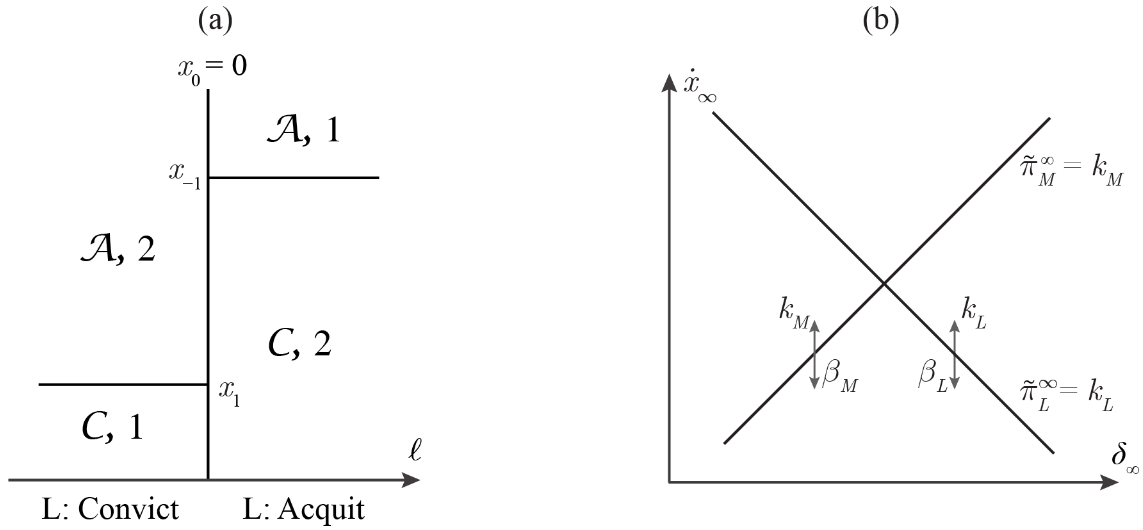

Figure 1: **Panel (a): Outcomes of a Contrarian Contradictory Equilibrium.** Contrarian Lones proposes to convict if his type indicates innocence, and to acquit if it indicates guilt. Weak types of Moritz $m \in [x_1, x_{-1}]$ contradict either one of Lones' initial proposals. **Panel (b): Asymptotic Indifference Curves in Continuous Time.** The limit gap δ_∞ corresponds to δ_{EI} in Figure 7, while the speed limit $\dot{x}_\infty$ corresponds to $\lim_{\eta\to 0}(\delta_{EI} + \delta_{MJ})/\eta$. The curves cross once and share the comparative statics of intransigent debate in Figure 7(b).

for all t.[3] A communicative equilibrium $(\ell_t), (m_t)$ is then characterized by

$$\tilde{\Pi}_M(\ell_t, m_t, \dot{\ell}_t) = k_m \qquad \text{and} \qquad \tilde{\Pi}_L(m_t, \ell_t, \dot{m}_t) = k_L. \tag{2}$$

Clearly, the decision payoff gain in (1) must be positive, i.e. $\Delta(\ell_t - m_t, \beta_M) > 0$ and $\Delta(m_t - \ell_t, \beta_M) > 0$, or equivalently $\ell_t - m_t < b_M$ and $m_t - \ell_t < b_L$. Thus, equilibrium debate must be intransigent; ambivalent debate with its large ex-post decision errors can only arise with the frictions imposed by discrete time debating.

