University of Groningen

Learning juror competence

Romeijn, J.W.; Atkinson, D.

Published in:
Philosophy, Politics and Economics

DOI:
[10.1177/1470594X10372317](https://doi.org/10.1177/1470594X10372317)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2011

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Romeijn, J. W., & Atkinson, D. (2011). Learning juror competence: a generalized Condorcet Jury Theorem. *Philosophy, Politics and Economics*, 10(3), 237-262. <https://doi.org/10.1177/1470594X10372317>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Learning juror competence: a generalized Condorcet Jury Theorem

Politics, Philosophy & Economics

10(3) 237–262

© The Author(s) 2011

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1470594X10372317

ppe.sagepub.com

Jan-Willem Romeijn

University of Groningen, The Netherlands

David Atkinson

University of Groningen, The Netherlands

Abstract

This article presents a generalization of the Condorcet Jury Theorem. All results to date assume a fixed value for the competence of jurors, or alternatively, a fixed probability distribution over the possible competences of jurors. In this article, we develop the idea that we can learn the competence of the jurors by the jury vote. We assume a uniform prior probability assignment over the competence parameter, and we adapt this assignment in the light of the jury vote. We then compute the posterior probability, conditional on the jury vote, of the hypothesis voted over. We thereby retain the central results of Condorcet, but we also show that the posterior probability depends on the size of the jury as well as on the absolute margin of the majority.

Keywords

Condorcet Jury Theorem, juror competence, Bayesian inference, Condorcet formula, majority voting

1. Introduction

In its classical formulation, the Condorcet Jury Theorem concerns collective judgement by majority voting. For example, a jury might have to decide whether it thinks that Jack is guilty or innocent of killing Jill. Every member of the voting body, in this case the jury, votes between two alternatives, of which only one is correct. The collective chooses the

Corresponding author:

Jan-Willem Romeijn, Oude Boteringestraat 52, 9712 GL, Groningen, The Netherlands

Email: j.w.romeijn@rug.nl

alternative that receives most votes. In the example, each juror votes for guilt or innocence, and the verdict that receives most votes is chosen. It is assumed that consulting a single voter gives a better chance of a correct result than flipping a fair coin: the chance that a juror correctly chooses between guilt and innocence is larger than a half. The Condorcet Jury Theorem then states that if we include ever more voters in the collective, then the chance that the collective chooses the correct alternative tends to one. Put simply, the Condorcet Jury Theorem reflects the wisdom of the crowd.

The classical formulation of the theorem concerns the chances of a correct decision only. In the example, it concerns the chance of a majority vote for guilt if in actual fact Jack is guilty. As such, it does not say anything about the probability of the alternatives being true or false after all the votes have been tallied. However, under fairly general conditions we can deduce that with increasing size of the voting body, the probability associated with the true alternative tends to one as well. We can even derive a simple formula for the probability of the two alternatives, the so-called Condorcet formula. If we assume that the alternatives are of equal probability before voting has taken place and if we further assume that the chance that a voter chooses correctly is independent of which alternative happens to be true (often referred to as the symmetry of the juror competence), then the probability of the alternatives is a function only of the absolute margin between the voters for each of the alternatives: the larger the numerical difference between those who voted for Jack's guilt against those who voted for his innocence, the stronger the support for Jack's guilt. The size of the jury does not come into this equation.

Despite possible appearances to the contrary, the fact that the support according to the Condorcet formula is independent of the jury size is in keeping with the Condorcet Jury Theorem itself. On the assumption of either alternative being true, and using the fact that the chance for voters of choosing that alternative is larger than a half, we can show that the expected absolute margin will increase at the same rate as the size of the voting body. The Condorcet formula thus yields the correct results if we also consider ever larger voting bodies, even though the situations covered by the Condorcet formula are increasingly rare. However, the absence of jury size in the Condorcet formula is not entirely uncontroversial. Say that we compare two juries, a very large one of 1000 voters which has a comparatively narrow margin of 10 in favour of guilt and a small jury of 10 which is unanimous in its verdict of guilt. According to the Condorcet formula, both juries convey the same probability onto the alternatives of guilt and innocence. This is often considered to be counter-intuitive.

The present article presents a generalization of the Condorcet Jury Theorem, in which jury size plays a more prominent role. The key idea is that the relative majority of the jury (that is, the proportion of jurors that makes up the majority) reveals something about the competence of the jurors. Effectively, the symmetry of competence is thereby dropped, and replaced by an assumption on how we may learn about competence. Most consequences of the Condorcet Jury Theorem will hold true under the generalized theorem as well. The main departures are the following.

1. The probability that the jury majority verdict is incorrect monotonically increases with the jury size, n , if the absolute margin, Δ , remains constant.

2. The probability that the jury majority verdict is incorrect tends to one-half as n tends to infinity, if Δ remains constant in this limit.
3. The probability that the jury majority verdict is incorrect tends to zero as n tends to infinity, if the fractional majority, $f = \Delta/n$, tends to a non-zero positive constant in this limit.

