A model of jury decisions where all jurors have the same evidence

1 May 2004

Forthcoming in *Synthese* / *Knowledge, Rationality and Action*

Franz Dietrich
Philosophy, Probability and Modelling Group
Center for Junior Research Fellows
University of Konstanz
78457 Konstanz, Germany
franz.dietrich@uni-konstanz.de

Christian List
Department of Government
London School of Economics
Houghton Street
London WC2A 2AE, U.K.
c.list@lse.ac.uk

Acknowledgments: Previous versions of this paper were presented at the Summer School on Philosophy and Probability, University of Konstanz, September 2002, and at the GAP5 Workshop on Philosophy and Probability, University of Bielefeld, September 2003. We thank the participants at these events and Luc Bovens, Branden Fitelson, Jon Williamson and the anonymous reviewers of this paper for comments and discussion. Franz Dietrich also thanks the Alexander von Humboldt Foundation, the German Federal Ministry of Education and Research, and the Program for the Investment in the Future (ZIP) of the German Government, for supporting this research.

A model of jury decisions where all jurors have the same evidence

Abstract. Under the independence and competence assumptions of Condorcet's classical jury model, the probability of a correct majority decision converges to certainty as the jury size increases, a seemingly unrealistic result. Using Bayesian networks, we argue that the model's independence assumption requires that the state of the world (guilty or not guilty) is the latest common cause of all jurors' votes. But often – arguably in all courtroom cases and in many expert panels – the latest such common cause is a shared 'body of evidence' observed by the jurors. In the corresponding Bayesian network, the votes are direct descendants not of the state of the world, but of the body of evidence, which in turn is a direct descendant of the state of the world. We develop a model of jury decisions based on this Bayesian network. Our model permits the possibility of misleading evidence, even for a maximally competent observer, which cannot easily be accommodated in the classical model. We prove that (i) the probability of a correct majority verdict converges to the probability that the body of evidence is not misleading, a value typically below 1; (ii) depending on the required threshold of 'no reasonable doubt', it may be impossible, even in an arbitrarily large jury, to establish guilt of a defendant 'beyond any reasonable doubt'.

Key words: Condorcet jury theorem, Bayesian networks, Parental Markov condition, conditional independence, interpretation of evidence

1. Introduction

Suppose a jury (committee, expert panel etc.) has to determine whether or not a defendant is guilty (whether or not some factual proposition is true). There are two possible states of the world: $x = 1$ (the defendant is guilty) and $x = 0$ (the defendant is not guilty). Given that the state of the world is x, each juror has the same probability (competence) $p > 1/2$ of voting for x and the votes of different jurors are independent from each other. Then the probability that a majority of jurors votes for x, given the state of the world x, converges to 1 as the number of jurors increases. This is the classical Condorcet jury theorem (e.g. Grofman, Owen and Feld 1983). The theorem implies that the reliability of majority decisions can be made arbitrarily close to certainty by increasing the jury size.

This result may seem puzzling. What if all jurors are tricked by the same evidence, which seems ever so compelling? What if, against all odds, the wind blows an innocent person's hair to the crime scene and the jurors believe that it could not have arrived there without the person's presence? What if the evidence is so confusing that, no matter how many jurors are consulted, there is not enough evidence to solve a case conclusively?

The classical Condorcet jury theorem suggests that we can rule out such scenarios by increasing the jury size sufficiently. Suppose each juror views the crime scene from a different perspective and obtains a separate item of evidence about the state of the world.

This requires that, for each additional juror, a new independent item of evidence is available. So there must exist arbitrarily many items of evidence as the jury size tends to infinity, which are confirmationally independent regarding the hypothesis that the defendant is guilty (on confirmational independence, see Fitelson 2001). Call this case A. Then the jury would be able to reach a correct decision with a probability approaching 1, by aggregating arbitrarily many independent items of evidence into a single verdict. But often there are not arbitrarily many independent items of evidence. Rather, the jury as a whole reviews the same body of evidence, such as that presented in the courtroom, which does not increase with the jury size. Each juror has to decide whether he or she believes that this evidence supports the hypothesis that the defendant is guilty. Call this case B. Arguably, decisions in most real-world juries and many committees and expert panels are instances of case B. Moreover, in most legal systems, there are 'rules of evidence' specifying what evidence is admissible in a court's decision and what evidence is not. Jurors are legally required to use only the evidence presented in the courtroom (typically the only evidence about a case jurors come to see) and to ignore any evidence obtained through other channels (in those rare cases where they have such evidence).

We argue that, while case A might satisfy the conditions of the classical Condorcet jury theorem, case B does not. We represent each case using Bayesian networks (Pearl 2000; Bovens and Olsson 2000; Corfield and Williamson 2001). Case A satisfies Condorcet's independence assumption, so long as a demanding condition holds: The state of the world is the latest common cause of the jurors' votes. In the corresponding Bayesian network, votes are direct causal descendants of the state of the world. This assumption, although implicit in the classical Condorcet jury model, is not usually acknowledged. Case B, by contrast, violates the classical independence assumption, as there exists an intermediate common cause between the state of the world and the jurors' votes, namely the body of evidence. In the corresponding Bayesian network, the jurors' votes are direct descendants of the body of evidence, which in turn is a direct descendant of the state of the world. This dependency structure has radical implications for the Condorcet jury theorem. The model developed in this paper is based on the Bayesian network of case B.

The main novelty of our model is that different jurors are independent not conditional on the state of the world, but conditional on the evidence. This follows from the requirement, formulated in terms of the Parental Markov Condition (defined below), that independence should be assumed conditional on the latest common cause. While in case A the latest common cause of the jurors' votes is the state of the world, in case B it is the shared body of evidence. Our model shows that, irrespective of the jury size and juror competence, the overall jury reliability at best approaches the probability that the evidence is not misleading, i.e. the probability that the evidence points to the truth from the perspective of a maximally competent 'ideal' observer, a value typically below one. We prove further that, depending on the required threshold of 'no reasonable doubt', it may be impossible, even in an arbitrarily large jury and even when there is unanimity, to establish guilt of a defendant 'beyond any reasonable doubt'. The results imply that, if real-world jury, committee or expert panel decisions are more similar to case B than to case A, the classical Condorcet jury theorem fails to apply to such decisions.

Previous work on dependencies between jurors' votes has focused on, first, opinion leaders – jurors who influence other jurors – (Grofman, Owen and Feld 1983; Nitzan and Paroush 1984; Owen 1986; Boland 1989; Boland, Proschan and Tong 1989; Estlund 1994) and, secondly, a lack of free speech that makes votes dependent on a few dominant 'schools of thought' (e.g. Lahda 1992). These sources of dependence differ from the one in our model. In the first case, the votes themselves are causally interdependent. In the second, some votes have an additional common cause: a common 'school of thought' that is independent from the state of the world. But in both cases, unlike in our model, votes are still direct descendants of the state of the world. As a consequence, existing models with dependencies have preserved the result that the probability of a correct majority decision converges to 1 as the jury size increases, so long as different jurors' votes are not too highly correlated. Further, these models do not impose an upper bound on the total evidence available to the jury, and they usually suggest that the difference between Condorcet's classical model and one with dependencies lies in a different (slower) convergence rate, but not in a different limit, as in our model. By contrast, the dependency structure of our Bayesian network model has been unexplored so far.

