Economics Working Paper Series
Faculty of Business, Economics and Law, AUT

Feddersen and Pesendorfer meet Ellsberg

Matthew Ryan

2019/07

Feddersen and Pesendorfer meet Ellsberg

Matthew Ryan*
School of Economics
Auckland University of Technology

This version: September 2019

Abstract

The Condorcet Jury Theorem formalises the "wisdom of crowds": binary decisions made by majority vote are asymptotically correct as the number of voters tends to infinity. This classical result assumes like-minded, expected utility maximising voters who all share a common prior belief about the right decision. Ellis (2016) shows that when voters have *ambiguous* prior beliefs – a (closed, convex) *set* of priors – and follow *maxmin expected utility (MEU)*, such wisdom requires that voters' beliefs satisfy a "disjoint posteriors" condition: different private signals lead to posterior sets with disjoint interiors. Both the original theorem and Ellis's generalisation assume symmetric penalties for wrong decisions. If, as in the jury context, errors attract asymmetric penalities, then it is natural to consider voting rules that raise the hurdle for the decision carrying the heavier penalty for error (such as conviction in jury trials). In a classical model, Feddersen and Pesendorfer (1998) have shown that, paradoxically, raising this hurdle may actually increase the likelihood of the more serious error. In particular, crowds are not wise under the unanimity rule: the probability of the more serious error does not vanish as the crowd size tends to infinity. We show that this "Jury Paradox" persists in the presence of ambiguity, whether or not juror beliefs satisfy Ellis's "disjoint posteriors" condition. We also characterise the strictly mixed equilibria of this model and study their properties. Such equilibria cannot exist in the absence of ambiguity but may exist for arbitrarily large jury size when ambiguity is present. In addition to "uninformative" strictly mixed equilibria, analogous to those exhibited by Ellis (2016), there may also exist strictly mixed equilibria which are "informative" about voter signals.

*My thanks to Andrew Ellis, Simona Fabrizi, Steffen Lippert, Aniol Llorente-Saguer and Addison Pan for helpful comments on previous drafts. Financial support from the Royal Society of New Zealand through Marsden Fund grant UOA1617 is gratefully acknowledged.

1 Introduction

McLennan (1998) distinguishes two types of Condorcet Jury Theorem (CJT). In McLennan's typology, a CJT of the *Second Type* asserts that, under certain conditions, majority voting by privately but imperfectly informed jurors produces a correct decision with probability approaching unity as the jury size increases. Such theorems provide one possible formalisation of the "wisdom of crowds".

Condorcet assumed "sincere" voting – each juror votes as if deciding the outcome unilaterally – but a CJT of the Second Type also obtains when jurors vote strategically, with attention confined to *symmetric Bayesian Nash equilibrium* (henceforth, "equilibrium") outcomes: see Austen-Smith and Banks (1996) and McLennan (1998).[1] More recently, Ellis (2016) proved a version of Condorcet's result when jurors have ambiguous prior beliefs about the defendant's guilt, but are not too imperfectly informed. For reasons to be explained shortly, this generalisation is non-trivial.

In the canonical jury model, jurors share a common prior probability on the defendant's innocence. They cast their votes without consultation,[2] after observing private signals drawn independently from a known state-conditional distribution, with the state being "guilty" or "innocent". Jurors also share the common objective of maximising the expected probability of a correct decision (i.e., convict if guilty and acquit if innocent). In Ellis's model, the common prior is replaced by a common *set* of priors; there is a common *interval* of probabilities on the innocent state. The canonical model is obtained when this interval is a singleton. After observing their private signals, jurors update each prior in the set and vote so as to maximise the *minimum* (over posteriors) of the expected probability of a correct decision.

Relative to the standard model, the introduction of ambiguity alters the formal analysis in two fundamental respects: first, the voting decision of an individual juror can no longer be determined by conditioning on her vote being pivotal; second, a voter's best response may necessitate randomisation. These features substantially complicate the analysis and Ellis does not give a complete characterisation of the equilibria of his model. However, he does prove some interesting general properties of these equilibria.

When the posterior probability interval for innocence contains a neighbourhood of $\frac{1}{2}$ following any signal – that is, when jurors "lack confidence" in their information (Ellis, 2016, Definition 4) – there exists an equilibrium in which each juror casts either vote with equal probability (Ellis, 2016, Proposition 1). Voting is completely uninformative about the jurors' signals. For jurors who lack confidence, such voting is both sincere and a *strict*

[1] McLennan (1998) shows that *ex ante* optimal equilibria may be non-symmetric, but most of the subsequent literature focusses on the symmetric case.

[2] For this reason, the jury analogy is somewhat misleading. It does, however, have the advantage of being memorable; it also nicely motivates the asymmetric loss functions that underpin the "Jury Paradox" to be discussed shortly.

best response. Ellis (2016, Theorem 1) further shows that if jurors lack confidence, then in *any* equilibrium jurors vote for the incorrect decision at least as often as for the correct one, conditional on either state. These results apply to juries of any size: ambiguity and lack of confidence completely undermine the "wisdom" of the crowd.

Conversely, *provided the posterior intervals for different signals have disjoint interiors*, then signals can be unambiguously ordered by the implied likelihood of innocence and a Condorcet-like Jury Theorem holds: as the number of jurors approaches infinity, *there exists* a sequence of equilibria along which the probability of a correct decision converges to unity (Ellis, 2016, Theorem 2). Note the two italicised caveats.

The results of Condorcet and Ellis assume that jurors apply equal utility penalties to each type of error: convicting the innocent and acquitting the guilty. However, Blackstone's maxim exhorts us to guard against the former error more strenuously than we guard against the latter, and therefore to apply a higher utility penalty to conviction of the innocent than to acquittal of the guilty. If we follow Blackstone's maxim then it also is natural to consider raising the voting hurdle for conviction; to trade off some decision accuracy for a reduced likelihood of the more grievous error. Surprisingly, however, Feddersen and Pesendorfer (1998) show that such a trade-off may be illusory. Raising the conviction hurdle may sometimes *increase* the likelihood of convicting an innocent defendant; even more paradoxically, the probability of convicting the innocent always remains asymptotically bounded away from zero under the unanimity rule.[3] We call this latter result the *Jury Paradox*. It stands in stark contrast to Condorcet's theorem.

In this paper we re-visit the Jury Paradox in the presence of ambiguity. We analyse the equilibria of Ellis's model when conviction requires unanimity rather than a simple majority of guilty votes. To bring Ellis's model in line with the framework of Feddersen and Pesendorfer, we make two other adjustments. First, we generalise Ellis's juror utility function so that convicting the innocent may attract a higher utility penalty than acquitting the guilty. Second, we specialise Ellis's information structure by assuming only two possible signal realisations, with state-independent signal distributions.

Feddersen and Pesendorfer's (1998) model is a special case of ours in which prior ambiguity vanishes (the prior interval is a singleton). We prove that the Jury Paradox persists in the more general model: the equilibrium probability of convicting an innocent defendant is strictly bounded away from zero independently of the jury size: see Theorem 5.1 below. This is the main result of the paper. Our analogue of Feddersen and Pesendorfer's paradox does not require either of the caveats in Ellis's version of Condorcet's theorem.

Along the way to proving our main result, we also characterise the *strictly mixed* equilibria of our model – the equilibria in which jurors randomise following *either* signal. Such equilibria are of particular interest since they cannot exist in the absence of ambiguity. In

[3] Provided, that is, one excludes the trivial equilibrium in which all jurors vote for acquittal irrespective of their private information.

our model, there may even be multiple strictly mixed equilibria. There may exist a strictly mixed equilibrium with uninformative (or, in the language of Feddersen and Pesendorfer, "non-responsive") voting, just as there is under the majority rule (Ellis, 2016, Proposition 1), and this equilibrium may co-exist with another strictly mixed equilibrium in which votes are informative ("responsive"): jurors randomise differently for different signals. Importantly for our purposes, responsive strictly mixed equilibria may exist asymptotically – for arbitrarily large jury size – so such equilibria cannot be ignored when proving the Jury Paradox.

The next section introduces our model: a slight modification of Ellis's which nests that of Feddersen and Pesendorfer as a special case. Section 3 describes the best response (to symmetric profiles) correspondence of a voter in our model, summarised by Figure 1. The latter figure is an important guide for the reader through the subsequent analysis. Section 4 characterises a range of equilibria of the model. It is not an exhaustive stocktake but does describe all of the strictly mixed equilibria, which come in a variety of forms; Figure 2 provides an overview of conditions for the existence of these equilibria. Our main result (the Jury Paradox) is contained in Section 5 and Section 6 concludes. Several appendices contain longer proofs and other technical details that would otherwise disrupt the flow of the text.

2 The model

Our basic notation is based on that of Ellis (2016), suitably adapted to our purpose, but we use the language of jury trials from Feddersen and Pesendorfer (1998). The reader is referred to these papers for further discussion of the model.

A set $I = \{1, 2, ..., N+1\}$ of jurors, with generic member i, makes a decision $d \in D = \{A, B\}$ by secret ballot.[4] Ellis (2016) requires N to be even since he focusses on majority rule. We focus instead on the *unanimity* rule, so N may be even or odd here, provided $N \geq 1$.[5] We interpret A as the decision to "acquit" the defendant; hence B corresponds to entering a conviction. We use the same notation for decisions and votes: each juror may vote A for acquittal (the "innocent" vote) or B for conviction (the "guilty" vote). Votes determine the decision via the unanimity rule: the defendant is acquitted – decision $d = A$ is made – unless *all* jurors vote for conviction, in which case decision $d = B$ is made.

The defendant may be innocent or guilty, represented by the state $s \in S = \{a, b\}$, where $s = a$ is the state of innocence and $s = b$ the state of guilt. (Think of b as the state in which the defendant is "bad".) Jurors share common ambiguous prior information about the state. The prior probability of $s = a$, denoted p, is commonly known to lie in the interval $[\underline{p}, \overline{p}] \subseteq (0, 1)$ but nothing more than this. Note that ambiguous beliefs are

[4]In Ellis the voters are selecting a candidate so he uses the notation $c \in C$ rather than $d \in D$.

[5]Ellis denotes the cardinality of I by $2n + 1$, for some positive integer n. We use N rather than $2n$ since we allow an even number of jurors.

objectively determined here; we have what Jaffray (1991) calls an environment of "imprecise risk". Subjective variation in the perception of ambiguity would, of course, be natural to assume, but our purpose is to establish, as a benchmark, the pure effects of ambiguity. In Feddersen and Pesendorfer's model, as in Ellis's, the precisely specified prior belief is also assumed to be commonly held, and therefore objective in the same sense.

Prior to casting her vote, each juror receives a private signal $t \in T = \{1, 2\}$. Conditional on $s \in S$, these signals are independently and identically distributed with $\Pr(1|a) = \Pr(2|b) = r \in \left(\frac{1}{2}, 1\right)$. This information structure is a special case of that of Ellis (2016), who allows T to be any finite set containing at least two elements, and the state-conditional signal distributions to be arbitrarily specified. We specialise to the two-signal case so that our results are comparable with those in Feddersen and Pesendorfer (1998).[6]

Let $\Omega = S \times T^I$ be the state space characterising all *ex ante* uncertainty. Each $p \in \left[\underline{p}, \overline{p}\right]$ determines a probability over Ω, with

$$\Pr(a, t_1, ..., t_{N+1}) = pr^{N+1-\sum_i (t_i - 1)} (1-r)^{\sum_i (t_i - 1)}$$

and

$$\Pr(b, t_1, ..., t_{N+1}) = (1-p) r^{\sum_i (t_i - 1)} (1-r)^{N+1-\sum_i (t_i - 1)}.$$

The set of such probabilities is closed and convex, and denoted by Π. Each juror uses *full Bayesian updating (FBU)* to condition on her privately observed signal: each prior is updated by Bayes' rule to form the juror's (private) set of posteriors. The posterior interval for the conditional probability $\Pr(a|t_i = t)$ is independent of i (since each signal is equally precise and signal distributions are identical across jurors) and is denoted by $\Pi_t = [\underline{\pi}_t, \overline{\pi}_t]$, with generic element π_t. Since $\left[\underline{p}, \overline{p}\right] \subseteq (0, 1)$ it follows that $\Pi_t \subseteq (0, 1)$ also.