The propensity is asymptotically stationary with limit $\tilde{\pi}_i^\infty(\delta, \dot{x}) \equiv \lim_{y\to\infty} \tilde{\Pi}_i(y - \delta, y, \dot{x})$, as in §3.5. The speeds of debate $\dot{\ell}_t, \dot{m}_t$ likewise converges to the same *limit speed* $\dot{x}_\infty$, while Lones' cutoff gap $\delta_t \equiv \ell_t - m_t$ converges to a constant *limit gap* $\delta_\infty \equiv \lim \delta_t$. In equilibrium, the speed limit and limit gap are uniquely determined by the jurors' *asymptotic indifference curves*, $\tilde{\pi}_L^\infty(\delta_\infty, \dot{x}_\infty) = 0$ and $\tilde{\pi}_M^\infty(-\delta_\infty, \dot{x}_\infty) = 0$, seen in Figure 1(b). Reading $\dot{x}_\infty$ as a function of δ_∞, Lones' curve decreases since $\tilde{\pi}_L$ rises in both of its arguments, while Moritz' curve increases since $\tilde{\pi}_M$ falls in δ_∞. As either juror grows more patient or biased, his curve shifts down, and the limit speed falls, as in Proposition 6.

Errors of impunity (miscarriages of justice) are positive when the cutoff gap δ_t is positive (negative). In the limit, δ_∞ $(-\delta_\infty)$ represents the eventual error of impunity (miscarriage of

[3] There may also exist deferential equilibria here. Suppose all remaining types of one juror, say Lones, dig in their heels at time t; Moritz' remaining types $m > m_t$ then concede immediately. The indifference condition for Lones' last conceding type ℓ_t then requires $\int_{m_t}^{\infty} \Delta(m - \ell_t, \beta_L) f(m|\ell_t, m \geq m_t) dm = 0$.

justice); clearly, only one of them can be positive. By Figure 1(b), eventual errors of impunity fall and eventual miscarriages of justice rise as Lones grows more biased/patient, and conversely for Moritz. This corroborates Proposition 5 for intransigent debate.

Nixon-China debate must end immediately in continuous time. Assume otherwise, that debate transpires until some $t > 0$, at which some types ℓ_{-t} and m_{-t} simultaneously concede. Given the cost of delay, hawk Lones must strictly prefer acquittal conditional on (ℓ_{-t}, m_{-t}), while dove Moritz strictly prefers conviction. This is impossible.[4]

Finally, consider Lones' behavior at $t = 0$. We argue that $\ell_0 = -\infty$, i.e. all types of Lones propose to convict; thus, Nixon-China debate never arises. Otherwise, if ℓ_0 is finite, this marginal type is indifferent between acquit and convict. If ℓ_0 proposes to acquit, all types of Moritz agree at once. If he proposes to convict, weak types of Moritz, $m < m_0$ concede to convict, but strong types $m > m_0$ hold out, yielding an acquittal when ℓ_0 concedes at once afterwards. Thus, ℓ_0 must be indifferent between the verdicts conditional on $m < m_0$. But his indifference condition in the natural debate requires that he strictly prefer to convict conditional on the event $m = m_0$, which is less favorable for guilt than the event $m < m_0$. This is a contradiction. All told, a communicative equilibrium in the continuous time game is characterized by cutoff functions (ℓ_t) and (m_t) that satisfy the ODE (2) together with the transversality condition $\lim_{t\to 0} \ell_t = \lim_{t\to 0} m_t = -\infty$.[5,6]

B.3 An Outsider's Perspective on Debate

To capture the ebb and flow of the debate, consider the belief of an uninformed observer, for whom debate is a form of social learning. For instance, suppose President Obama is initially unaware of intelligence, and is watching two informed members of his security council argue whether to get bin Laden. A hawk favors action and a dove favors inaction. Should the president worry about acting on the latest twist in the debate, aware that either party may simply be playing the devil's advocate? He must predicate his behavior on the *public posterior* — i.e., the guilt chance given the history and equilibrium strategies.

In informational herding models of social learning, the public belief process is a martingale that always favors the action just taken. This need not hold in our debate. Firstly, the public posterior moves in the direction of the last argument. After the hawk has just argued for action in period t, it conditions on $\ell \geq x_t$ and $m \geq x_{t-1}$, but when the dove next argues for inaction,

[4] With unbiased jurors, this logic applies to both subgames. Equilibrium is then unique, with all types of Moritz conceding at once to Lones. So purely informational debate is not possible in continuous time.