2. The model

2.1 The classical jury model

There are n jurors, labelled $i = 1, 2, ..., n$. The state of the world is represented by a binary variable X taking the values 0 (not guilty) or 1 (guilty). The jurors' votes are represented by the binary random variables $V_1, V_2, ..., V_n$. Each V_i takes the values 0 (a 'not guilty' vote) or 1 (a 'guilty' vote). A juror i's judgment is correct if and only if the value of V_i coincides with that of X. We use capital letters to denote random variables and small letters to denote particular values. Condorcet's classical model assumes the following.[1]

Independence Given the State of the World (I|X). The votes $V_1, V_2, ..., V_n$ are independent from each other, conditional on the state of the world X.

This implicitly assumes that each juror's vote is directly probabilistically caused by the state of the world,[2] and is therefore independent from the other jurors' votes once the state of the world is given.

Competence Given the State of the World (C|X). For each state of the world $x \in \{0, 1\}$ and all jurors $i = 1, 2, ..., n$, $p = P(V_i=x|X=x) > 1/2$.

Each juror's vote is thus a signal about the state of the world, where the signal is noisy, but biased towards the truth, as $p > 1/2$. The Condorcet jury theorem states that majority voting over such independent signals reduces the noise. More precisely, let $V = \Sigma_{i=1,...,n} V_i$ be the number of votes for 'guilty'. Then $V > n/2$ means that there is a majority for 'guilty', and $V < n/2$ means that there is a majority for 'not guilty'.

Theorem 1. (Condorcet jury theorem) *If (I|X) and (C|X) hold, then* $P(V>n/2|X=1)$ *and* $P(V<n/2|X=0)$ *converge to* 1 *as n tends to infinity*.[3]

2.2 Bayesian networks

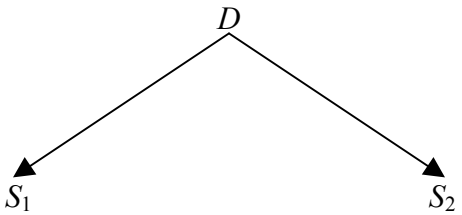
Bayesian networks can graphically represent the (probabilistic) causal relations between the different variables such as X and $V_1, V_2, ..., V_n$. A *Bayesian network* is a directed

acyclic graph, consisting of a finite number of nodes and arrows. The nodes represent the variables, and arrows (→) between nodes represent direct causal dependencies.[4] The direction of an arrow represents the direction of causality. For example, a connection of the form $X \rightarrow V_1$ means 'X (directly) causally affects V_1'. Here X is a *parent* of V_1, and V_1 is a *child* of X. One node is a *descendant* of another, the *ancestor*, if there exists a sequence of arrows connecting the two nodes, where each arrow points away from the ancestor node and towards the descendant node. One node is a *non-descendant* of another if there exists no such sequence. So the descendant relation is the transitive closure of the child relation. *Acyclicity* of the graph means that no node is its own descendant. A *Bayesian tree* is a Bayesian network in which every variable has at most one parent. Many joint probability distributions of the variables at the nodes are consistent with a given Bayesian network. Here, consistency with the network means that the following condition is satisfied (for details on Bayesian networks, see Pearl 2000, ch. 1):

Parental Markov Condition (PM). Any variable is independent from its non-descendants (except itself), conditional on its parents.[5]

For example, consider a medical condition (say a flu) that can cause two symptoms in a patient (a sore throat and a fever). Consider the Bayesian tree of diagram 1, which contains three variables D, S_1 and S_2, each of which takes the value 0 or 1: D is 1 if the patient has the condition and 0 otherwise; S_1 is 1 if the patient has the first symptom and 0 otherwise; and S_2 is 1 if the patient has the second symptom and 0 otherwise.

Diagram 1: A simple Bayesian tree



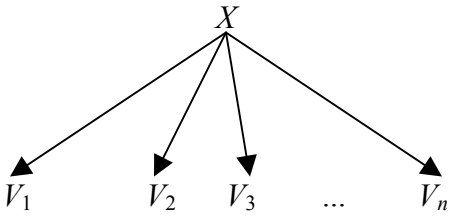
This Bayesian tree, in which the symptoms S_1 and S_2 are direct descendants of condition D, expresses that both symptoms are direct consequences of condition D, rather than being commonly caused by some intermediate symptom S of the condition. The two symptoms are not independent unconditionally: A sore throat increases the chance of

having a flu, which in turn increases the chance of having a fever. The Parental Markov Condition says that the two symptoms are independent conditional on their common cause: Given that you have a flu ($D = 1$), having a sore throat and having a fever are independent from each other, and given that you have no flu ($D = 0$) having a sore throat and having a fever are also independent from each other.

2.3 The classical jury model revisited

Diagram 2 shows the Bayesian tree corresponding to the classical Condorcet jury model.

Diagram 2: Bayesian tree for the classical Condorcet jury model



The votes V_1, V_2, ..., V_n are non-descendants of each other and each have X as a parent. So the Parental Markov Condition holds if and only if V_1, V_2, ..., V_n are independent from each other, conditional on X, which is exactly the independence condition of the classical jury model. So an alternative statement of that model can be given in terms of the Bayesian tree in diagram 2 together with conditions (PM) and (C|X).

The Bayesian tree in diagram 2 has the property that the state of the world X is the latest common cause of the jurors' votes. In case B in the introduction, this property is violated. So, if real-world jury decisions are more like case B than case A, they are not adequately captured by the classical model.

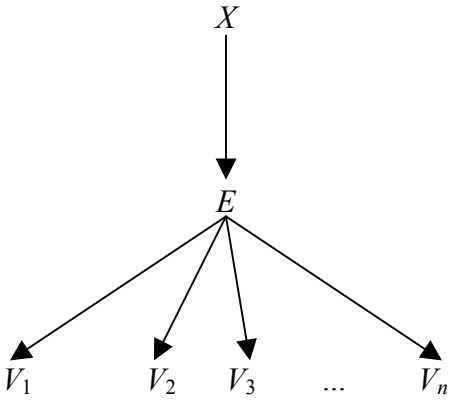
2.4 The new model

The new model gives up the assumption that the state of the world is the latest common cause of the jurors' votes. Instead, we assume that there exists an intermediate common cause between the state of the world and the votes. For simplicity, we describe that intermediate common cause as the *body of evidence*.

To illustrate why introducing a common body of evidence creates a dependency between votes that contradicts Condorcet's independence assumption, imagine that you, an external observer, know that the defendant is truly guilty, and you learn that the first 10 jurors have wrongly voted for 'not guilty'. From this, you will infer that the jurors' common evidence is highly misleading, which in turn implies that the 11th juror is also likely to vote for 'not guilty'. This contradicts the classical condition of independence given the state of the world, according to which the first 10 votes provide no information for predicting the 11th vote once you know what the true state of the world is.

We represent the common body of evidence by a random variable, E, which takes values in some set $\mathbf{E}$. Diagram 3 shows the Bayesian tree corresponding to the new model.

Diagram 3: Bayesian tree for the new model



The value of E can be interpreted as the totality of available information about the state of the world the jurors are exposed to, including the testimony of witnesses, jury deliberation, the appearance of the defendant in court (relaxed or stressed, smiling or serious etc.). In Bayesian tree terms, E is a child of the state of the world and a parent of the jurors' votes. What matters is not the particular nature of E, which will usually be complex, but the fact that every juror is exposed to the same body of evidence.[6] We do not make any particular assumption about the set $\mathbf{E}$ of possible bodies of evidence, which may be finite, countably infinite, or even uncountably infinite.