Voters have *maxmin expected utility (MEU)* preferences and share a common Bernoulli utility function, $u : D \times S \to \mathbb{R}$, specified as follows: $u(A, a) = u(B, b) = 1$, $u(A, b) = 0$ and $u(B, a) = -c$, where $c \geq 0$. Thus, all agree that A is the correct decision in state a and B is the correct decision in state b. Ellis's model is the special case in which $c = 0$. When $c > 0$ errors attract asymmetric penalties: convicting the innocent results in lower utility than acquitting the guilty.

Our model also nests that of Feddersen and Pesendorfer (1998). Their juror utility function is obtained by first applying the positive, affine transformation $u \longmapsto (c+2)^{-1}(u-1)$ and then defining

$$q = \frac{c+1}{c+2} \tag{1}$$

Our model is therefore equivalent to Feddersen and Pesendorfer's when $\underline{p} = \overline{p} = \frac{1}{2}$.[7] In particular:

$$\pi u(B, a) + (1-\pi) u(B, b) \geq \pi u(A, a) + (1-\pi) u(A, b)$$

[6] Proposition 2 in Ellis (2016) also concerns this special case.

[7] Feddersen and Pesendorfer also allow $c \in (-1, 0)$ (i.e., $q \in \left(0, \frac{1}{2}\right)$) but this case is symmetric to ours – it effectively reverses the roles of the states – so nothing is lost by omitting it.

$$\Leftrightarrow \quad \pi \leq \frac{1}{2+c} \quad \Leftrightarrow \quad 1-\pi \geq q \tag{2}$$

Thus, B is an optimal decision if the decision-maker assesses that state $s=a$ has probability *no greater than* $(2+c)^{-1}$; equivalently, that state $s=b$ has probability *no less than* q. One can think of Feddersen and Pesendorfer's parameter q as quantifying the meaning of "beyond reasonable doubt". Belief in guilt must clear this hurdle in order for conviction to be the preferred decision. In Feddersen and Pesendorfer's model, this belief is the probability that the juror assigns to $s=b$ after conditioning on her private signal *and the event that her vote is pivotal* (given the voting strategies of the other voters). As Ellis (2016) notes, matters are more complicated when jurors have ambiguous beliefs. Since a voter has a *set* of posteriors after conditioning on her signal, and since the "operative" posterior probability from this set may be *decision-contingent*, it is no longer possible to condition on pivotality to determine optimal voting behaviour. We discuss this phenomenon in more detail below.[8]

A *strategy* for voter i is a mapping from T to $[0,1]$, specifying the probability of casting a B vote conditional on each possible signal realisation. Let σ_t^i denote the probability that $i \in I$ votes B after observing $t \in T$, and let $\sigma^i = (\sigma_1^i, \sigma_2^i)$ denote i's strategy.

Following Feddersen and Pesendorfer (1998) and Ellis (2016), we focus exclusively on *symmetric* equilibria, in which each voter follows the same strategy. Consider a generic voter i who believes that the other voters follow a common strategy, denoted $\sigma = (\sigma_1, \sigma_2)$. Let ρ_s denote the probability that voter i's vote is pivotal, conditional on being in state $s \in S$; let θ_s denote the probability that i is not pivotal and (in addition) a "correct" decision is made, again conditional on being in state $s \in S$. After observing her private signal $t \in T$, voter i therefore chooses $\sigma_t^i \in [0,1]$ to maximise

$$\min_{\pi_t \in \Pi_t} \pi_t \left[\rho_a \left(1-\sigma_t^i - c\sigma_t^i\right) + \theta_a - (1-\rho_a-\theta_a)\, c\right] \;+\; (1-\pi_t)\left[\rho_b \sigma_t^i + \theta_b\right]$$

Of course, both ρ_s and θ_s depend on σ but (following Ellis) we suppress this dependence in the notation for convenience. In particular, the unanimity rule implies that

$$\rho_a = \left[r\sigma_1 + (1-r)\,\sigma_2\right]^N \tag{3}$$

and

$$\rho_b = \left[(1-r)\,\sigma_1 + r\sigma_2\right]^N \tag{4}$$

while $\theta_a = 1-\rho_a$ and $\theta_b = 0$. In other words, σ_t^i solves

$$\max_{\sigma_t^i \in [0,1]} \min_{\pi_t \in \Pi_t} V\left(\sigma_t^i, \sigma; \pi_t\right) \tag{5}$$

[8]See also Pan (2019).

where

$$V\left(\sigma_t^i, \sigma; \pi_t\right) \;=\; \pi_t\left[1-(1+c)\,\rho_a\sigma_t^i\right] \;+\; (1-\pi_t)\,\rho_b\sigma_t^i \tag{6}$$

It is possible that minimising posterior in Π_t may vary with σ_t^i, so we cannot condition on pivotality when solving (5).

Since we focus on symmetric equilibria, we usually omit the player superscript on strategies, and abuse notation by using $\sigma = (\sigma_1, \sigma_2)$ to denote both the strategy of a generic voter in a symmetric profile and the symmetric profile itself. We are interested in symmetric strategy profiles (σ_1, σ_2) that form a Bayesian Nash equilibrium of this voting game, in the sense that $\sigma_t^i = \sigma_t$ solves (5) for each $t \in T$. The term "equilibrium" will be used to indicate a symmetric profile that is a Bayesian Nash equilibrium.

To characterise equilibria we must first describe best responses to symmetric profiles. This is done in the next section, and summarised by Figure 1.

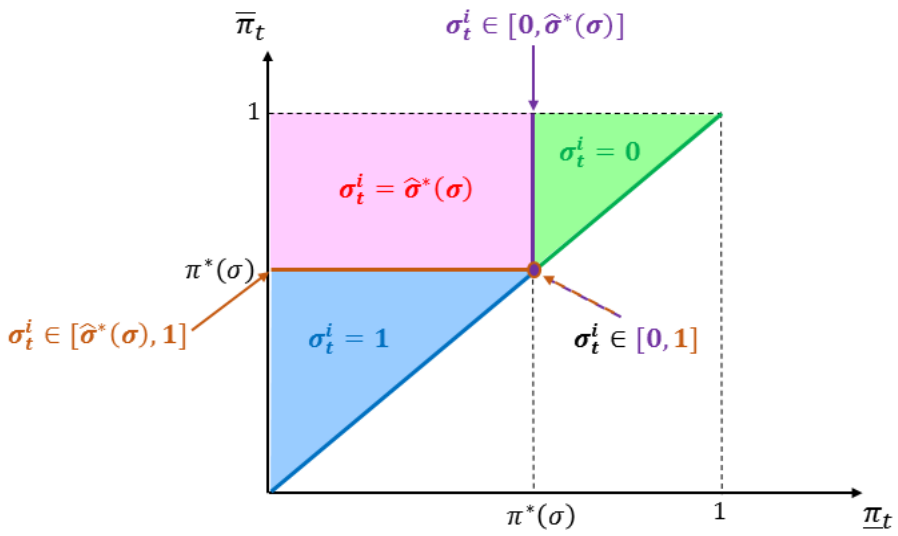


Figure 1: Optimal responses

3 The best response correspondence

The symmetric profile $\sigma = (0, 0)$ is always an equilibrium of our game. If $\sigma_1 = \sigma_2 = 0$ then $\rho_a = \rho_b = 0$ and anything is a best response for voter i, since the defendant will be acquitted no matter how i votes. From now on we exclude this (uninteresting) equilibrium and focus on best responses to symmetric profiles with $\sigma_1 + \sigma_2 > 0$, so that $\rho_a > 0$ and $\rho_b > 0$ (i.e., profiles such that each voter has a positive probability of being pivotal conditional on either state).

Consider Figure 1. In this figure,

$$\pi^* (\sigma) \;=\; \frac{\rho_b}{\rho_b + (1+c)\,\rho_a} \;=\; \frac{1}{1 + (1+c)\,(\rho_a/\rho_b)} \tag{7}$$

and $\hat{\sigma}^* (\sigma) = \min\{\sigma^* (\sigma), 1\}$, where

$$\sigma^* (\sigma) \;=\; \frac{1}{\rho_b + (1+c)\,\rho_a} \;=\; \frac{\pi^* (\sigma)}{\rho_b} \tag{8}$$

Note that $\pi^* (\sigma)$ and $\sigma^* (\sigma)$ are well-defined when $\rho_a > 0$ and $\rho_b > 0$. The notation emphasises the implicit dependence of π^* and σ^* on σ.

We use Figure 1 to identify i's best response(s) to $\sigma \neq (0,0)$ as follows. First, calculate $\pi^* (\sigma)$ and $\hat{\sigma}^* (\sigma)$ using (7) and (8). Next, for each $t \in T$ locate the point $(\underline{\pi}_t, \overline{\pi}_t)$ in Figure 1 – it will obviously sit on or above the 45 degree line. The coloured region into which $(\underline{\pi}_t, \overline{\pi}_t)$ falls determines the optimal value(s) for σ_t^i as indicated in Figure 1.

Figure 1 will play a critical role in the arguments to follow. Its formal derivation is described in Appendix A.

Recall that $\sigma = (0,0)$ is an equilibrium of our model. We will identify any other equilibria with the help of Figure 1. Let us therefore pause to highlight the main features of this figure.

- First, since $\pi^* (\sigma) \in (0,1)$ for any $\sigma \neq (0,0)$, there is always a non-empty (green) triangular region in which it is uniquely optimal to vote for A (acquittal) following signal t, and a non-empty (blue) triangular region in which it is uniquely optimal to vote for B (conviction) following signal t.

- Along the 45 degree line in Figure 1 there is no ambiguity and we recover the best response correspondence for the model of Feddersen and Pesendorfer (1998). Voter i strictly prefers to vote for A following signal t when $\underline{\pi}_t = \overline{\pi}_t > \pi^* (\sigma)$, strictly prefers to vote for B when $\underline{\pi}_t = \overline{\pi}_t < \pi^* (\sigma)$ and is indifferent amongst all $\sigma_t^i \in [0,1]$ when $\underline{\pi}_t = \overline{\pi}_t = \pi^* (\sigma)$.[9]

- Once we allow for ambiguous priors – that is, once we move above the 45 degree line in Figure 1 – a much richer picture emerges. Suppose $\sigma^* (\sigma) < 1$. Since (8) implies

[9]Let $\mathcal{P}_i$ denote the event in which i is pivotal, so that $\rho_a = \Pr[\mathcal{P}_i | a]$. It is easy to check that

$$\Pr[a \mid \mathcal{P}_i \text{ and } t_i = t] \;=\; \frac{\rho_a \Pr[a|t_i = t]}{\rho_a \Pr[a|t_i = t] + \rho_b \Pr[b|t_i = t]}.$$

$\sigma^*(\sigma) > 0$ when $\sigma \neq (0,0)$, there exist optimal responses with $0 < \sigma_t^i < 1$ whenever $(\underline{\pi}_t, \overline{\pi}_t)$ lies in the rectangle between the two triangular regions, but only one point in this rectangle – the southeast corner – at which *all* $\sigma_t^i \in [0,1]$ are optimal responses. Moreover, at each point within the (pink) interior of the rectangle, $\sigma_t^i = \sigma^*(\sigma)$ is the *unique* best response. This response perfectly hedges against uncertainty by equalising the expected payoff in each state: the value of $V(\sigma^*(\sigma), \sigma; \pi_t)$ is the same for all π_t.