The exclusive dependence on the absolute margin, as revealed by the Condorcet formula, is thus seen to be an artefact of idealizing assumptions, and not something inherent to jury verdicts.

2. Condorcet revisited

In this section, we introduce Condorcet's Jury Theorem (see Dietrich, 2008; Ladha, 1992; List, 2008; List and Goodin, 2001). We specifically present the results of List (2004), according to which the posterior probability of the hypothesis voted over, conditional on the jury vote, only depends on the absolute margin of votes in favour and against the hypothesis.

2.1. Condorcet's Theorem

Let H^1 be the hypothesis that Jack murdered Jill and H^0 the hypothesis that he did not, so $\{H^0, H^1\}$ is a partition. Suppose that of a jury of n members trying Jack a number n_1 vote that H^1 is true, while the remaining $n_0 = n - n_1$ members vote that H^0 is true. For both $j = 0$ and $j = 1$, we assume that if H^j is in fact true, the eventuality that jury member i votes for H^j being true, denoted by V_{ij} , has some fixed chance q_j , which we will call the competence of the jurors on the hypothesis H^j . For all $i, i' = 1, \dots, n$ and $j, j' = 0, 1$, if $i' \neq i$, we set

$$p(V_i^j | H^j \cap V_{i'}^{j'}) = p(V_i^j | H^j) = q_j. \quad (1)$$

The left equality says that the jurors all vote independently and the right one that they vote with fixed competences q_j , for both $j = 0, 1$. Note that the competences of the jurors with respect to H^0 and H^1 do not refer to a general ability to judge. They are specific for the hypotheses, and the competences for the hypotheses H^0 and H^1 can differ: jurors might be more accurate if H^0 is true than if H^1 is so. Finally, we assume that the competences will be greater than one half, $q_j > 1/2$, so the judgement of jury members is better than the result of tossing a fair coin.

We can now introduce Condorcet's Jury Theorem. Say that Jack is indeed guilty, H^1 , so that the probability of any jury member voting in favour of Jack's guilt, V_i^1 , namely q_1 , is greater than one-half. Now for an ever larger jury size n , consider the relative frequency of voters in favour, that is, $f_i = \frac{n_1}{n} = 1 - f_0$. By the law of large numbers, the difference $f_1 - q_1$ tends to zero. Because $q_1 > 1/2$, we have that the probability of a correct majority vote, $n_1 > n_0$, tends to 1 in the limit. We refer to Ladha (1993), Berend and Paroush (1998), and Dietrich (2008) for proofs of more general versions of this theorem.

All these versions concern cases in which the competences are sampled from a fixed distribution, thus modelling the heterogeneity of the jury.

2.2. Inverse probability

List (2004) emphasizes the importance of an inverted version of this result. Rather than calculating the probability of a majority of votes V_i^j given the truth of H^j , we want to know the probability of the correctness of the hypothesis H^j , given a majority of votes V_i^j . For convenience we denote the difference between votes for H^1 and H^0 , also called the ‘absolute margin of the majority’, by $\Delta = n_1 - n_0$. We denote the entire jury vote by $V_{n\Delta} = \cap_{i=1}^n V_i^{u(i)}$. Here $u(i) = 1$ if jury member i voted for H^1 and $u(i) = 0$ if she voted for H^0 , so that $n_1 = \sum_i u(i)$. By Bayes’ Theorem we have

$$\begin{aligned} p(H^1|V_{n\Delta}) &= \frac{p(\cap_i V_i^{u(i)}|H^1)p(H^1)}{p(V_{n\Delta})} \\ &= \frac{\prod_i p(\cap_i V_i^{u(i)}|H^1)p(H^1)}{p(V_{n\Delta})} \\ &= \frac{q_1^{n_1}(1-q_1)^{n_0}p(H^1)}{p(V_{n\Delta})}, \end{aligned}$$

where $p(H^1)$ is the prior probability of the hypothesis H^1 . For H^0 we can derive a similar expression, replacing q_1 by q_0 and swapping the roles of n_0 and n_1 . The denominator $p(V_{n\Delta}) = p(\cap_i V_i^{u(i)})$ is the same in both equations.

We can avoid calculating the denominator $p(V_{n\Delta})$ by using the posterior odds instead of the posterior probability:

$$\frac{p(H^1|V_{n\Delta})}{p(H^0|V_{n\Delta})} = \frac{q_1^{n_1}(1-q_1)^{n_0}p(H^1)}{q_0^{n_0}(1-q_0)^{n_1}p(H^0)}. \quad (2)$$

These are the odds that Jack is guilty, given the jury verdict. Since $n_0 = (n - \Delta)/2$ and $n_1 = (n + \Delta)/2$, the posterior odds depend both on the absolute margin Δ and on the jury size n .