The probability distribution of E depends on the state of the world: The distribution of E given guilt ($X = 1$) is different from that given innocence ($X = 0$). In the case of guilt, it is

usually more likely that the body of evidence will point towards guilt than in the case of innocence. For instance, the defendant might be more likely to fail a lie detector test in the case of guilt than in the case of innocence. We prove that the Parental Markov Condition, when applied to the Bayesian tree in diagram 3, has two implications:

Common Signal (S). The joint probability distribution of the votes V_1, V_2, …,V_n given both the evidence E and the state of the world X is the same as that given just the evidence E.

So, the votes are only indirectly caused by the state of the world: They depend on the state of the world only through the body of evidence. Once the evidence is given, what the state of the world is makes no difference to the probabilities of the jurors' votes.[7]

Independence Given the Evidence (I|E). The votes V_1, V_2, ..., V_n are independent from each other, conditional on the body of evidence E.

So, the votes are independent from each other not once the state of the world is given, but once the evidence is given. Technically, this is described by saying that the consequences are *screened off* by their common cause, which means that the consequences (here the votes) become independent when we conditionalize on their common cause. We have:

Proposition 1. *(PM) holds if and only if (S) and (I|E) hold.*
Proof. All proofs are given in the appendix. ■

The important part of proposition 1 is that (PM) entails (S) and (I|E), which provides a justification for using (S) and (I|E) in our jury model. We have also proved the reverse entailment to show that all theorems using (S) and (I|E) could equivalently use (PM).

In the new model, each juror's vote is a signal, not primarily about the state of the world, but about the body of evidence, which in turn is a signal about the state of the world.[8] Both signals are noisy: The body of evidence is a noisy signal about the state of the world; and a juror's vote is a noisy signal about the body of evidence. But both signals are typically biased towards the truth: The body of evidence is more likely to suggest

guilt than innocence in cases of guilt; and an individual juror's vote is more likely to reflect an 'ideal' interpretation of the evidence than not. We address these issues below.

In essence, our new jury theorem shows that majority voting reduces the noise in one set of signals – in the jurors' interpretation of the body of evidence – but not in the other – in the body of evidence as a signal about the state of the world.[9]

Let us introduce our assumption about juror competence formally. Recall that, in the classical model, competence was modelled by each juror's probability $p > 1/2$ of making a correct decision, conditional on the state of the world. We here define competence as the probability of giving an 'ideal' interpretation of the evidence, conditional on that evidence. Specifically, we assume that, for any body of evidence $e \in \mathbf{E}$, there exists an 'ideal' interpretation, denoted $f(e)$, that a hypothetical ideal observer of e would give. This ideal observer does not know the true state of the world, but gives the ideal (best possible) interpretation of the available evidence; $f(e) = 1$ means that the ideal observer would vote for 'guilty', and $f(e) = 0$ means that the ideal observer would vote for 'not guilty'. We call $f(e)$ the *ideal vote* – as opposed to the *correct vote*, which is the vote matching the true state of the world.[10] While knowledge of the true state of the world x would allow a correct vote, the ideal vote results from the best possible interpretation of the evidence e. The ideal vote and the correct vote differ in the case of misleading evidence, such as when an innocent person's hair is blown to the crime scene (and the person has no other alibi etc.). Our competence assumption states that the probability that juror i's vote matches the ideal vote $f(e)$ given the evidence e exceeds 1/2. Informally, each juror is better than random at arriving at an 'ideal' interpretation of the evidence.[11]

Competence Given the Evidence (C|E). For all jurors $i = 1, 2, ..., n$ and each body of evidence $e \in \mathbf{E}$, $p_e := P(V_i = f(e) | E = e) > 1/2$. The value of p_e may depend on e.

We assume – for simplicity and following the classical model – that the value of p_e is the same for all jurors i.[12] But we allow that p_e may depend on e. If the body of evidence e is easily interpretable, for instance in the case of overwhelming evidence for innocence, the probability that an individual juror's vote matches the ideal vote – here $f(e)=0$ – might be

high, say p_e=0.95. If the body of evidence e is confusing or ambiguous, that probability might be only p_e=0.55. Thus competence is a family of probabilities, containing one p_e for each $e \in \mathbf{E}$. The term 'competence' here corresponds to the ability to interpret the different possible bodies of evidence $e \in \mathbf{E}$ in a way that matches the ideal interpretation. For simplicity, one might replace (C|E) with the stronger (and less realistic) assumption of homogeneous competence, according to which p_e is the same for all possible $e \in \mathbf{E}$.

Homogeneous Competence Given the Evidence (HC|E). For all jurors i = 1, 2, ..., n, $p := P(V_i = f(e) | E = e) > 1/2$, for each body of evidence $e \in \mathbf{E}$. The value of p does not depend on e.

3. The probability distribution of the jury's vote

We consider the model based on diagram 3 – assuming (PM) and hence (S) and (I|E) – and derive the probability distribution of the jury's vote $V = \Sigma_{i=1,\ldots,n} V_i$ given the state of the world. This distribution depends crucially on two parameters: $p^{(1)} := P(f(E)=1|X=1)$ and $p^{(0)} := P(f(E)=0|X=0)$. The first is the probability that the evidence is not misleading (that it points to the truth for an ideal observer) in the case of guilt; the second is the probability that the evidence is not misleading in the case of innocence. Our first result addresses the case of homogeneous competence (HC|E).

Theorem 2. *If we have (S), (I|E) and (HC|E), the probability of obtaining precisely v out of n votes for 'guilty' given guilt is*

$$P(V=v|X=1) = p^{(1)}\binom{n}{v}p^{v}(1-p)^{n-v} + (1-p^{(1)})\binom{n}{v}p^{n-v}(1-p)^{v};$$

and the probability of obtaining precisely v out of n votes for 'guilty' given innocence is

$$P(V=v|X=0) = p^{(0)}\binom{n}{v}p^{n-v}(1-p)^{v} + (1-p^{(0)})\binom{n}{v}p^{v}(1-p)^{n-v}.$$

By theorem 2, if there is a non-zero probability of misleading evidence – specifically if $0<p^{(1)}<1$ or $0<p^{(0)}<1$ – the jury's vote V given the state of the world X does not have a binomial distribution, in contrast to the classical Condorcet jury model. The reason for this is that the votes V_1, V_2, ..., V_n, while *independent* given the evidence, are *dependent*

given the state of the world. The sum of dependent Bernoulli variables does not in general have a binomial distribution. If, on the other hand, the probability of misleading evidence is zero – i.e. $p^{(1)}=1$ and $p^{(0)}=1$ – the probabilities in theorem 2 reduce to those in the classical Condorcet jury model. Our next result describes the probability $P(V=v|X=x)$ for the more general case where we assume (C|E) rather than (HC|E).

Since E is a random variable, E induces a random variable p_E which takes as its value the competence p_e associated with the value e of E. To avoid confusion with the random variable E, we write the *expected value operator* as $Exp(.)$ rather than $E(.)$.

Theorem 3. *If we have (S), (I|E) and (C|E), the probability of obtaining precisely v out of n votes for 'guilty' given guilt is*

$$P(V=v|X=1) = p^{(1)}\binom{n}{v}Exp(p_E^{\,v}(1-p_E)^{n-v}\,|f(E)=1 \text{ and } X=1)$$
$$+ (1- p^{(1)})\binom{n}{v}Exp(p_E^{\,n-v}(1-p_E)^{v}|f(E)=0 \text{ and } X=1);$$

and the probability of obtaining precisely v out of n votes for 'guilty' given innocence is

$$P(V=v|X=0) = p^{(0)}\binom{n}{v}Exp(p_E^{\,n-v}(1-p_E)^{v}|f(E)=0 \text{ and } X=0)$$
$$+ (1- p^{(0)})\binom{n}{v}Exp(p_E^{\,v}(1-p_E)^{n-v}\,|f(E)=1 \text{ and } X=0).$$

Note that, in theorems 2 and 3, by summing $P(V=v|X=1)$ over all $v > n/2$, we obtain the probability of a simple majority for 'guilty' given guilt; and, by summing $P(V=v|X=0)$ over all $v < n/2$, we obtain the probability of a simple majority for 'not guilty' given innocence. The present results allow us to compare the probability of a correct jury verdict in our model – specifically in the case of homogeneous competence – with that in the classical Condorcet jury model for the same fixed level of juror competence p.