4 Equilibria

Since $r > \frac{1}{2}$ it is easy to verify that $(\underline{\pi}_1, \overline{\pi}_1) \gg (\underline{\pi}_2, \overline{\pi}_2)$: the point $(\underline{\pi}_1, \overline{\pi}_1)$ will lie *strictly to the northeast* of the point $(\underline{\pi}_2, \overline{\pi}_2)$ when plotted in Figure 1. We therefore deduce the following intuitive but important fact:

Lemma 4.1 *If $\sigma = (\sigma_1, \sigma_2) \neq (0,0)$ is an equilibrium, then $\sigma_2 \geq \sigma_1$.*

In the language of Feddersen and Pesendorfer (1998), an equilibrium is *non-responsive* if $\sigma_2 = \sigma_1$ and *responsive* if $\sigma_2 > \sigma_1$. We analyse these two classes of equilibria in Sections 4.1 and 4.2 respectively. As a benchmark, recall that the (symmetric Bayesian Nash) equilibria of Feddersen and Pesendorfer's (1998) model exhibit the following features: (i) $\sigma_1 = \sigma_2 \in \{0,1\}$ in any non-responsive equilibrium; (ii) $\sigma_2 = 1$ in any responsive equilibrium; and (iii) responsive equilibria are unique when they exist. It turns out that none of these properties generalises to an environment with ambiguity.[10]

4.1 Non-responsive equilibria

As already noted, our voting game always has a trivial non-responsive equilibrium in which $\sigma_1 = \sigma_2 = 0$. At the other extreme, if $\overline{\pi}_1$ is sufficiently low, there exists another non-responsive equilibrium with $\sigma_1 = \sigma_2 = 1$.

Therefore, using (1) and the facts that $\rho_a > 0$ and $\rho_b > 0$, we have:

$$\begin{aligned} \Pr[a|t_i = t] \leq \pi^*(\sigma) \quad &\Leftrightarrow \quad \Pr[a \mid \mathcal{P}_i \text{ and } t_i = t] \leq \frac{1}{2+c} \\ &\Leftrightarrow \quad \Pr[b \mid \mathcal{P}_i \text{ and } t_i = t] \geq q \end{aligned}$$

The latter is the form in which the condition appears in Feddersen and Pesendorfer (1998).

[10] Proposition 4.1 shows that (i) does not hold; Proposition 4.2 proves that (ii) does not hold; Figure 2 illustrates the possibility of multiple responsive equilibria (though Lemma 4.3 establishes that there is at most one responsive equilibrium with $\sigma_2 = 1$, as in Feddersen and Pesendorfer's model).

Lemma 4.2 *The symmetric profile with $\sigma_1 = \sigma_2 = 1$ is an equilibrium iff*

$$\overline{\pi}_1 \leq \frac{1}{2+c}.$$

Proof: From Figure 1, $\sigma = (1,1)$ is an equilibrium iff

$$\overline{\pi}_1 \leq \pi^* ((1,1)) \quad \Leftrightarrow \quad \overline{\pi}_1 \leq \frac{1}{2+c}$$

where we have used (7). $\square$

There may also exist a strictly mixed non-responsive equilibrium.

Proposition 4.1 *There is at most one equilibrium with $0 < \sigma_1 = \sigma_2 < 1$. Such an equilibrium exists iff*

$$\underline{\pi}_1 \leq \frac{1}{2+c} \leq \overline{\pi}_2$$

in which case it is given by

$$\sigma_1 = \sigma_2 = \left(\frac{1}{2+c}\right)^{\frac{1}{N+1}}.$$

Proof: From Figure 1 and the fact that $(\underline{\pi}_1, \overline{\pi}_1) \gg (\underline{\pi}_2, \overline{\pi}_2)$, if σ is an equilibrium with $0 < \sigma_1 = \sigma_2 < 1$ then it must be the case that $\sigma_1 = \sigma_2 = \sigma^*$. Moreover, if $\sigma_1 = \sigma_2$ then $\rho_a = \rho_b$ and hence

$$\pi^* (\sigma) = \frac{1}{2+c} \tag{9}$$

from (7). We must therefore have

$$\underline{\pi}_1 \leq \frac{1}{2+c} \leq \overline{\pi}_2 \tag{10}$$

for such an equilibrium to exist (recall Figure 1). This proves the necessity of condition (10).

Provided (10) holds, it follows that an equilibrium with $\sigma_1 = \sigma_2 = x \in (0,1)$ exists iff $x = \sigma^* ((x,x))$. Using (8) and (9), this is equivalent to

$$\begin{aligned} x = \frac{\pi^*((x,x))}{\rho_b} \quad &\Leftrightarrow \quad x = \frac{1}{(2+c)\left[(1-r)x + rx\right]^N} \\ &\Leftrightarrow \quad x = \frac{1}{(2+c)x^N} \\ &\Leftrightarrow \quad x = \left(\frac{1}{2+c}\right)^{\frac{1}{N+1}} \end{aligned}$$

Since

$$\left(\frac{1}{2+c}\right)^{\frac{1}{N+1}} \in (0,1)$$

we have established the remaining claims in the proposition. □

When $c = 0$, Proposition 4.1 gives an analogue of Ellis (2016, Proposition 1) for the unanimity rule (and our specialised information structure). Under condition (10), there exists a strictly mixed equilibrium in which private signals are ignored and the associated equilibrium probability of conviction is $\frac{1}{2}$. Moreover, if

$$\underline{\pi}_1 < \frac{1}{2} < \overline{\pi}_2 \tag{11}$$

(i.e., when voters "lack confidence", in the language of Ellis), then the equilibrium identified in Proposition 4.1 is *strict*, in the sense that best responses (conditional on either signal) are *unique*. The same is true for Ellis' result under majority rule.[11] Lack of confidence creates uncertainty about the "right" way to vote: since $(\underline{\pi}_1, \overline{\pi}_1) \gg (\underline{\pi}_2, \overline{\pi}_2)$, and recalling (2), condition (11) means that neither signal allows voters to exclude posteriors that justify a strict preference for $d = A$ nor posteriors that justify a strict preference for $d = B$. In this situation, uncertainty-averse voters may strictly prefer to replace uncertainty with risk by voting randomly.

If $c > 0$ then Proposition 4.1 establishes the existence of a qualitatively similar equilibrium, albeit (not surprisingly) with a lower probability of conviction than when $c = 0$. However, the conviction probability is still positive *in either state* and this probability is *independent of* N.

4.2 Responsive equilibria

In Feddersen and Pesendorfer (1998) all responsive equilibria have $\sigma_2 = 1$. Such equilibria may also exist in our model. As in Feddersen and Pesendorfer's model, they are unique when they exist.

Lemma 4.3 *There is at most one responsive equilibrium with $\sigma_2 = 1$.*

Proof: If $\sigma = (\sigma_1, 1)$ then

$$\frac{\rho_a}{\rho_b} = \left[\frac{r\sigma_1 + (1-r)}{(1-r)\,\sigma_1 + r}\right]^N$$

[11] Proposition 4.1 (for $c = 0$) is somewhat stronger than Ellis' result (albeit for a restricted information structure), in that Proposition 4.1 provides necessary and sufficient conditions for existence of equilibria with $\sigma_1 = \sigma_2 \in (0,1)$, and also establishes the uniqueness of such equilibria when they exist.

which is strictly increasing in σ_1. It follows that $\pi^*((\sigma_1, 1))$ is strictly decreasing in σ_1. Using (8) we have

$$\sigma^*((\sigma_1, 1)) = \frac{\pi^*((\sigma_1, 1))}{[(1-r)\sigma_1 + r]^N}.$$

so $\sigma^*((\sigma_1, 1))$ is also strictly decreasing in σ_1. Suppose $\sigma' = (\sigma_1', 1)$ and $\sigma'' = (\sigma_1'', 1)$ are both equilibria with $0 \leq \sigma_1' < \sigma_1'' < 1$. Then $\pi^*(\sigma') > \pi^*(\sigma'')$ and $\sigma^*(\sigma') > \sigma^*(\sigma'')$. Hence $\hat{\sigma}^*(\sigma') \geq \hat{\sigma}^*(\sigma'')$. By inspection of Figure 1 we must therefore have $\sigma_1' \geq \sigma_1''$, which is the desired contradiction. □

Calculating the unique value of σ_1 in such an equilibrium is less straightforward than in Feddersen and Pesendorfer (1998), since there may be *many* σ_1 values for which randomisation is an optimal response to $\sigma = (\sigma_1, 1)$ given $t = 1$ (see Figure 1). Fortunately, it is unnecessary to characterise σ_1 in order to establish our main result, nor to identify the precise conditions under which a responsive equilibrium with $\sigma_2 = 1$ exists. We therefore eschew any further discussion of these matters here, though the interested reader may find such discussion in Fabrizi et al. (2019b).

In the presence of ambiguity, there may also exist responsive equilibria with $\sigma_2 < 1$. Provided N is sufficiently large, such equilibria must have $\sigma_1 > 0$.

Lemma 4.4 *If N is sufficiently large, there cannot exist a responsive equilibrium with $\sigma_1 = 0$.*

Proof: Suppose σ satisfies $0 = \sigma_1 < \sigma_2$. Then (7) and (8) give:

$$\pi^*(\sigma) = \frac{1}{1 + (1+c)\left(\frac{1-r}{r}\right)^N} \to 1 \quad \text{as } N \to \infty$$

(where we have used the fact that $r > \frac{1}{2}$). Since $\overline{\pi}_1 < 1$ it follows, by inspection of Figure 1, that σ cannot be an equilibrium if N is sufficiently large. □

Since our interest is in asymptotic properties of equilibria, we may therefore ignore responsive equilibria with $\sigma_1 = 0$. It remains to consider responsive equilibria which are also *strictly mixed.* That is, equilibria with $0 < \sigma_1 < \sigma_2 < 1$. It turns out that such equilibria may exist in our model, even asymptotically. The following result is proved in Appendix B.

Proposition 4.2 *A responsive, strictly mixed equilibrium exists only if $\overline{\pi}_2 \geq \underline{\pi}_1$. When $\overline{\pi}_2 \geq \underline{\pi}_1$ there exist functions $\alpha_1(N)$ and $\alpha_2(N)$ satisfying*

$$\frac{1}{2+c} < \alpha_2(N) \leq \alpha_1(N) \leq \frac{r^N}{r^N + (1-r)^N(1+c)}$$

for all N, such that a responsive, strictly mixed equilibrium exists iff either

$$\frac{1}{2+c} < \underline{\pi}_1 < \alpha_1(N) \tag{I}$$

or

$$\frac{1}{2+c} < \overline{\pi}_2 < \alpha_2(N) \tag{II}$$

The functions α_1 and α_2 are (implicitly) defined in Appendix B. We also provide a characterisation (again, implicit) of each strictly mixed equilibrium. As the proof of Proposition 4.2 makes clear, when $\underline{\pi}_1 < \overline{\pi}_2$ and conditions (I) and (II) are *both* satisfied, there exist exactly two strictly mixed responsive equilibria; if $\underline{\pi}_1 = \overline{\pi}_2$ then condition (II) implies condition (I), and if (I) holds there is a continuum of strictly mixed responsive equilibria; otherwise, there exists at most one strictly mixed responsive equilibrium. It is also worth noting that if condition (II) holds and

$$\underline{\pi}_1 \leq \frac{1}{2+c}$$

then a strictly mixed responsive equilibrium co-exists with a non-responsive equilibrium of the sort described in Proposition 4.1.

Figure 2 summarises the conditions for the existence of strictly mixed equilibria, both responsive (Proposition 4.2) and non-responsive (Proposition 4.1). These conditions restrict $\underline{\pi}_1$ (measured on the horizontal axis of Figure 2) and $\overline{\pi}_2$ (measured on the verical axis). A necessary condition for any strictly mixed equilibrium to exist is that $\overline{\pi}_2 \geq \underline{\pi}_1$ so only points on or above the diagonal are relevant; moreover, we have $\underline{\pi}_1 > 0$ and $\overline{\pi}_2 < 1$ by assumption. If $(\underline{\pi}_1, \overline{\pi}_2)$ lies in the red area (including the red boundaries) then a non-responsive, strictly mixed equilibrium exists. If $(\underline{\pi}_1, \overline{\pi}_2)$ lies *strictly* between the horizontal blue and red lines, or *strictly* between the vertical red and purple lines, then a responsive strictly mixed equilibrium exists. In regions B and C of Figure 2 there exist multiple strictly mixed equilibria: one responsive and one non-responsive equilibrium in region B;[12] two responsive equilibria in region C.[13]

The equilibria of Proposition 4.2 are interesting in themselves. They are unique to the model with ambiguity but Ellis (2016) does not identify any analogous equilibria in the context of majority voting (which is not to say that none exists). However, they are also relevant to our main purpose. The following lemma, which is proved in Appendix C, verifies that such equilibria may exist for arbitrarily large N; in particular, region D in Figure 2 does not vanish in the limit as $N \to \infty$.