[5] We do not prove existence of such an equilibrium here; by footnote 4, existence requires at least one juror to be biased. In proving existence, it seems useful to revert the (unbounded) log-likelihood ratio types ℓ, m to probabilities via $\lambda \equiv e^{\lambda}/(1 + e^{\lambda})$ and $\mu \equiv 1/(1 + e^{m})$, in order to transform the transversality conditions into boundary conditions $\lambda_0 = 0$ and $\mu_0 = 1$.

[6] We do not attempt to show that the discrete time equilibria converge to continuous time equilibria as the period length η vanishes. However, we argued in §5.2 and prove in §A.9 that discrete time equilibria for small $\eta > 0$ share many salient features of the continuous time equilibria discussed here: Namely, communicative natural debate is intransigent (after a short initial phase), the set of conceding types vanishes for small $\eta > 0$, and the initial cutoff x_0 tends to $-\infty$.

it conditions on $m \geq x_{t+1}$, and thus moves in favor of inaction.

For short period lengths $\eta > 0$, a sharper result is possible. With symmetric types, the public belief after the hawk has argued for action in period $2t$ favors action iff $x_{2t} > x_{2t-1}$, and it favors inaction after the dove's period $2t+1$ argument iff $x_{2t+1} > x_{2t}$, i.e. if the cutoff staircase in Figure 5 straddles the 45-degree diagonal. For symmetric jurors, i.e. $\beta_L = \beta_M$ and $\kappa_L = \kappa_M$, these inequalities follow from Lemma A.5(c).

But now consider a more extreme (or more patient) hawk. As β_L increases, he pushes harder for action while the dove — who eventually resorts to intransigence — softens his stance. The cutoff staircase in Figure 5 eventually shifts to the left of the 45-degree line, i.e. $x_{2t} < x_{2t-1}$; formally this follows from Lemma A.5(d). So even when the hawk had the last word, the president puts greater weight on the earlier proposal to acquit by the more moderate dove. While the public posterior is still a martingale, it only moves in the direction of the last proposal, but always favors inaction as long as debate continues.[7]

B.4 Informativeness of Jurors' Signals

Propositions 4-6 analyze delay and decision errors as a function of jurors' patience and bias. It is natural to ask how changing the informativeness of jurors' signals impacts our equilibrium. A juror i grows more informed if his unconditional type distribution f shifts out in the MLRP sense; in particular, the inverse limit hazard rate γ increases. Akin to the rise in waiting costs considered in Proposition 4, we can show for unbiased jurors that cutoff types shift out as jurors grow more informed. But this no longer implies that debate speeds up, for now there is more probability mass in the signal tails. To see why this tradeoff cannot be easily signed, consider two extreme cases. Jurors should intuitively agree immediately if each is perfectly informed or if each lacks any information, because debate is costly and offers no gain. So the speed of debate is non-monotonic in the informativeness of jurors' types, and we can make no predictions.

Turning to decision errors considered in Proposition 5, we can show that as, say, Lones grows more informed, the eventual miscarriage of justice δ_{MJ} increases, whereas δ_{EI} is U-shaped, decreasing when Lones is intransigent ($\delta_{MJ} < \bar{b}_L$) and increasing when Lones is ambivalent. Thus, surprisingly both errors eventually *rise* as jurors grow more informed. But this paradox is due to our definition of decision errors with reference to the optimal verdict conditional on (ℓ, m). The true decision costs must account for the expected loss of this baseline, which clearly falls as jurors grow more informed. Loosely, as jurors grow more informed, the decision is farther from a more ambitious goal post.

[7] More strongly, it may even favor acquittal after the dove concedes, i.e. conditional on $\ell > x_{2t}$ and $m \in [x_{2t-1}, x_{2t+1}]$ if β_L sufficiently exceeds β_M and waiting costs are small.