Corollary 1. *Suppose we have (S), (I|E) and (HC|E). Let v>n/2. Then the probability of obtaining precisely v out of n votes for 'guilty' given guilt satisfies*

$$P(V=v|X=1) \leq \binom{n}{v}p^{v}(1-p)^{n-v},$$

and so the probability of obtaining a majority for 'guilty' given guilt satisfies

$$P(V>n/2|X=1) \leq \sum_{v>n/2}\binom{n}{v}p^{v}(1-p)^{n-v}.$$

The left-hand sides of the two inequalities correspond to our new model, the right-hand sides to the classical model. So corollary 1 implies that the probability of a majority for 'guilty' given guilt in our new model is less than or equal to that in Condorcet's model. Similarly, the probability of a majority for 'not guilty' given innocence in our new model is less than or equal to that in the classical model. The probability of a correct jury verdict is equal in the two models if and only if the probability of misleading evidence is zero. Unless the evidence always 'tells the truth' – unless $p^{(1)}=p^{(0)}=1$ – the jury in our new model will reach a correct verdict with a lower probability than in the classical model.

4. A modified jury theorem

We now state our modified jury theorem. Its first part is concerned with the probability that the majority of jurors matches the ideal vote, and its second part with the more important probability that the majority of jurors matches the true state of the world.

Theorem 4. *Suppose we have (S), (I|E) and (C|E).*

(i) *Let W be the number of jurors i whose vote V_i coincides with the ideal vote $f(E)$. For each $x\in\{0,1\}$, $P(W > n/2|X=x)$ converges to 1 as n tends to infinity.*

(ii) *$P(V>n/2|X=1)$ converges to $p^{(1)}$ as n tends to infinity, and $P(V<n/2|X=0)$ converges to $p^{(0)}$ as n tends to infinity.*

Part (i) states that, given the state of the world, the probability that the majority verdict matches the ideal interpretation of the evidence converges to 1 as n tends to infinity. But the ideal interpretation may not be correct. Part (ii) states that the probability that the majority verdict matches the true state of the world (given that state) converges to the probability that the ideal interpretation of the evidence is correct, i.e. that the evidence is not misleading. Reformulating part (i), the probability of no simple majority for the *ideal* interpretation of the evidence converges to 0. Reformulating part (ii), the probability of no simple majority matching the true state of the world converges to the probability that the evidence is misleading, i.e. that the ideal interpretation of the evidence is incorrect.

This theorem allows the interpretation that, by increasing the jury size, it is possible to approximate the ideal interpretation of the evidence, no more and no less. The problem of

insufficient or misleading evidence cannot be avoided by adding jurors. Irrespective of the jury size, the probability of a correct majority decision at most approaches the probability that the evidence 'tells the truth', i.e. that its ideal interpretation matches the state of the world. Since there is typically a nonzero probability of misleading evidence – i.e. a nonzero probability that the evidence, even when ideally interpreted, points to 'guilt' when the defendant is innocent or vice-versa – the probability that the jury will fail to track the truth converges to a nonzero value as the jury size increases, regardless of how large the competence parameters p_e are in condition (C|E).[13]

5. Reasonable doubt

We now discuss the implications of our findings for the Bayesian question of when a jury is capable of establishing guilt of a defendant 'beyond any reasonable doubt'. So far we have been concerned with the 'classical' probability of a particular voting outcome – for instance, a majority for 'guilty' – conditional on the state of the world. But in a jury context, we may also be interested in the Bayesian probability of a particular state of the world – for instance, the guilt of the defendant – conditional on a particular voting outcome. Suppose we initially attach a certain prior probability to the hypothesis that the defendant is guilty. We may then ask: Given that the jury has produced a particular majority for guilty, what is the posterior probability that the defendant is truly guilty? Reformulated in degree of belief terms, the question is this: What degree of belief can we attach to the hypothesis that the defendant is truly guilty, given that we have observed a particular voting outcome in the jury, such as an overwhelming majority for 'guilty'?

Formally, the probability we are concerned with here is not $P(V=v|X=x)$, but $P(X=x|V=v)$. Note the reversed order of conditionalization. Let $r = P(X=1)$ denote the prior probability that the defendant is guilty. We assume that there is prior uncertainty about the guilt of the defendant, i.e. $0 < r < 1$. Below we also assume nonzero probabilities of misleading evidence, i.e. $0 < p^{(1)}, p^{(0)} < 1$. In the classical model – assuming (I|X) – we have:

$$P(X = 1|V = v) = \frac{rp^{2v-n}}{rp^{2v-n} + (1\text{-r})(1-p)^{2v-n}} \text{ (List 2004a).}$$

We can easily see that, for a sufficiently large jury and a sufficiently large majority, $P(X = 1|V = v)$ can take a value arbitrarily close to 1. In the limiting case where all jurors vote unanimously, the posterior belief converges to the alternative ('guilty' or 'innocent') supported by all jurors: $P(X = 1|V = n)$ converges to 1, and $P(X = 1|V = 0)$ converges to 0, as n tends to infinity. It is important to keep this implication of the classical model in mind when we see the results for our modified model. To simplify the exposition, we only consider the case of homogeneous competence here, i.e. (HC|E). The general case is technically more involved, but essentially analogous.

Theorem 5. *If we have (S), (I|E) and (HC|E), then the probability that the defendant is guilty given that precisely v out of n jurors have voted for 'guilty' is*

$$P(X = 1|V = v) = \frac{1}{1+\frac{1-r}{r}\times\frac{(1-p^{(0)})(p/(1-p))^{2v-n}+p^{(0)}}{p^{(1)}(p/(1-p))^{2v-n}+(1-p^{(1)})}}.$$

How confident in the correctness of a jury verdict can we ever be, given these Bayesian considerations? More formally, how close to 1 or 0 can the posterior probability $P(X = 1|V = v)$ ever get? Possibly never very close, unlike in the classical model. Consider the best-case scenario, where all jurors vote unanimously, either for 'guilty' or for 'innocent'. These two cases correspond to $V = n$ and $V = 0$. Using theorem 5 we can determine the posterior probability of guilt given $V=n$ and the posterior probability of guilt given $V = 0$.

Corollary 2. *Suppose we have (S), (I|E) and (HC|E). Then*:

(a) The probability that the defendant is guilty given a unanimous 'guilty' vote is

$$P(X = 1|V = n) = \frac{1}{1+\frac{1-r}{r}\times\frac{(1-p^{(0)})(p/(1-p))^{n}+p^{(0)}}{p^{(1)}(p/(1-p))^{n}+(1-p^{(1)})}},$$ *which converges to*

$$\frac{1}{1+((1-r)/r)((1-p^{(0)})/p^{(1)})} = P(X=1|f(E)=1)\ (<1) \quad \textit{as n tends to infinity.}$$

(b) The probability that the defendant is guilty given a unanimous 'not guilty' vote is

$$P(X = 1|V = 0) = \frac{1}{1+\frac{1-r}{r}\times\frac{(1-p^{(0)})((1-p)/p)^{n}+p^{(0)}}{p^{(1)}((1-p)/p)^{n}+(1-p^{(1)})}},$$ *which converges to*

$$\frac{1}{1+((1-r)/r)\,(p^{(0)}/(1-p^{(1)}))} = P(X=1|f(E)=0)\ (>0) \quad \textit{as n tends to infinity.}$$

By contrast, in the classical model $P(X=1|V=n)$ converges to 1 and $P(X=1|V=0)$ converges to 0, as n tends to infinity.