[12] Where B is taken to include the red vertical boundary (excluding its end points) but none of its other boundaries.

[13] Where C is taken to include the green boundary (excluding its end points) but none of its other boundaries.

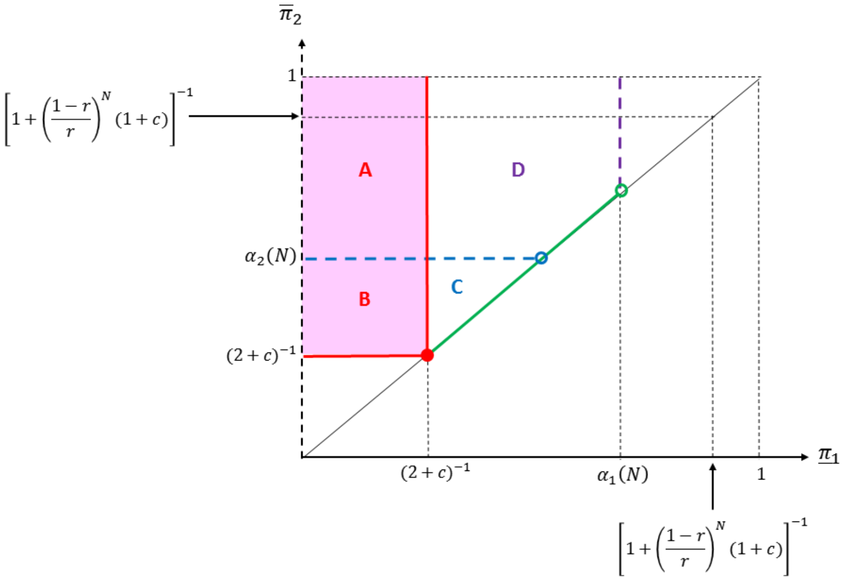


Figure 2: Parameter regions in which strictly mixed equilibria exist

Lemma 4.5 *There exists some $\eta > 0$ such that*

$$\alpha_1(N) \geq \frac{1}{2+c} + \eta$$

for all N, where α_1 is the function whose existence is established in Proposition 4.2.

It follows that if $\overline{\pi}_2 \geq \underline{\pi}_1$ and

$$\frac{1}{2+c} < \underline{\pi}_1 < \frac{1}{2+c} + \eta$$

there exists a strictly mixed responsive equilibrium for any N.

It is worth pausing to understand the nature of such an equilibrium – one whose existence is guaranteed by condition (I) of Proposition 4.2. This nature is spelled out in detail in Appendix B,[14] but let us summarise its key features here. Suppose $\sigma = (\sigma_1, \sigma_2)$ denotes such an equilibrium. Then signal $t = 2$ puts a juror in two minds – one end of the posterior interval supports vote A while the other end supports vote B – so the juror randomises to perfectly hedge against this uncertainty: $\sigma_2 = \sigma^*(\sigma) \in (0,1)$. Signal $t = 1$

[14]And summarised in Figure 6(II).

supports vote A but not decisively: the juror would be indifferent about how to vote if she held posterior $\underline{\pi}_1$. It is therefore optimal to vote for A but also to vote for B with any probability up to $\sigma^*(\sigma)$. Thus, $\sigma_1 \in (0, \sigma^*(\sigma))$ consistent with motivating a type $t = 2$ juror to hedge.

The indifference of the type $t = 1$ juror fixes the ratio ρ_a/ρ_b, and hence also the ratio σ_1/σ_2.[15] If parameter N increases, then the required ratio σ_1/σ_2 will also increase:[16] since ρ_a/ρ_b is less than 1 in any responsive equilibrium, raising N reduces the likelihood of being pivotal in $s = a$ relative to being pivotal in $s = b$, which necessites an off-setting increase in σ_1/σ_2 (the probability of voting to convict given $t = 1$ relative to voting to convict given $t = 2$). Let $\sigma' = (\sigma_1', \sigma_2')$ denote the new equilibrium strategy (for the higher value of N), we must have $\sigma_1'/\sigma_2' > \sigma_1/\sigma_2$. To ensure that the type $t = 2$ juror still wants to hedge, we must also have $\sigma_1' > \sigma_1$: if $\sigma_1' \leq \sigma_1$ then $\sigma_2' < \sigma_2$ (given $\sigma_1'/\sigma_2' > \sigma_1/\sigma_2$) which implies $\sigma^*(\sigma') > \sigma^*(\sigma)$ from (8), so $\sigma_2' < \sigma_2 = \sigma^*(\sigma) < \sigma^*(\sigma')$.[17]

Since $\sigma_1'/\sigma_2' > \sigma_1/\sigma_2$, the new equilibrium strategies are closer to non-responsiveness than the old, which is why, asymptotically, an equilibrium in this class can only exist for parameter values sufficiently close to the (red) western boundary of the C and D regions in Figure 2. It is also a driving force behind our version of the Jury Paradox. Even for these exotic equilibria, one type of juror must be maintained in a state of indifference, which requires a fixed value for the likelihood ratio ρ_a/ρ_b. As the number of jurors increases, this forces voting behaviour to become increasingly uninformative about private signals. A similar phenomenon drives Feddersen and Pesendorfer's result. The hard work is to show that informativeness vanishes sufficiently quickly to overcome the increasingly demanding unanimity requirement such that error probabilities remain bounded away from zero. This work is undertaken in the next section.

5 The Jury Paradox

We are now ready to prove our main result. Recall the Jury Paradox of Feddersen and Pesendorfer (1998): excluding the trivial equilibrium in which $\sigma = (0, 0)$, the equilibrium probability that a convicted defendant is innocent is bounded away from zero independently of N.[18] In other words, under the unanimity rule, the probability that a conviction is erroneous does not vanish in the limit. In this section we show that this result survives the introduction of prior ambiguity into Feddersen and Pesendorfer's model. The greater

[15] As described by the locus $\Omega(\underline{\pi}_1)$ in Figure 7 of Appendix B.

[16] As N increases, the line $\Omega(\underline{\pi}_1)$ in Figure 7 gets flatter.

[17] Consider equation (CI) in Appendix B. If we fix σ_2 and increase N, then σ_1 must increase to maintain the equality. Hence, the curve of solutions to (C1) in Figure 7 must move rightwards. Since the curve $\Omega(\underline{\pi}_1)$ is getting flatter, the value of σ_1 must increase at the intersection.

[18] Feddersen and Pesendorfer (1998) actually bound this probability for any Bayesian Nash equilibrium – *symmetric or otherwise* – in which there is a positive probability of conviction (*ibid.*, Proposition 1).

range, and more complex structure, of equilibria in the presence of ambiguity is what renders this exercise non-trivial.

Let $\mathcal{C}$ denote the event that the defendant is convicted. For any *non-responsive* equilibrium, the probability of conviction is independent of whether the defendant is guilty or innocent, so the probability that a convicted defendant is innocent satisfies $\Pr(a|\mathcal{C}) \geq \underline{p} > 0$ for any N and any prior $p \in [\underline{p}, \overline{p}]$. For any non-responsive equilibrium $\sigma \neq (0,0)$, the probability of conviction is strictly positive and bounded away from zero independently of N: Proposition 4.1 ensures that $\Pr(\mathcal{C}) \geq (2+c)^{-1}$ when $\sigma_1 = \sigma_2 \in (0,1)$, and the same bound obviously holds if $\sigma = (1,1)$. Thus, the probability that an innocent defendant is convicted is bounded away from zero independently of N for any non-responsive equilibrium other than $\sigma = (0,0)$. In the rest of this section, we show that the same is true for responsive equilibria.

We start by bounding the probability of innocence *conditional* on a conviction being entered.

Proposition 5.1 *There exists some $\kappa > 0$, which does not depend on N, such that in any responsive equilibrium,* $\Pr(a|\mathcal{C}) \geq \kappa$ *for any prior $p \in [\underline{p}, \overline{p}]$.*

Proof: From Figure 1, if σ is a responsive equilibrium then $\pi^*(\sigma) \leq \overline{\pi}_1$. Hence, using (7),

$$\ell(\overline{\pi}_1) \leq \frac{\rho_a}{\rho_b} \tag{12}$$

where

$$\ell(x) = \frac{1-x}{x(1+c)} \tag{13}$$

Since $\overline{\pi}_1 \in (0,1)$ it follows that $\ell(\overline{\pi}_1) > 0$. Fix some prior $p \in [\underline{p}, \overline{p}]$. We can use (12) and Bayes' Rule to bound $\Pr(a|\mathcal{C})$ as follows:

$$\begin{aligned}
\Pr(a|\mathcal{C}) &= \frac{p\Pr(\mathcal{C}|a)}{p\Pr(\mathcal{C}|a) + (1-p)\Pr(\mathcal{C}|b)} \\
&= \frac{p(\rho_a/\rho_b)^{\frac{N+1}{N}}}{p(\rho_a/\rho_b)^{\frac{N+1}{N}} + (1-p)} \\
&\geq \frac{p\ell(\overline{\pi}_1)^{\frac{N+1}{N}}}{p\ell(\overline{\pi}_1)^{\frac{N+1}{N}} + (1-p)}
\end{aligned}$$

Letting

$$\kappa = \frac{\underline{p}\min\left\{\ell(\overline{\pi}_1), \ell(\overline{\pi}_1)^2\right\}}{\underline{p}\min\left\{\ell(\overline{\pi}_1), \ell(\overline{\pi}_1)^2\right\} + \left(1-\underline{p}\right)} > 0$$

gives a bound that is independent of N and $p \in [\underline{p}, \overline{p}]$. $\square$

It follows immediately that:

Corollary 5.1 *There exists some $\kappa > 0$, which does not depend on N, such that for any equilibrium $\sigma \neq (0,0)$,*

$$\Pr(a|\mathcal{C}) \ \geq \ \min\{\kappa, \ \underline{p}\} > 0$$

for any prior $p \in [\underline{p}, \overline{p}]$.

It remains to show that the asymptotic probability of conviction is non-zero for responsive equilibria.

For this purpose, it is convenient to divide the responsive equilibria into two categories: those with $\sigma_2 = 1$ and those with $\sigma_2 < 1$. Equilibria in the former category have the same qualitative features as the responsive equilibria in Feddersen and Pesendorfer (1998), and are unique when they exist (Lemma 4.3). The following is proved in Appendix D.

Proposition 5.2 *There exists $\gamma > 0$ such that any responsive equilibrium with $\sigma_2 = 1$ satisfies*

$$\Pr(\mathcal{C}|a) \ \geq \ \left[\frac{(2r-1)\,\ell(\overline{\pi}_1)^{\frac{1}{N}}}{r-(1-r)\,\ell(\overline{\pi}_1)^{\frac{1}{N}}}\right]^{N+1} \geq \gamma \tag{14}$$

if N is sufficiently large.[19]

It follows that the probability of convicting an innocent defendant in such an equilibrium is at least $\underline{p}\gamma > 0$ for any prior $p \in [\underline{p}, \overline{p}]$ and any sufficiently large N. It further follows that the probability of conviction in any responsive equilibrium with $\sigma_2 = 1$ is bounded away from zero independently of $N \in \{1, 2, ...\}$ and $p \in [\underline{p}, \overline{p}]$.

It remains to consider the responsive equilibria with $\sigma_2 < 1$. Recalling Lemma 4.4, any such equilibrium must be strictly mixed when N is large. Lemma 4.5 confirms that strictly mixed responsive equilibria may exist asymptotically. The following result, whose proof may be found in Appendix E, shows that along any sequence of strictly mixed responsive equilibria with jury size converging to infinity, the probability of conviction is bounded away from zero uniformly in $p \in [\underline{p}, \overline{p}]$.

Proposition 5.3 *Suppose $\sigma^{(k)}$ is a strictly mixed responsive equilibrium for a jury of size $N_k \in \{1, 2, ...\}$, with conviction probability $\Pr(\mathcal{C}) = \gamma_k(p)$ for prior $p \in [\underline{p}, \overline{p}]$. If $N_k \to \infty$ as $k \to \infty$, then*

$$\min_{p \in [\underline{p}, \overline{p}]} \left[\liminf_{k \to \infty} \gamma_k(p)\right] \ > \ 0.$$

[19] Recall that $\ell(x)$ is defined in (13).

Let us summarise the story so far. Excluding the trivial equilibrium, Corollary 5.1 says that the probability a convicted defendant is innocent is bounded away from zero independently of N and $p \in \left[\underline{p}, \overline{p}\right]$. The probability of conviction is likewise uniformly bounded away from zero for non-trivial equilibria that are non-responsive (recall the discussion prior to Proposition 5.1) and for responsive equilibria (Propositions 5.2 and 5.3). We have therefore arrived at our main result:

Theorem 5.1 *For each $N \in \{1, 2, ...\}$, let $\sigma^{(N)} \neq (0, 0)$ be an equilibrium for a jury of size N and let $\Pr(\mathcal{C}|a) = \psi^{(N)}(p)$ be the probability of convicting an innocent defendant in $\sigma^{(N)}$ given prior $p \in \left[\underline{p}, \overline{p}\right]$. Then*

$$\min_{p \in \left[\underline{p}, \overline{p}\right]} \left[\liminf_{N \to \infty} \psi^{(N)}(p)\right] > 0.$$

Theorem 5.1 gives an analogue of Feddersen and Pesendorfer's (1998) Jury Paradox for juries with ambiguous priors. Its statement is a little more cumbersome in the presence of ambiguity, since responsive equilibria need not be unique even for arbitrarily large N. However, its spirit is intact: under the unanimity rule, the probability of convicting an innocent defendant is bounded away from zero, and this bound can be chosen independently of the prior (within the allowable set $\left[\underline{p}, \overline{p}\right]$).

Theorem 5.1 holds even if $c = 0$. In this case, Ellis's (2016) Theorem 2 tells us that decisions by majority rule are asymptotically correct along *some* sequence of equilibria, *provided* Π_1 *and* Π_2 *have disjoint interiors*. Our result does not require either italicised caveat, but when Π_1 and Π_2 have disjoint interiors it follows that majority rule may eliminate the possibility of convicting the innocent, at least asymptotically, whereas the unanimity rule cannot.

6 Concluding remarks

Ellis (2016) established a qualified version of Condorcet's favourable result for majority voting (the "Jury Theorem") in the presence of ambiguous prior beliefs. We have proved an unqualified extension of Feddersen and Pesendorfer's (1998) *unfavourable* result on the unanimity rule (the "Jury Paradox") to the environment with ambiguous priors.

An interesting by-product of this work has been a complete characterisation of the strictly mixed equilibria of the voting game under unanimity. We have shown that *responsive* strictly mixed equilibria may exist and may be non-unique; that they may co-exist with *non-responsive* strictly mixed equilibria; and that they may exist for arbitrarily large jury size. Fabrizi et al. (2019b) study responsive equilibria of the more conventional (i.e., Feddersen and Pesendorfer) variety: those with $\sigma_2 = 1$. It turns out that these have somewhat curious comparative static properties as the level of ambiguity changes; properties which we plan to test experimentally.

This paper, like Ellis (2016), restricts the locus of ambiguity to the prior. However, it is equally natural to consider the possibility that jurors may have ambiguous assessments of the precision of signals: ambiguity about the conditional probabilities, $\Pr(s|t)$. Fabrizi et al. (2019a) consider a variation on the present model in which jurors have precise prior beliefs but entertain a set of values for r. They remain certain that everyone's signal has the same precision and that signal draws are independent. For "conventional" equilibria (i.e., those with $\sigma_2 = 1$), it is shown, both theoretically and experimentally, that voting is more informative under ambiguity than it would be if jurors had precise beliefs about r, specified as the mid-point of the interval under ambiguity.

References

Austen-Smith, D., and Banks, J. S. (1996). Information aggregation, rationality, and the Condorcet jury theorem. *American Political Science Review*, 90(1), 34-45.

Ellis, A. (2016). Condorcet meets Ellsberg. *Theoretical Economics,* 11(3), 865-895.

Fabrizi, S., Lippert, S., Pan, A. and Ryan, M.J. (2019a). Unanimous jury voting with an ambiguous likelihood. Unpublished working paper, University of Auckland.

Fabrizi, S., Lippert, S., Pan, A. and Ryan, M.J. (2019b). Unanimity under ambiguity. Unpublished working paper, Auckland University of Technology.

Feddersen, T., and Pesendorfer, W. (1998). Convicting the innocent: The inferiority of unanimous jury verdicts under strategic voting. *American Political Science Review*, 92(1), 23-35.

Jaffray, J. Y. (1991). Linear utility theory and belief functions: a discussion. In *Progress in Decision, Utility and Risk Theory* (pp. 221-229). Springer, Dordrecht.

McLennan, A. (1998). Consequences of the Condorcet jury theorem for beneficial information aggregation by rational agents. *American Political Science Review*, 92(2), 413-418.

Pan, A. (2019). A note on pivotality. *Games*, 10(2), 24.

Appendix A

In this Appendix we describe the derivation of Figure 1.

Let us start by observing that (6) may be written as follows:

$$V\left(\sigma_t^i, \sigma; \pi_t\right) \;=\; \pi_t + \sigma_t^i\left(\left(1-\pi_t\right)\rho_b - \left(1+c\right)\pi_t\rho_a\right)$$

Recalling (7), it follows that $V\left(\sigma_t^i, \sigma; \pi_t\right)$ is *strictly decreasing* in σ_t^i iff $\pi_t > \pi^*(\sigma)$; *strictly increasing* in σ_t^i iff $\pi_t < \pi^*(\sigma)$; and *constant* in σ_t^i iff $\pi_t = \pi^*(\sigma)$. Thus, if all the posteriors in Π_t are (strictly) on the same side of $\pi^*(\sigma)$, then voter i's response follows the same logic as in Feddersen and Pesendorfer (1998): recall footnote 9. In particular, if $\pi^*(\sigma) < \underline{\pi}_t$ then the function

$$V^*\left(\sigma_t^i, \sigma\right) \equiv \min_{\pi_t \in \Pi_t} V\left(\sigma_t^i, \sigma; \pi_t\right) \tag{15}$$

is the lower envelope of functions that are strictly decreasing in σ_t^i so V^* itself is strictly decreasing in σ_t^i. It follows that $\sigma_t^i = 0$ is the unique best response to σ (conditional on $t_i = t$) when $\pi^*(\sigma) < \underline{\pi}_t$. Conversely, if $\pi^*(\sigma) > \overline{\pi}_t$ then (15) is strictly increasing in σ_t^i so $\sigma_t^i = 1$ is the unique best response to σ (again, conditional on $t_i = t$). *This gives the green and blue regions (respectively) in Figure 1.*

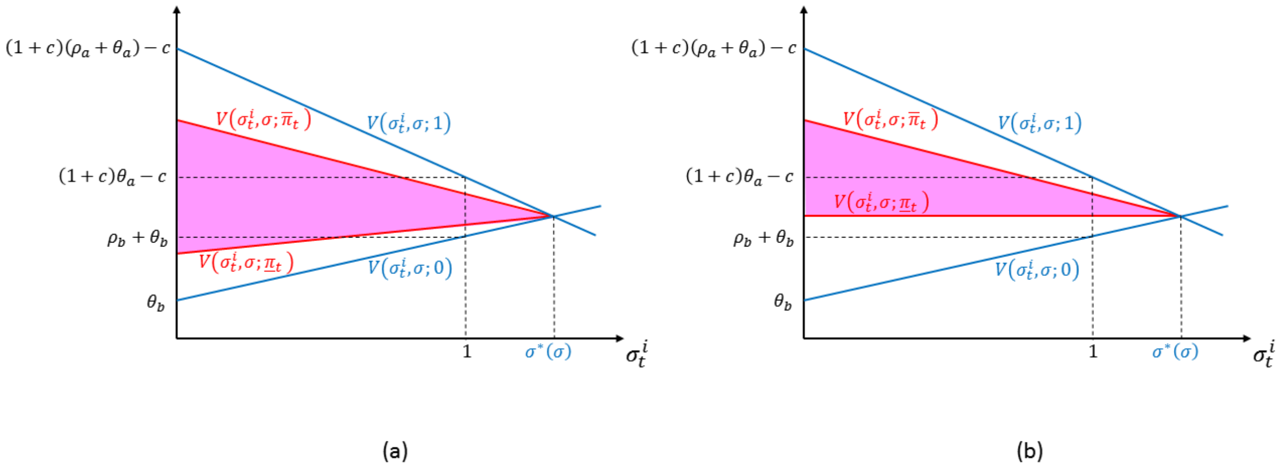


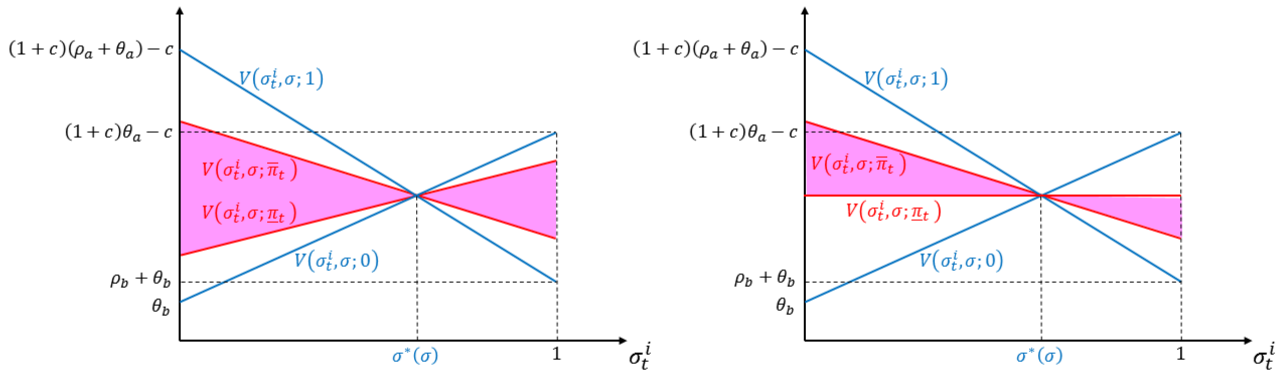
Figure 3: Best responses when $\underline{\pi}_t \leq \pi^*(\sigma) \leq \overline{\pi}_t$ and $\sigma^*(\sigma) \geq 1$. In case (a), $\underline{\pi}_t < \pi^*(\sigma)$ so $\sigma_t^i = 1$ is optimal. In case (b), $\underline{\pi}_t = \pi^*(\sigma)$ so any $\sigma_t^i \in [0,1]$ is optimal.