So, as n increases, [the probability that the defendant is guilty given a unanimous vote for 'guilty'] converges to [the probability that the defendant is guilty given that the evidence points towards guilt]. Likewise, as n increases, [the probability that the defendant is guilty given a unanimous vote for 'not guilty'] converges to [the probability that the defendant is guilty given that the evidence points towards innocence].

Corollary 2 describes the bounds on the posterior probability that $X = 1$ or $X = 0$, given the verdict of a large jury, by assuming the unrealistic case that V/n tends to 1 or 0. But this case occurs with probability 0 (unless $p = 1$), since with probability 1 the proportion of 'guilty'-votes V/n converges to either p or $1-p$. The former is the case if $f(E)=1$, the latter if $f(E)=0$. However, even in these two realistic cases – V/n converging to p and V/n converging to $1-p$ – the posterior probability of guilt, given the jury verdict, converges to exactly the same limits as in corollary 2.

Corollary 3. *Suppose we have (S), (I|E) and (HC|E). Let* $v_1, v_2, ..., v_n \in \{0,1\}$ *and put* $q_n := (v_1+v_2+...+v_n)/n$ *for all n. Then the probability that the defendant is guilty given that a proportion of* q_n *of the jurors have voted for 'guilty' – where* q_n *converges to either p or* $1-p$ *as n tends to infinity – is as follows*:

(a) *If* q_n *converges to p, then* $P(X = 1|V/n = q_n)$ *converges to* $P(X=1|f(E)=1)$ (<1), *as n tends to infinity (as in case (a) of corollary 2).*

(b) *If* q_n *converges to* $1-p$, *then* $P(X = 1|V/n = q_n)$ *converges to* $P(X=1|f(E)=0)$ (> 0), *as n tends to infinity (as in case (b) of corollary 2).*

The convergence results of corollaries 1 and 2 are identical, showing that in sufficiently large juries it is irrelevant whether the jury supports an alternative unanimously or by a proportion close to p (the exact meaning of 'close' depends on n and on the distance of p to 1/2).

Now we are in a position to state the key implication of these results: It may be impossible, even in an arbitrarily large jury and even when there is unanimity for 'guilty', to establish guilt of a defendant 'beyond any reasonable doubt'. More precisely, suppose that the jury's overall decision (or the judge's decision based on the jury vote) is required to satisfy the following decision principle:

> Convict the defendant *if and only if* the posterior probability of guilt, given the jury vote, exceeds c, where c is some fixed parameter close to 1 (e.g. c = 0.95).

The parameter c captures the threshold of reasonable doubt: Only a posterior probability of guilt above c is interpreted as representing a degree of belief beyond reasonable doubt. By corollary 2, we can immediately see that, if $P(X=1 \mid f(E)=1) \leq c$, then conviction will *never* be possible according to the decision principle just introduced. No matter how large the jury is and no matter how large the majority for 'guilty' is, the jury vote will never justify a degree of belief greater than c that the defendant is guilty, and hence will never establish guilt of the defendant beyond any reasonable doubt. So, if $P(X=1 \mid f(E)=1) \leq c$, even a unanimous vote for 'guilty' in a ten-million-member jury will be insufficient for conviction – in sharp contrast to what Condorcet's classical model implies.

6. Summary

Using Bayesian networks, we have developed a new model of jury decisions. The model can represent a jury, committee or expert panel deciding on whether or not some factual proposition is true, and where the decision is made on the basis of shared evidence. We have suggested that our model is more realistic than the classical Condorcet jury model. First, it captures the empirical fact that in real-world jury, committee or expert panel decisions the state of the world is typically not the latest common cause of the jurors' votes, but there exists some intermediate common cause: the body of evidence, as described here. Secondly, in legal contexts, the model captures the requirement that jurors must not use any evidence other than that presented in the courtroom. This means that, *even if*, hypothetically, the jurors could each obtain an independent signal about the state of the world (without any intermediate common cause between different such

signals), they would be required by law not to use such information. Our model makes two key assumptions:

- The Parental Markov Condition, applied to the Bayesian tree in diagram 3, which has two implications:
    - Common Signal: The jurors' votes depend on the true state of the world only through the available body of evidence.
    - Independence Given the Evidence: The votes of different jurors are independent from each other given the available body of evidence.
- Competence Given the Evidence: For each body of evidence, each juror has a probability greater than 1/2 of matching the ideal interpretation of that evidence. In the 'homogeneous' case, juror competence is the same for all possible bodies of evidence; in the 'heterogeneous' case, it may depend on the evidence.

Then:

- The probability of correct majority decision (given the state of the world) is typically less than, and at most equal to, the corresponding probability in the classical Condorcet jury model.
- As the jury size increases, the probability of a correct majority decision (given the state of the world) converges to the probability that the evidence is not misleading. Unless the evidence is never misleading, the probability of a correct majority decision is strictly less than one.
- Depending on the required threshold of 'no reasonable doubt', it may be impossible, even in an arbitrarily large jury and even when the jury unanimously votes for 'guilty', to establish guilt of a defendant 'beyond any reasonable doubt'.

Our model reduces to the classical Condorcet jury model *if and only if* we assume *both* that the evidence is never misleading *and* that juror competence is the same for all possible bodies of evidence (homogeneous competence). *If* these assumptions are inadequate in real-world jury, committee or expert panel decisions, then the classical Condorcet jury model, as it stands, fails to apply to such decisions.

References

Austen-Smith, D., and J. Banks (1996), Information Aggregation, Rationality, and the Condorcet Jury Theorem, *American Political Science Review* 90: 34-45.

Boland, P. J. (1989), Majority Systems and the Condorcet Jury Theorem, *Statistician* 38: 181-189.

Boland, P. J., F. Proschan and Y. L. Tong (1989), Modelling dependence in simple and indirect majority systems, *Journal of Applied Probability* 26: 81-88.

Bovens, L. and E. Olsson (2000), Coherentism, reliability and Bayesian networks, *Mind* 109: 685-719.

Corfield, D. and J. Williamson (eds.) (2001), *Foundations of Bayesianism*, Dordrecht (Kluwer).

Dietrich, F. (2003), General Representation of Epistemically Optimal Procedures, *Social Choice and Welfare*, forthcoming.

Estlund, D. (1994), Opinion leaders, independence and Condorcet's jury theorem, *Theory and Decision* 36: 131-162.

Fitelson, B. (2001), A Bayesian Account of Independent Evidence with Application, *Philosophy of Science* 68 (Proceedings): S123-S140.

Grofman, B., G. Owen and S. L. Feld (1983), Thirteen theorems in search of the truth, *Theory and Decision* 15: 261-278.

Lahda, K. K. (1992), The Condorcet Jury Theorem, Free Speech, and Correlated Votes, *American Journal of Political Science* 36: 617-634.

List, C., and R. E. Goodin (2001), Epistemic Democracy: Generalizing the Condorcet Jury Theorem, *Journal of Political Philosophy* 9: 277-306.