It remains to consider the case in which $\underline{\pi}_t \leq \pi^*(\sigma) \leq \overline{\pi}_t$: *the case described by the pink rectangle in Figure 1 and its (brown and purple) boundaries.* To analyse this case, it is convenient to re-write (6) as follows:

$$V\left(\sigma_t^i, \sigma; \pi_t\right) = \sigma_t^i \rho_b + \pi_t \left(1 - (1+c)\,\rho_a \sigma_t^i - \rho_b \sigma_t^i\right)$$

Recalling (8), we observe that

$$1 - (1+c)\,\rho_a \sigma^*(\sigma) - \rho_b \sigma^*(\sigma) = 0$$



(i) (ii)

Figure 4: Best responses when $\underline{\pi}_t \leq \pi^*(\sigma) \leq \overline{\pi}_t$ and $\sigma^*(\sigma) < 1$. In case (i), $\underline{\pi}_t < \pi^*(\sigma)$ so $\sigma_t^i = \sigma^*(\sigma)$ is optimal. In case (ii), $\underline{\pi}_t = \pi^*(\sigma)$ so any $\sigma_t^i \in [0, \sigma^*(\sigma)]$ is optimal.

and hence

$$V^*\left(\sigma_t^i, \sigma\right) = \begin{cases} V\left(\sigma_t^i, \sigma; \underline{\pi}_t\right) & \text{if } \sigma_t^i \leq \sigma^*(\sigma) \\ V\left(\sigma_t^i, \sigma; \overline{\pi}_t\right) & \text{if } \sigma_t^i \geq \sigma^*(\sigma) \end{cases}$$

In other words, $\underline{\pi}_t$ is the uniquely most "pessimistic" posterior in Π_t when $\sigma_t^i < \sigma^*(\sigma)$ and $\overline{\pi}_t$ is the uniquely most "pessimistic" posterior when $\sigma_t^i > \sigma^*(\sigma)$. Moreover, by choosing $\sigma_t^i = \sigma^*(\sigma)$ voter i can perfectly hedge against uncertainty, since $V(\sigma^*(\sigma), \sigma; \pi_t)$ is independent of π_t. The latter hedging possibility may provide strict incentives to randomise, as noted by Ellis (2016). The t-conditional best response(s) to any σ satisfying $\underline{\pi}_t \leq \pi^*(\sigma) \leq \overline{\pi}_t$ may now be characterised as follows:

- If $\sigma^*(\sigma) \geq 1$ then $V^*(\sigma_t^i, \sigma) = V(\sigma_t^i, \sigma; \underline{\pi}_t)$ for any $\sigma_t^i \in [0,1]$, so $\sigma_t^i = 1$ is uniquely optimal if $\underline{\pi}_t < \pi^*(\sigma)$, and any $\sigma_t^i \in [0,1]$ is optimal if $\underline{\pi}_t = \pi^*(\sigma)$. See Figure 3.
- If $\sigma^*(\sigma) < 1$ then $V^*(\sigma_t^i, \sigma)$ is *non-decreasing* in σ_t^i when $\sigma_t^i < \sigma^*(\sigma)$, since $\underline{\pi}_t \leq \pi^*(\sigma)$, and *non-increasing* in σ_t^i when $\sigma_t^i > \sigma^*(\sigma)$, since $\overline{\pi}_t \geq \pi^*(\sigma)$. Hence: (i) $\sigma_t^i = \sigma^*(\sigma)$ is uniquely optimal if $\underline{\pi}_t < \pi^*(\sigma) < \overline{\pi}_t$; (ii) any $\sigma_t^i \in [0, \sigma^*(\sigma)]$ is optimal if $\underline{\pi}_t = \pi^*(\sigma)$; and (iii) any $\sigma_t^i \in [\sigma^*(\sigma), 1]$ is optimal if $\overline{\pi}_t = \pi^*(\sigma)$. Figure 4 illustrates cases (i) and (ii).[20] Case (iii) is symmetric to (ii).

[20] Think of the horizontal axes in Figures 3 and 4 as the ground, with a see-saw above. The fulcrum of the see-saw is at $\sigma^*(\sigma)$ and the set Π_t determines its range of motion, with $\pi^*(\sigma)$ corresponding to the see-saw in a horizontal position, and values of π_t above (respectively, below) $\pi^*(\sigma)$ corresponding to

Appendix B

In this Appendix we prove Proposition 4.2.

We start with an important preliminary result. To facilitate its statement, let us define

$$\Gamma = \left\{ (\sigma_1, \sigma_2) \in [0,1]^2 \mid \sigma_1 \le \sigma_2 \;\text{ and }\; \sigma_2 > 0 \right\}$$

Set Γ contains all the strategies with $\sigma_1 \le \sigma_2$ excluding $\sigma = (0,0)$.

Lemma 6.1 *Let $\Omega(x) = \{\sigma \in \Gamma \mid \pi^*(\sigma) = x\}$. Then: (i) $\Omega(x) = \emptyset$ iff*

$$x \notin \left[\frac{1}{2+c}, \; \frac{r^N}{r^N + (1-r)^N (1+c)} \right];$$

(ii) $\Omega(x) = \{\sigma \in \Gamma \mid \sigma_1 = 0\}$ if

$$x = \frac{r^N}{r^N + (1-r)^N (1+c)};$$

and (iii) $\Omega(x) = \{\sigma \in \Gamma \mid \sigma_2 = \lambda(x)\sigma_1\}$ otherwise, where

$$\lambda(x) = \frac{r - (1-r)\,\ell(x)^{\frac{1}{N}}}{r\ell(x)^{\frac{1}{N}} - (1-r)} \in [1, \infty) \quad \text{and} \quad \ell(x) = \frac{1-x}{x(1+c)}.$$

Proof: Since $\sigma \in \Gamma$ we have

$$\pi^*(\sigma) = x \quad \Leftrightarrow \quad \frac{\rho_a}{\rho_b} = \ell(x) \tag{16}$$

from (7).[21] The likelihood ratio ρ_a/ρ_b is equal to

$$\left[\frac{ry + (1-r)(1-y)}{(1-r)y + r(1-y)} \right]^N$$

where $y = \sigma_1/(\sigma_1 + \sigma_2) \in \left[0, \frac{1}{2}\right]$. It follows that

$$\left(\frac{1-r}{r} \right)^N \le \frac{\rho_a}{\rho_b} \le 1$$

a downward (respectively, upward) slope from left to right. The function $V^*\left(\sigma_t^i, \sigma\right)$ describes the lower envelope of the see-saw's trajectory over the range $\sigma_t^i \in [0,1]$.

[21] In particular, there is a one-to-one relationship between $\pi^*(\sigma)$ and the likelihood ratio ρ_a/ρ_b. If $\sigma \in \Omega(x)$ then this ratio is given by $\ell(x)$.

Using these inequalities and (16) we deduce that $\Omega(x) \neq \emptyset$ iff

$$\frac{1}{2+c} \leq x \leq \frac{r^N}{r^N + (1-r)^N (1+c)}.$$

This proves (i). When

$$x = \frac{r^N}{r^N + (1-r)^N (1+c)}$$

we have

$$\ell(x) = \left(\frac{1-r}{r}\right)^N$$

Using (3)-(4) it is clear that $\sigma \in \Gamma$ satisfies

$$\frac{\rho_a}{\rho_b} = \left(\frac{1-r}{r}\right)^N$$

iff $\sigma_1 = 0$. This proves (ii). Finally, if

$$\frac{1}{2+c} \leq x < \frac{r^N}{r^N + (1-r)^N (1+c)}$$

then (3)-(4) and (16) imply

$$\pi^*(\sigma) = x \quad \Leftrightarrow \quad \sigma_2 = \lambda(x)\sigma_1.$$

with $\lambda(x) \in [1, \infty)$. This establishes (iii) and completes the proof of Lemma 6.1. $\square$

Figure 5 illustrates $\Omega(x)$.

Now consider Proposition 4.2. It is obvious from Figure 1 that $\overline{\pi}_2 \geq \underline{\pi}_1$ is necessary for the existence of a responsive, strictly mixed equilibrium: otherwise it would not be possible for both $(\underline{\pi}_1, \overline{\pi}_1)$ and $(\underline{\pi}_2, \overline{\pi}_2)$ to both lie in the rectangular region between the green and blue triangles in Figure 1. We assume $\overline{\pi}_2 \geq \underline{\pi}_1$ henceforth.

Suppose $\hat{\sigma} \in \Gamma' = \{\sigma \in \Gamma \mid 0 < \sigma_1 < \sigma_2 < 1\}$ (i.e., $\hat{\sigma}$ is strictly mixed and responsive). By inspection of Figure 1 it is evident that in order for $\hat{\sigma}$ to be an equilibrium, we must have $\overline{\pi}_2 = \pi^*(\hat{\sigma})$ or $\underline{\pi}_1 = \pi^*(\hat{\sigma})$ or both. If both conditions hold, then $\overline{\pi}_2 = \underline{\pi}_1$. Let us temporarily assume that $\overline{\pi}_2 > \underline{\pi}_1$. The case $\overline{\pi}_2 = \underline{\pi}_1$ will be considered later.

If $\pi^*(\hat{\sigma}) = \underline{\pi}_1 < \overline{\pi}_2$, then $\hat{\sigma}$ will be an equilibrium iff $\hat{\sigma}_2 = \sigma^*(\hat{\sigma})$. See Figure 6(I). Similarly, if $\pi^*(\hat{\sigma}) = \overline{\pi}_2 > \underline{\pi}_1$, then $\hat{\sigma}$ will be an equilibrium iff $\hat{\sigma}_1 = \hat{\sigma}^*(\sigma)$. See Figure 6(II). If $\pi^*(\hat{\sigma}) = x$ then the condition $\hat{\sigma}_t = \sigma^*(\hat{\sigma})$ may be expressed, using (8), as follows:

$$[(1-r)\hat{\sigma}_1 + r\hat{\sigma}_2]^N \hat{\sigma}_t = x.$$

Thus, when $\overline{\pi}_2 > \underline{\pi}_1$ there exists a strictly mixed responsive equilibrium iff at least one of the following two conditions is met:

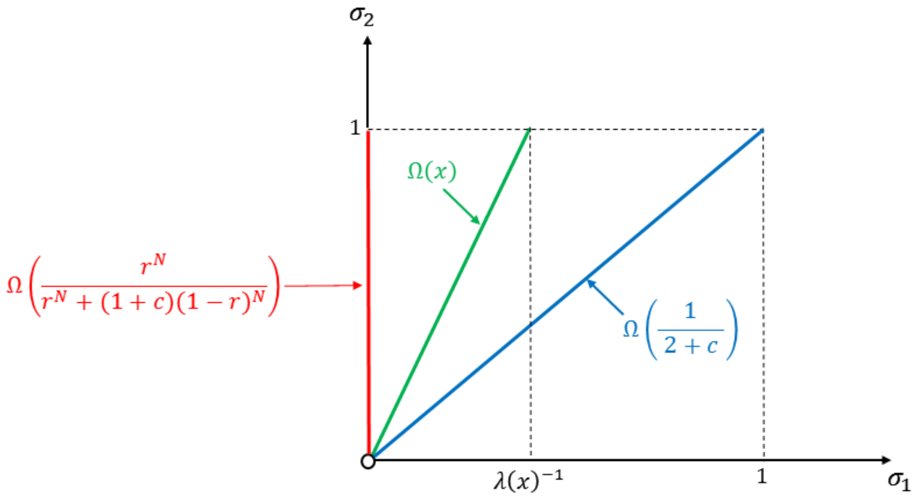


Figure 5: Note that the points in $\Omega(x)$ form a line, and this line gets steeper as x increases

(I′) There exists $\sigma \in \Omega(\underline{\pi}_1) \cap \Gamma'$ satisfying

$$[(1-r)\sigma_1 + r\sigma_2]^N \sigma_2 = \underline{\pi}_1 \tag{CI}$$

(II′) There exists $\sigma \in \Omega(\overline{\pi}_2) \cap \Gamma'$ satisfying

$$[(1-r)\sigma_1 + r\sigma_2]^N \sigma_1 = \overline{\pi}_2 \tag{CII}$$

Consider condition (I′). Note first, using Figure 5, that $\Omega(\underline{\pi}_1) \cap \Gamma' \neq \emptyset$ iff

$$\frac{1}{2+c} < \underline{\pi}_1 < \frac{r^N}{r^N + (1-r)^N (1+c)} \tag{17}$$

Now consider Figure 7. The set Γ' is the *interior* of the red triangle. The line corresponding to $\Omega(\underline{\pi}_1)$ is indicated, under the assumption that $\underline{\pi}_1$ satisfies (17). The other locus in Figure 7 describes the set of solutions to (CI): it is easily verified that the implicit function defined by (CI) is convex with a strictly negative slope, and that

$$\sigma_1 = \sigma_2 = \underline{\pi}_1^{\frac{1}{N+1}}$$

lies on the graph of this function. Condition (I′) requires that these two loci intersect in Γ'. Such an intersection will obviously be unique if it exists. An intersection in Γ' will

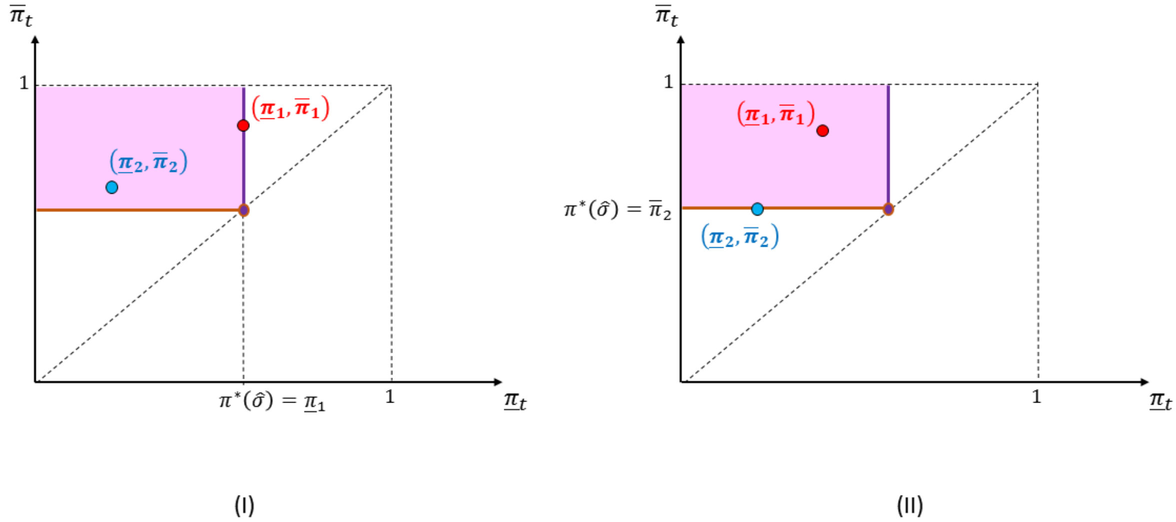


Figure 6: Strictly mixed responsive equilibria when $\overline{\pi}_2 > \underline{\pi}_1$

obtain iff the $\Omega(\underline{\pi}_1)$ locus hits the top of the red triangle *strictly to the right* of the locus of solutions to (CI), as illustrated in Figure 7.

Recall from Figure 5 that the locus $\Omega(\underline{\pi}_1)$ hits the top of the red triangle at $\lambda(\underline{\pi}_1)^{-1}$. Given $x \in (0,1)$, let $h_I(x)$ denote the value of σ_1 that solves

$$[(1-r)\,\sigma_1 + r]^N = x \tag{18}$$

(noting that this solution might be negative).[22] *Hence, condition (I′) is satisfied iff (17) and*

$$h_I(\underline{\pi}_1) < \frac{1}{\lambda(\underline{\pi}_1)} \tag{19}$$

Using the fact that $h_I(x)$ is strictly increasing we may re-write condition (19) as follows:

$$\begin{aligned}
\left[(1-r)\,\lambda(\underline{\pi}_1)^{-1} + r\right]^N > \underline{\pi}_1 \quad &\Leftrightarrow \quad \frac{2r-1}{r-(1-r)\,\ell(\underline{\pi}_1)^{1/N}} > \underline{\pi}_1^{1/N} \\
&\Leftrightarrow \quad r\underline{\pi}_1^{1/N} - (1-r)\left(\frac{1-\underline{\pi}_1}{1+c}\right)^{1/N} < 2r-1 \qquad (20)
\end{aligned}$$

where the first equivalence uses the definition of $\lambda(\underline{\pi}_1)$ and the second uses the definition of $\ell(\underline{\pi}_1)$. It is obvious that the left-hand side of (20) is strictly increasing in $\underline{\pi}_1$ so there

[22] That is, the locus of solutions to (CI) may hit the left-hand edge of the triangle rather than the top edge.

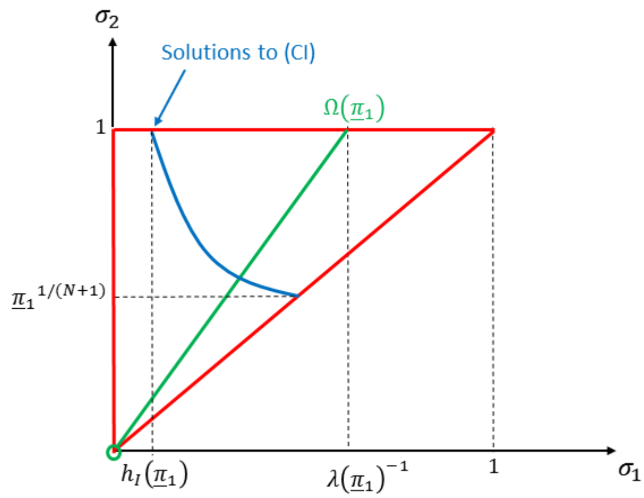


Figure 7: Strategy σ satisfies (I′) iff it lies in the interior of the red triangle at the intersection of the green and blue curves

is some $\beta_1(N)$ such that (20) is equivalent to $\underline{\pi}_1 < \beta_1(N)$, where the notation emphasises the dependence of this upper bound on N. Letting

$$\alpha_1(N) = \min\left\{\beta_1(N)\,,\ \frac{r^N}{r^N + (1-r)^N(1+c)}\right\}$$

we have therefore shown that (I′) holds iff (I). (We will verify the stated properties of α_1 shortly.)

Next, consider condition (II′). We have $\Omega(\overline{\pi}_2) \cap \Gamma' \neq \emptyset$ iff

$$\frac{1}{2+c} < \overline{\pi}_2 < \frac{r^N}{r^N + (1-r)^N(1+c)} \tag{21}$$

This time $\sigma \in \Omega(\overline{\pi}_2) \cap \Gamma'$ is an equilibrium if it sits at the intersection of $\Omega(\overline{\pi}_2)$ and the locus defined by (CII). Given $x \in (0,1)$, let $h_{II}(x)$ denote the value of σ_1 that solves

$$[(1-r)\sigma_1 + r]^N \sigma_1 = x \tag{22}$$

It is obvious that this solution exists and is unique. Moreover, $h_{II}(x) \in (0,1)$ for any $x \in (0,1)$. Reasoning as for condition (I′), *mutatis mutandis*, we deduce that: *condition (II′) is satisfied iff (21) and*

$$h_{II}(\overline{\pi}_2) < \frac{1}{\lambda(\overline{\pi}_2)} \tag{23}$$

This time, we use the facts that $\lambda(x)$ is strictly increasing and $h_{II}(\overline{\pi}_2) \in (0,1)$ to re-write condition (23) as follows:

$$\begin{aligned}
\pi^*((h_{II}(\overline{\pi}_2),1)) > \overline{\pi}_2 \quad &\Leftrightarrow \quad \left[\frac{rh_{II}(\overline{\pi}_2)+(1-r)}{(1-r)h_{II}(\overline{\pi}_2)+r}\right]^N < \ell(\overline{\pi}_2) \\
&\Leftrightarrow \quad [rh_{II}(\overline{\pi}_2)+(1-r)]^N h_{II}(\overline{\pi}_2) < \overline{\pi}_2 \ell(\overline{\pi}_2) \\
&\Leftrightarrow \quad [rh_{II}(\overline{\pi}_2)+(1-r)]^N h_{II}(\overline{\pi}_2) < \frac{1-\overline{\pi}_2}{1+c} \qquad (24)
\end{aligned}$$

where the second equivalence uses the definition of $h_{II}(\overline{\pi}_2)$. Since $h_{II}(x)$ is strictly increasing, it is easy to see that there is some $\beta_2(N)$ such that (24) is equivalent to $\overline{\pi}_2 < \beta_2(N)$. Letting

$$\alpha_2(N) = \min\left\{\beta_2(N),\ \frac{r^N}{r^N+(1-r)^N(1+c)}\right\}$$

we deduce that (II′) holds iff (II).

Let us now verify the stated properties of the functions α_1 and α_2. Comparing (22) and (18), it is obvious that $h_I(x) < h_{II}(x)$ for all $x \in (0,1)$. Thus, if $\overline{\pi}_2 = z$ satisfies (23) then $\underline{\pi}_1 = z$ satisfies (19). It follows that $\beta_1(N) \geq \beta_2(N)$. To see why $\beta_2(N) > (2+c)^{-1}$, set $\overline{\pi}_2 = (2+c)^{-1}$ in (24) to obtain:

$$\left[rh_{II}\left(\frac{1}{2+c}\right)+(1-r)\right]^N h_{II}\left(\frac{1}{2+c}\right) < \frac{1}{2+c}.$$

This inequality must hold since, by the definition of h_{II}, we have:

$$\begin{aligned}
&\left[(1-r)h_{II}\left(\frac{1}{2+c}\right)+r\right]^N h_{II}\left(\frac{1}{2+c}\right) = \frac{1}{2+c} \\
\Rightarrow \quad &\left[rh_{II}\left(\frac{1}{2+c}\right)+(1-r)\right]^N h_{II}\left(\frac{1}{2+c}\right) < \frac{1}{2+c}
\end{aligned}$$

where we have used $r > \frac{1}{2}$ and $h_{II}\left((2+c)^{-1}\right) \in (0,1)$. Hence, $\beta_2(N) > (2+c)^{-1}$ and we have therefore established that $\alpha_1(N) \geq \alpha_2(N) > (2+c)^{-1}$ for all N.

Finally, let us return to the scenario in which $\overline{\pi}_2 = \underline{\pi}_1$. Then σ is a strictly mixed responsive equilibrium iff

$$\sigma \in \Omega(\underline{\pi}_1) \cap \Gamma'$$

and $\sigma_1 \leq \sigma^*(\sigma) \leq \sigma_2$. See Figure 8(a). Note that

$$\sigma_1 \leq \sigma^*(\sigma) \leq \sigma_2 \quad \Leftrightarrow \quad [(1-r)\sigma_1 + r\sigma_2]^N \sigma_1 \leq \underline{\pi}_1 \leq [(1-r)\sigma_1 + r\sigma_2]^N \sigma_2$$

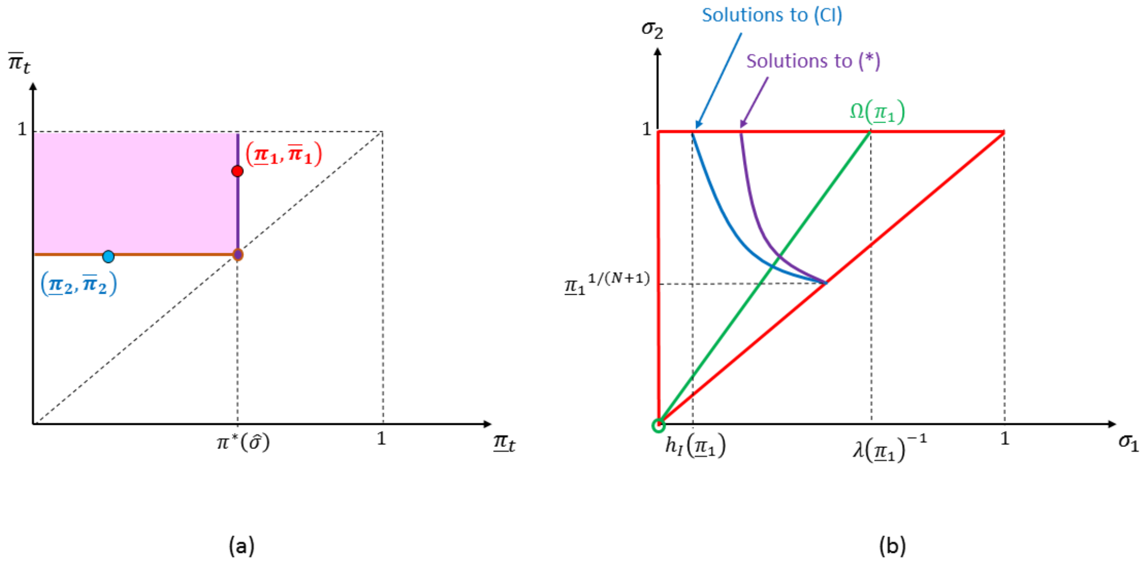


Figure 8: The case $\overline{\pi}_2 = \underline{\pi}_1$.