List, C. (2004a), On the Significance of the Absolute Margin, *British Journal for the Philosophy of Science,* forthcoming.

List, C. (2004b), The Epistemology of Special Majority Voting, *Social Choice and Welfare*, forthcoming.

Nitzan, S., and J. Paroush (1984), The significance of independent decisions in uncertain dichotomous choice situations, *Theory and Decision* 17: 47-60.

Owen, G. (1986), Fair Indirect Majority Rules, in B. Grofman and G. Owen (eds.), *Information Pooling and Group Decision Making*, Greenwich, CT (Jai Press).

Pearl, J. (2000), Causality: models, reasoning, and inference, Cambridge (C.U.P.).

Appendix: Proofs

Proof of proposition 1.

(i) First assume (PM). Let $e \in \mathbf{E}$ be any body of evidence. We show that given $E=e$ the variables $V_1,\dots, V_n, X$ (votes and state of the world) are independent, which implies in particular that given $E=e$ the votes $V_1,\dots, V_n$ are independent (Independence Given the Evidence (I|E)) and that given $E=e$ the vote vector $(V_1,\dots, V_n)$ is independent of X (which is equivalent to Common Signal (S)).

To show that given $E=e$ the variables $V_1,\dots, V_n, X$ are independent, let $v_1,\dots, v_n, x \in \{0,1\}$ be any possible realizations of these variables. First, we apply (PM) on the first juror's vote V_1: Since E is the only parent of V_1 and all of $V_2,\dots, V_n, X$ are non-descendants of V_1, by (PM), given $E=e$, V_1 is independent of the vector of variables $(V_2,\dots, V_n, X)$, i.e.

(1) $P(V_1= v_1, \dots, V_n= v_n, X=x|E=e)=P(V_1= v_1|E=e)P(V_2= v_2, \dots, V_n= v_n, X=x|E=e)$.

Next, we apply (PM) on V_2 to decompose the second term of the last product: Since E is the only parent of V_2 and all of $V_3,\dots, V_n, X$ are non-descendants of V_2, by (PM), given $E=e$, V_2 is independent of the vector of variables $(V_3,\dots, V_n, X)$, i.e.

$$P(V_2= v_2, \dots, V_n= v_n, X=x|E=e)= P(V_2= v_1|E=e)P(V_3= v_3, \dots, V_n= v_n, X=x|E=e).$$

Substituting this into (1), we obtain

$$P(V_1= v_1, \dots, V_n= v_n, X=x|E=e)= P(V_1= v_1|E=e)P(V_2= v_2|E=e) \times P(V_3= v_3, \dots, V_n= v_n, X=x|E=e).$$

By continuing to decompose joint probabilities, one finally arrives at

$$P(V_1= v_1, \dots, V_n= v_n, X=x|E=e)=P(V_1= v_1|E=e)\times \dots \times P(V_n= v_n|E=e)P(X=x|E=e),$$

which establishes the independence of $V_1,\dots, V_n, X$.

(ii) Now assume (S) and (I|E). Let $e \in \mathbf{E}$ be any realization of E. To show (PM) we have to go through all nodes of the tree. What (PM) states for the top node X is vacuously true since X has no non-descendants (except itself). Regarding E, its only non-descendant (except itself) is its parent X, and of course, given X, E is independent of X since X is deterministic. Finally, consider vote V_1 (the proof for any other vote $V_2, \dots, V_n$ is analogous). We have to show that V_1 is independent of its vector of non-descendants $(V_2, \dots, V_n, X)$ given its parent $E=e$. (We have excluded E from the vector of non-

descendants because E is deterministic given $E=e$.) Let $v_2,\dots, v_n, x\in\{0,1\}$ be any realizations of $V_2, \dots, V_n, X$. By (S), given $E=e$, $(V_2, \dots, V_n)$ is independent of X, and so

$$P(V_2= v_2, \dots, V_n= v_n, X=x|E=e)= P(V_2= v_2, \dots, V_n= v_n|E=e)P(X=x|E=e).$$

Now we can apply (I|E) to decompose the first factor in the last product, which yields

$$P(V_2= v_2, \dots, V_n= v_n, X=x|E=e)= P(V_2= v_2 |E=e)\times\dots\times P(V_n= v_n|E=e)P(X=x|E=e).$$

This shows the independence of $(V_2,\dots, V_n, X)$ given $E=e$. ■

An alternative proof of proposition 1 might be given using the criterion of *d*-separation or the theory of *semi-graphoids*.

Proof of theorem 2.

By (HC|E), each body of evidence $e\in\mathbf{E}$ is equally easy to interpret ideally, and so we assume for simplicity that $\mathbf{E} = \{0, 1\}$, where $e = 0$ is the evidence ideally interpreted as suggesting innocence $f(e)=0$, and $e = 1$ is the evidence ideally interpreted as suggesting guilt $f(e)=1$. By (HC|E) and (I|E), if $E=1$ then the votes $V_1, V_2, \dots, V_n$ are independently Bernoulli distributed, with a probability p of $V_i = 1$ and a probability $1-p$ of $V_i = 0$ for each i. If $E=0$ then the votes $V_1, V_2, \dots, V_n$ are also independently Bernoulli distributed, with a probability p of $V_i = 0$ and a probability $1-p$ of $V_i = 1$ for each i. Hence, given $E=1$, the jury's vote $V = \Sigma_{i=1,\dots,n}V_i$ has a Binomial distribution with parameters n and p. And given $E=0$, V has a Binomial distribution with parameters n and $1-p$:

$$(2)\qquad P(V=v|E=1) = \binom{n}{v}p^v(1-p)^{n-v},\ \ P(V=v|E=0)=\binom{n}{v}p^{n-v}(1-p)^v.$$

Now, the probability of obtaining precisely v out of n votes for 'guilty' given the state of the world x is:

$$P(V=v|X=x)=P(V=v|E=1 \textit{ and } X=x)P(E=1|X=x)+P(V=v|E=0 \textit{ and } X=x)P(E=0|X=x).$$

By (S), conditionalizing on *both* $E=e$ *and* $X=x$ is equivalent to conditionalizing only on $E=e$, so that:

$$P(V=v|X=x) = P(V=v|E=1)P(E=1|X=x) + P(V=v|E=0)P(E=0|X=x).$$

Explicitly, taking the two cases $x=0$ and $x=1$,

$$P(V=v|X=1) = P(V=v|E=1)p^{(1)} + P(V=v|E=0)(1- p^{(1)});$$

$$P(V=v|X=0) = P(V=v|E=0)p^{(0)} + P(V=v|E=1)(1-p^{(0)}).$$

Recall that $p^{(1)} := P(f(E)=1|X=1)$ and $p^{(0)} := P(f(E)=0|X=0)$, and here $f(E)=E$. Now theorem 2 in the case $\mathbf{E} = \{0, 1\}$ follows from (2). The general case follows from theorem 3 below. ■

Proof of theorem 3.