In other words, σ must lie in the region whose western boundary is the locus determined by (CI) and whose eastern boundary is the locus described by the equation

$$[(1-r)\,\sigma_1 + r\sigma_2]^N \,\sigma_1 = \underline{\pi}_1 \tag{$*$}$$

See Figure 8(b). This region has a non-empty intersection with $\Omega\left(\underline{\pi}_1\right) \cap \Gamma'$ iff (19) holds, which is equivalent to (I).

Appendix C

In this Appendix we prove Lemma 4.5.

First, let $\varepsilon > 0$ be small enough to satisfy[23]

$$r \ln\left(\frac{1}{2+c} + \varepsilon\right) < (1-r)\ln\left(\frac{1}{2+c} - \frac{\varepsilon}{1+c}\right) \tag{25}$$

[23]There exists such an ε since setting $\varepsilon = 0$ in (25) gives:

$$(2r-1)\ln\left(\frac{1}{2+c}\right) < 0$$

which is obviously true.

We claim that

$$\underline{\pi}_1 = \frac{1}{2+c} + \varepsilon \tag{26}$$

(and hence any smaller value of $\underline{\pi}_1$) satisfies (20) when N is sufficiently large, and hence that

$$\beta_1(N) \geq \frac{1}{2+c} + \varepsilon$$

for sufficiently large N. To verify this claim, substitute (26) into (20) to get

$$r\left(\frac{1}{2+c} + \varepsilon\right)^{1/N} - (1-r)\left(\frac{1}{2+c} - \frac{\varepsilon}{1+c}\right)^{1/N} \ < \ 2r - 1 \tag{27}$$

Let $g(N)$ denote the left-hand side of (27). It is obvious that

$$\lim_{N\to\infty} g(N) \ = \ 2r - 1 \tag{28}$$

To prove our claim, it suffices to show that g is *strictly increasing* when N is sufficiently large, since g then approaches this limit *from below* and it follows that (27) must hold for large N.

To see that g is strictly increasing for large N, let

$$\overline{g}(x) = r\left(\frac{1}{2+c} + \varepsilon\right)^{x} - (1-r)\left(\frac{1}{2+c} - \frac{\varepsilon}{1+c}\right)^{x}.$$

Then

$$\overline{g}'(x) = r\left(\frac{1}{2+c} + \varepsilon\right)^{x} \ln\left(\frac{1}{2+c} + \varepsilon\right) - (1-r)\left(\frac{1}{2+c} - \frac{\varepsilon}{1+c}\right)^{x} \ln\left(\frac{1}{2+c} - \frac{\varepsilon}{1+c}\right)$$

Hence:

$$\lim_{x\to 0} \overline{g}'(x) \ = \ r\ln\left(\frac{1}{2+c} + \varepsilon\right) - (1-r)\ln\left(\frac{1}{2+c} - \frac{\varepsilon}{1+c}\right)$$

which is strictly less than zero by (25). It follows that $g(N)$ is strictly increasing for large N.

Using the fact that $r \in \left(\frac{1}{2}, 1\right)$ we have

$$\lim_{N\to\infty} \frac{r^N}{r^N + (1-r)^N(1+c)} \ = \ \lim_{N\to\infty} \frac{1}{1 + \left(\frac{1-r}{r}\right)^N (1+c)} = 1 \tag{29}$$

and hence

$$\alpha_1(N) \geq \frac{1}{2+c} + \varepsilon$$

for all sufficiently large N. Since $\alpha_1(N) > (2+c)^{-1}$ for *all* N, there must exist some $\eta > 0$ such that

$$\alpha_1(N) \geq \frac{1}{2+c} + \eta$$

for all N.

Appendix D

In this Appendix we prove Proposition 5.2.

By Lemmas 4.1 and 4.4, if N is sufficiently large, any responsive equilibrium with $\sigma_2 = 1$ must have $\sigma_1 \in (0,1)$. Hence, if $(\sigma_1, 1)$ is a responsive equilibrium and N is sufficiently large, then

$$\underline{\pi}_1 \leq \pi^*((\sigma_1, 1)) \leq \overline{\pi}_1 \tag{30}$$

(see Figure 1). Recalling (29), we have

$$\overline{\pi}_1 < \frac{r^N}{r^N + (1-r)^N(1+c)}$$

for sufficiently large N. Combining this fact with (30) and Figure 5 we see that

$$0 < \lambda^{-1}(\overline{\pi}_1) \leq \sigma_1$$

in any responsive equilibrium with $\sigma_2 = 1$, provided N is sufficiently large. In such an equilibrium:[24]

$$\Pr(\mathcal{C}|a) \ \geq \ \left[r\lambda(\overline{\pi}_1)^{-1} + (1-r)\right]^{N+1} \ = \ \left[\frac{(2r-1)\,\ell(\overline{\pi}_1)^{\frac{1}{N}}}{r-(1-r)\,\ell(\overline{\pi}_1)^{\frac{1}{N}}}\right]^{N+1} > 0$$

[24] Using (1) we see that

$$\frac{1}{c+1} = \frac{1-q}{q}$$

so

$$\left[\frac{(2r-1)\,\ell(\overline{\pi}_1)^{\frac{1}{N}}}{r-(1-r)\,\ell(\overline{\pi}_1)^{\frac{1}{N}}}\right]^{N+1} = \left[\frac{(2r-1)\left(\frac{(1-q)(1-\overline{\pi}_1)}{q\overline{\pi}_1}\right)^{\frac{1}{N}}}{r-(1-r)\left(\frac{(1-q)(1-\overline{\pi}_1)}{q\overline{\pi}_1}\right)^{\frac{1}{N}}}\right]^{N+1}.$$

If $\overline{\pi}_1 = r$ this is identical to the expression for $l_I(r, g, N+1)$ on p.26 of Feddersen and Pesendorfer (1998). Moreover, if $\overline{\pi}_1 = r$ then condition (31) is equivalent to $q > 1-r$, which is necessary in Feddersen and Pesendorfer's model to ensure the existence of a responsive equilibrium; otherwise, if $q \leq 1-r$, then $\sigma = (1,1)$ is an equilibrium of the Feddersen and Pesendorfer model. Likewise, when (31) is violated in our model, $\sigma = (1,1)$ is an equilibrium but no responsive equilibrium exists.

where the final inequality uses $\lambda^{-1}(\overline{\pi}_1) > 0$. It follows that

$$\ell(\overline{\pi}_1) < 1 \quad \Leftrightarrow \quad \overline{\pi}_1 > \frac{1}{2+c} \tag{31}$$

in any responsive equilibrium with $\sigma_2 = 1$ and large N:[25] if $\ell(\overline{\pi}_1) \geq 1$, then $2r-1 > 0$ and

$$\left[\frac{(2r-1)\,\ell(\overline{\pi}_1)^{\frac{1}{N}}}{r-(1-r)\,\ell(\overline{\pi}_1)^{\frac{1}{N}}}\right]^{N+1} > 0$$

would imply $r-(1-r)\,\ell(\overline{\pi}_1)^{\frac{1}{N}}$ and hence

$$\Pr(\mathcal{C}|a) \;\geq\; \left[\frac{(2r-1)\,\ell(\overline{\pi}_1)^{\frac{1}{N}}}{r-(1-r)\,\ell(\overline{\pi}_1)^{\frac{1}{N}}}\right]^{N+1} \;\geq\; \left[\frac{(2r-1)\,\ell(\overline{\pi}_1)^{\frac{1}{N}}}{r\ell(\overline{\pi}_1)^{\frac{1}{N}}-(1-r)\,\ell(\overline{\pi}_1)^{\frac{1}{N}}}\right]^{N+1} = 1$$

which is impossible (since conviction cannot be certain, conditional on $s = a$, in an equilibrium with $\sigma_1 < 1$).

To complete the proof we will show that

$$\left[\frac{(2r-1)\,\ell(\overline{\pi}_1)^{\frac{1}{N}}}{r-(1-r)\,\ell(\overline{\pi}_1)^{\frac{1}{N}}}\right]^{N+1} = \left[\frac{(2r-1)}{r-(1-r)\,\ell(\overline{\pi}_1)^{\frac{1}{N}}}\right]^{N+1} \ell(\overline{\pi}_1)^{\frac{N+1}{N}} \tag{32}$$

has a strictly positive limit as $N \to \infty$ when $0 < \ell(\overline{\pi}_1) < 1$. Since

$$\lim_{N\to\infty} \ell(\overline{\pi}_1)^{\frac{N+1}{N}} \;=\; \ell(\overline{\pi}_1) > 0$$

the limit of (32) is strictly positive iff

$$\lim_{N\to\infty} \left[\frac{(2r-1)}{r-(1-r)\,\ell(\overline{\pi}_1)^{\frac{1}{N}}}\right]^{N+1} \;>\; 0 \tag{33}$$

We follow the logic on p.32 of Feddersen and Pesendorfer (1998) to establish (33) as follows:

$$\begin{aligned}
\lim_{N\to\infty} \left[\frac{(2r-1)}{r-(1-r)\,\ell(\overline{\pi}_1)^{\frac{1}{N}}}\right]^{N+1} &\geq \lim_{N\to\infty} \left[\frac{(2r-1)}{r-(1-r)\left[1-(N+1)^{-1}\ln\ell(\overline{\pi}_1)\right]}\right]^{N+1} \\
&= \lim_{N\to\infty} \left[1+(1-r)(N+1)^{-1}\right]^{-(N+1)} \\
&= \exp\left[-\left(\frac{1-r}{2r-1}\right)\ln\ell(\overline{\pi}_1)\right] \\
&= \ell(\overline{\pi}_1)^{-(1-r)/(2r-1)}
\end{aligned}$$

[25] In fact, Fabrizi et al. (2019b) show that this is necessary for any N.

where the inequality and second equality use Feddersen and Pesendorfer's (1998) expressions (6) and (5) respectively.

Appendix E

In this Appendix we prove Proposition 5.3.

Suppose $\sigma^{(k)}$ is a strictly mixed responsive equilibrium for a jury of size $N_k \in \{1, 2, ...\}$, with $N_k \to \infty$ as $k \to \infty$. It follows that $\underline{\pi}_1 \leq \overline{\pi}_2$. From the proof of Proposition 4.2, each $\sigma^{(k)}$ satisfies

$$\left[(1-r)\,\sigma_1^{(k)} + r\sigma_2^{(k)}\right]^{N_k} \sigma_2^{(k)} = \underline{\pi}_1$$

or

$$\left[(1-r)\,\sigma_1^{(k)} + r\sigma_2^{(k)}\right]^{N_k} \sigma_1^{(k)} = \overline{\pi}_2.$$

Therefore $\sigma_t^{(k)} \to 1$ as $k \to \infty$ for each $t \in T$, and hence

$$\liminf_{k\to\infty} \left[(1-r)\,\sigma_1^{(k)} + r\sigma_2^{(k)}\right]^{N_k} \geq \underline{\pi}_1.$$

Since

$$\lim_{k\to\infty} \left[(1-r)\,\sigma_1^{(k)} + r\sigma_2^{(k)}\right] = 1$$

we have

$$\liminf_{k\to\infty} \left[(1-r)\,\sigma_1^{(k)} + r\sigma_2^{(k)}\right]^{N_k+1} \geq \underline{\pi}_1.$$

Note that, for equilibrium $\sigma^{(k)}$,

$$\Pr\left(\mathcal{C}|b\right) = \left[(1-r)\,\sigma_1^{(k)} + r\sigma_2^{(k)}\right]^{N_k+1}$$

so the probability of conviction in $\sigma^{(k)}$ is at least

$$(1-\overline{p})\left[(1-r)\,\sigma_1^{(k)} + r\sigma_2^{(k)}\right]^{N_k+1}$$

for any $p \in \left[\underline{p}, \overline{p}\right]$. Thus:

$$\min_{p\in\left[\underline{p},\overline{p}\right]} \left[\liminf_{k\to\infty} \gamma_k\left(p\right)\right] \geq (1-\overline{p})\,\underline{\pi}_1 > 0.$$