First, we use the law of iterated expectations to write

$$P(V=v|X=x) = Exp(P(V=v|E \text{ and } X=x)|X=x).$$

By (S) we have $P(V=v|E \text{ and } X=x) = P(V=v|E)$, so that we deduce

(3) $$P(V=v|X=x) = Exp(P(V=v|E)|X=x).$$

By (C|E) and (I|E), conditional on E the votes $V_1, V_2, ..., V_n$ are independent and Bernoulli distributed with parameter p_E if $f(E)=1$ and $1-p_E$ if $f(E)=0$. Hence the sum V has a binomial distribution with first parameter n and second parameter p_E if $f(E)=1$ and $1-p_E$ if $f(E)=0$:

$$P(V=v|E) = \begin{cases} \binom{n}{v} p_E^{v}(1-p_E)^{n-v} & \text{if } f(E)=1 \\ \binom{n}{v} p_E^{n-v}(1-p_E)^{v} & \text{if } f(E)=0. \end{cases}$$

In other words,

$$P(V=v|E) = \binom{n}{v} p_E^{v}(1-p_E)^{n-v} 1_{\{f(E)=1\}} + \binom{n}{v} p_E^{n-v}(1-p_E)^{v} 1_{\{f(E)=0\}},$$

where $1_{\{f(E)=1\}}$ and $1_{\{f(E)=0\}}$ are characteristic functions (1_A is the random variable defined as 1 if the event A holds and as 0 if it does not).

So, by (3) and the linearity of the (conditional) expectation operator $Exp(.|X=x)$,

$$P(V=v|X=x) = P(f(E)=1|X=x)\binom{n}{v} Exp(p_E^{v}(1-p_E)^{n-v}|f(E)=1 \text{ and } X=x)$$
$$+ P(f(E)=0|X=x)\binom{n}{v} Exp(p_E^{n-v}(1-p_E)^{v}|f(E)=0 \text{ and } X=x). \blacksquare$$

Proof of corollary 1.

Suppose (HC|E) holds. Assume that $v>n/2$ (a majority for 'guilty'). Then

$$p^{n-v}(1-p)^{v} = p^{v}(1-p)^{n-v}((1-p)/p)^{2v-n} \leq p^{v}(1-p)^{n-v},$$

since $2v-n>0$ and $p>1/2$. So, by the formula for $P(V=v|X=1)$ in theorem 2, we deduce.

$$P(V=v|X=1) \leq p^{(1)}\binom{n}{v}p^{v}(1-p)^{n-v} + (1-p^{(1)})\binom{n}{v}p^{v}(1-p)^{n-v} = \binom{n}{v}p^{v}(1-p)^{n-v}, \text{ as required. } \blacksquare$$

Proof of theorem 4.

(i) We conditionalize on E. By (C|E) and (I|E), W is the sum of n independent Bernoulli variables with parameter p_E. The weak law of large numbers implies that the average W/n converges in probability to p_E. Since $p_E>1/2$, it follows that

$$lim_{n\to\infty}P(W>n/2|E)=1.$$

Applying the (conditional) expectation operator on both sides (which corresponds to averaging with respect to E), we obtain

$$Exp(lim_{n\to\infty}P(W>n/2|E) \mid X=x))=Exp(1 \mid X=x)=1.$$

By the dominated convergence theorem, we can interchange the expectation operator with the limit operator on the left hand side, so that

$$lim_{n\to\infty}\ Exp(P(W>n/2|E) \mid X=x) = 1.$$

By (S) we can replace $P(W>n/2|E)$ by $P(W>n/2|E \textit{ and } X=x)$. This leads to

$$lim_{n\to\infty}\ Exp(P(W>n/2|E \textit{ and } X=x) \mid X=x) = 1,$$

and hence by the law of iterated expectations

$$lim_{n\to\infty}\ P(W>n/2 \mid X=x) = 1.$$

(ii) Using the weak law of large numbers in a similar way as in (i), it is possible to prove that the probability $P(V>n/2|E)=P(V/n>1/2|E)$ converges to 1 if f(E)=1 and to 0 if $f(E)=0$ (as n tends to infinity). Hence

(4) $$lim_{n\to\infty}\ P(V>n/2 \mid E)=1_{\{f(E)=1\}},$$

where $1_{\{f(E)=1\}}$ is the random variable defined as 1 if $f(E)=1$ and as 0 if $f(E)=0$.

By the law of iterated expectations,

$$P(V>n/2|X=1)=Exp(P(V>n/2|E \textit{ and } X=1)|X=1), \text{ which by (S) simplifies to:}$$

(5) $$P(V>n/2|X=1)= Exp(P(V>n/2|E)|X=1).$$

Further, we have

$$P(f(E)=1|X=1)=Exp(1_{\{f(E)=1\}}|X=1)=Exp(lim_{n\to\infty}\ P(V>n/2 \mid E)|X=1),$$

where the last step uses (4). We now interchange the expectation operator with the limit (by the dominated convergence theorem) and then use (5) to obtain

$$P(f(E)=1|X=1)=lim_{n\to\infty}\ Exp(P(V>n/2 \mid E)|X=1)= lim_{n\to\infty}\ P(V>n/2|X=1).$$

As for the case $X=0$, it can be shown similarly that

$$P(f(E)=0|X=0)=lim_{n\to\infty}\ P(V<n/2|X=0). \blacksquare$$

The complexity of this proof is due to the fact that the set of possible evidences **E** is arbitrarily large (and endowed with some σ-algebra). For finite or countable **E**, (conditional) expectation operators could be replaced by summations.

Proof of theorem 5.

By Bayes's theorem, for any v,

$$P(X=1|V=v)=\frac{rP(V=v|X=1)}{rP(V=v|X=1)+(1-r)P(V=v|X=0)}.$$

Dividing numerator and denominator by $rP(V=v|X=1)$, we get

$$P(X=1|V=v)=\frac{1}{1+{}^{(1-r)}/_{r}\,{}^{P(V=v|X=0)}/_{P(V=v|X=1)}}.$$

We use theorem 2 for expressing $P(V=v|X=1)$ and $P(V=v|X=0)$, and we then simplify:

$$\frac{P(V=v|X=0)}{P(V=v|X=1)}=\frac{(1-p^{(0)})\binom{n}{v}p^{v}(1-p)^{n-v}+p^{(0)}\binom{n}{v}p^{n-v}(1-p)^{v}}{p^{(1)}\binom{n}{v}p^{v}(1-p)^{n-v}+(1-p^{(1)})\binom{n}{v}p^{n-v}(1-p)^{v}}$$

$$=\frac{(1-p^{(0)})(p/(1-p))^{2v-n}+p^{(0)}}{p^{(1)}(p/(1-p))^{2v-n}+(1-p^{(1)})}. \blacksquare$$

Proof of corollary 2.

To prove part (a), note that the convergence to

$$\frac{1}{1+{}^{(1-r)}/_{r}(1-p^{(0)})/p^{(1)}}$$

is clear because $(p/(1-p))^{n}\to\infty$, so that the ratio

$$\frac{(1-p^{(0)})(p/(1-p))^{n}+p^{(0)}}{p^{(1)}(p/(1-p))^{n}+(1-p^{(1)})}$$

is asympotically equivalent to

$$\frac{(1-p^{(0)})(p/(1-p))^{n}+0}{p^{(1)}(p/(1-p))^{n}+0}=\frac{1-p^{(0)}}{p^{(1)}}.$$

The rest follows from

$$\frac{1}{1+{}^{(1-r)}/_{r}(1-p^{(0)})/p^{(1)}} = \frac{1}{1+[P(X=0)/P(X=1)] \times P(f(E)=1|X=0)/P(f(E)=1|X=1)}$$

$$= \frac{P(X=1)\ P(f(E)=1|X=1)}{P(X=1)\ P(f(E)=1|X=1) + P(X=0)\ P(f(E)=1|X=0)}.$$

Part (b) has an analogous proof. ■

Proof of corollary 3.

In the formula of theorem 5, we replace v by nq_n. If $q_n \to p(>1/2)$, then the term $[p/(1-p)]^{2nq_n-n} = [p/(1-p)]^{2n(q_n-1/2)}$ tends to ∞. So the ratio

$\frac{(1-p^{(0)})(p/(1-p))^{2nq_n-n}+p^{(0)}}{p^{(1)}(p/(1-p))^{2nq_n-n}+(1-p^{(1)})}$ is asymptotically equivalent to $\frac{(1-p^{(0)})(p/(1-p))^{2nq_n-n}+0}{p^{(1)}(p/(1-p))^{2nq_n-n\ n}+0} = \frac{1-p^{(0)}}{p^{(1)}}$,

which proves (a). The proof of (b) is analogous. ■

---

[1] All conditions are formulated for a given group size n rather than beginning with 'For all n'. However, in many of our results, the group size is not fixed and tends to infinity. In these results, we implicitly assume that all conditions begin with 'For all n' (and that the competence parameter in the competence conditions is the same for all n). Compare List (2004b).

[2] We hereafter mean 'probabilistically caused' when we use the expression 'caused'. Probabilistic causation means that the cause affects the *probabilities* of consequences, whereas deterministic causation means that the cause determines the consequence *with certainty*. Probabilistic causation can arise for at least two reasons. Metaphysical reasons: The process in question may be genuinely indeterministic; causes determine consequences only with probabilities strictly between 0 and 1, but not with certainty. Epistemic reasons: The process in question may or may not be deterministic at the most fundamental level, but due to its complexity we may not be able to include, or fully describe, all relevant causal factors in the network representation; thus probabilities come into play. We here remain neutral on which of these two reasons apply, though it is obvious that any theoretical representation of jury decisions will be underdescribed and thus epistemically limited. (Our definition of probabilistic causation allows the special case where the net causal effect on probabilities is zero, because positive and negative causes may cancel each other out.)

[3] Several generalizations of the classical Condorcet jury model have been discussed in the literature. We have already referred to existing discussions of dependencies between different jurors' votes. Cases where different jurors have different competence levels are discussed, for instance, in Grofman, Owen and Feld (1983), Boland (1989) and Dietrich (2003). Cases where jurors vote strategically rather than sincerely are discussed, for instance, in Austen-Smith and Banks (1996). Cases where choices are not binary are discussed, for instance, in List and Goodin (2001). Cases where juror competence depends on the jury size are discussed, for instance, in List (2004b).

[4] Sometimes Bayesian networks are assumed to contain more information: Each node in the graph is endowed with a probability distribution of the variable at this node conditional on the node's parents (or unconditionally if there are no parents).

[5] To specify a joint probability distribution of the variables satisfying the Parental Markov Condition, it is sufficient to specify a distribution of each variable conditional on its parents (an unconditional distribution if there are no parents). The product of all these conditional probability functions then yields a joint distribution of all variables that satisfies the Parental Markov Condition.

[6] So all jurors base their votes solely on the same value e of E. Differences between jurors' votes are not the result of the jurors' independent – and thus potentially different – access to the state of the world (as in the classical model), but the result of different interpretations of the same evidence e. One juror might interpret the defendant's smile as a sign of innocence, whereas another might give the opposite interpretation.

[7] An equivalent statement of (S) is the following: Given E, the vector of votes $(V_1, V_2, \ldots, V_n)$ is independent of the state of the world X.

[8] One can imagine cases where (part of) the evidence E is *not* caused by the state of the world X. For instance, if X is the fact of whether or not the defendant has committed a given crime, then the information that the defendant bought a gun in a nearby shop two days before the crime may be evidence for guilt. But this evidence cannot be caused by the crime since the gun purchase happened *before* the crime. Rather, the causal link between the gun purchase and the crime goes in the other direction. To capture such cases, one might want to replace our causal relation $X \rightarrow E$ by some other causal relation between X and E, e.g. by $X \leftarrow E$, or by a bidirectional causal relation $X \leftrightarrow E$, or by a common parent of X and E. The theorems and corollaries of this paper still apply to such modified Bayesian trees (provided that the state X remains related to the votes *only through* the evidence E). The reason is that the Parental Markov Condition (PM) still implies Common Signal (S) and Independence Given the Evidence (I|E), so that (S) and (I|E) remain justified assumptions.

[9] This model captures not only the empirical fact that in real world jury decisions the available evidence is usually finite and limited, but also the legal norm, mentioned in the introduction, that jurors are not allowed to obtain or use any evidence other than that presented in the courtroom, or to discuss the case with any persons other than the other jurors.

[10] Different interpretations of the ideal vote $f(e)$ may be given. One is that the ideal vote is 1 if and only if the *objective* probability of guilt given the evidence e exceeds some threshold. Here the ideal interpreter is assumed to know the objective likelihoods (of the evidence given guilt and given innocence) and the objective prior probability of guilt. Another interpretation, which does not require an objective prior of guilt but a shared prior of guilt, is to assume that the ideal interpreter uses the group's shared (perhaps not objective) prior probability of guilt to calculate the posterior probability of guilt given the evidence. We can give a Bayesian account of both interpretations. Assume that the set $\mathbf{E}$ of all possible bodies of evidence is countable. Suppose that, by knowing the evidence-generating stochastic process, the ideal observer knows the probabilities $P(E=e|X=1)$ and $P(E=e|X=0)$. Suppose, further, that the ideal observer assigns the (objective or shared) prior probability $r:=P(X=1)$ to the proposition that the defendant is guilty. Then, using Bayes's theorem, the ideal observer can calculate the posterior probability that the defendant is guilty, given the evidence e, i.e. $P(X=1|E=e) = rP(E=e|X=1) / (rP(E=e|X=1) + (1-r)P(E=e|X=0))$. Furthermore, the group (or the ideal observer) might set a (normative) threshold of when to accept, beyond any reasonable doubt, that the defendant is guilty, given the evidence e. Now the ideal vote is a 'guilty' vote if $P(X=1|E=e) > 1-\varepsilon$ (for a suitable $\varepsilon > 0$) and a 'not guilty' vote otherwise. The prior probability r represents the degree of belief the ideal observer assigns to the guilt of the defendant *before* having seen any evidence. The value of ε represents how demanding the criterion of 'beyond any reasonable doubt' is.

[11] We also allow that not all jurors have observed the entire evidence e. For instance, some jurors might have missed the smile of the defendant. What matters is not that all jurors base their vote on the *full* evidence e, but that they use information *contained* in e. A juror's information is thus limited by e, which represents the maximally available information for *any* jury size.

[12] This assumption is a technical simplification, but involves no real loss of generality. As in the classical Condorcet jury model (e.g. Boland 1989), our model can be generalized by allowing differently competent jurors, so that the competence $P(V_i=f(e)|E=e)$ depends also on i, denoted $p_{e,i}$. Our asymptotic results then remain true if we replace (C|E) (respectively (HC|E)) by the weaker competence assumption that the limiting average competence, $lim_{n\rightarrow\infty}\sum_{all\ i} p_{e,i}/n$, exceeds 1/2. In corollary 3 one has to interpret p as the limiting average competence across jurors; since corollary 3 requires (HC|E), this limiting average competence does not depend on e here.

[13] It is possible to prove a slightly stronger result than theorem 4. Given the state of the world x, the ratio V/n converges with probability 1 to the random variable defined by p_E (>1/2) if $f(E)=1$ and $1-p_E$ (<1/2) if $f(E)=0$ (<1/2). Among these two possible limits the one that corresponds to a majority for the correct alternative happens with probability $p^{(x)} = P(f(E)=x|X=x)$. Hence, with probability 1, there is convergence to a stable majority as the jury size increases, where this majority supports the correct alternative with the probability that the evidence 'tells the truth'.