For the posterior odds we can derive an inverse variant of Condorcet’s Theorem. If we let the jury size n go to infinity, and assume a fixed relative frequency $f_1 > 1/2$, we can derive the exact conditions under which the posterior odds for H^1 will tend to infinity as well. We can write

$$\frac{p(H^1|V_{n\Delta})}{p(H^0|V_{n\Delta})} = \left(\frac{q_1^{f_1}(1-q_1)^{(1-f_1)}}{q_0^{(1-f_1)}(1-q_0)^{f_1}} \right)^n \frac{p(H^1)}{p(H^0)} = L^n \frac{p(H^1)}{p(H^0)}.$$

For n going to infinity, these odds will tend to infinity on condition that $L > 1$, independently of the prior odds for H_1 . Now if $1/2 < q_1 \leq q_0 < 1$ (that is, the jurors are at least as competent at acquitting an innocent Jack as they are at condemning a guilty one), then this condition is satisfied. If, however, $1/2 < q_0 < q_1 < 1$ (that is, the jurors are

less competent at acquitting an innocent Jack than they are at condemning a guilty one), then the odds of a correct verdict tend to infinity in the infinite n limit only if f_1 is large enough.¹

2.3. Condorcet's formula

By making two further assumptions we can arrive at the result that is central to List (2004): the so-called Condorcet formula. First, we assume that the priors of H^0 and H^1 are equal (that is, $p(H^1) = p(H^0)$), although this is not a crucial assumption. It means that initially we judge it equally probable that Jack committed the murder as that he did not. Second, and this is crucial for List's result, we assume that the competences of jury members with respect to H^0 and H^1 are equal (that is, $q_0 = q_1 = q$), meaning that all jury members are precisely as reliable in condemning a murderer as they are in acquitting an innocent suspect. With these two assumptions, Equation (2) simplifies to the Condorcet formula

$$\frac{p(H^1|V_{n\Delta})}{p(H^0|V_{n\Delta})} = \left(\frac{q}{1-q} \right)^\Delta. \quad (3)$$

Because of the requirement $q > 1/2$, we have that $\left(\frac{q}{1-q} \right) > 1$ so that the posterior odds that Jack killed Jill are larger than 1 if $\Delta > 0$ and smaller than 1 if $\Delta < 0$. The posterior odds depend only on the absolute margin between the numbers of correct and incorrect votes brought out by the jury members, and not on their total number. This is perfectly consistent with both the Condorcet Theorem and its inverse version for posterior odds: for increasing jury size n and fixed competence q_1 , the expected value of Δ increases with n , and for increasing jury size n and fixed relative frequency f_1 , the value of Δ naturally also increases with n .²

Now focus on the fact that, for a given Δ , the posterior odds do not depend on the jury size. Imagine there are two juries, one with 10 members and one with 100 members. Suppose that both juries vote on the guilt of Jack, and that the 10-member jury unanimously votes for guilt while the 100-member jury casts 56 votes in favour of Jack's guilt and 44 against. Which of the two juries then makes the guilt of Jack more probable? Well, the majority in the former is less than the majority in the latter (that is, $\Delta_{10} = < 12 = \Delta_{100}$), so that, according to Equation (3), the probability of Jack's guilt is greater for the larger than for the smaller of the two juries. Hence, if we want to have as much certainty as we can get, apparently we should prefer the verdict of the larger jury.

As has been observed in Bovens and Hartmann (2003) and in Goodin and Estlund (2004), and as we will further argue below, there is something suspect in this conclusion. Indeed, we surmise that Equation (3) is too strong an idealization: it is based on the unwarranted assumption of a symmetrical and fixed competence. In the following, we provide a model in which both of these assumptions are dropped. Notably, a similar approach was taken by Goodin and Estlund, who developed a model in which the jury vote is first used to estimate the competence of the jurors, after which the estimated jury vote is employed to determine the support that the vote lends to the alternatives voted over. However, as we will explain in more detail in Section 4, their approach has some serious shortcomings.

3. A counter-intuitive consequence

In this section, we argue, by means of a classical statistical analysis, that there is indeed something suspect about the Condorcet formula in Equation (3). The results of the analysis will not be used in the remainder of this article, but we think they provide a good illustration of the problem that we aim to address. According to the Condorcet formula, only the absolute margin matters when one assesses the probability that a jury vote lends to the hypothesis voted over. The problem is that no account has been taken of how probable the jury votes are.

3.1. *Intuitive characterization*

We first present the problem with the Condorcet formula in non-technical terms. As was noted in the foregoing, if two juries reach a verdict on the case of Jack and Jill, then the Condorcet formula tells us that the support each of these juries lend to their verdict is determined by the difference in the number of voters for the two alternatives voted over. Of course, if one jury consisted of jurors who are better informed and more confident in their individual judgement than the other jury, then this might well influence the support that each of the juries confers. In particular, it may so happen that the competence of members of one jury is greater than the competence of members of the other jury.

Now we turn to the example of the two juries presented in the foregoing. We might already feel that the smaller, but unanimous jury somehow has an edge over the large and divided one. We can advance a variety of reasons for that intuition. In this article, we will develop the view that a jury vote tells us something more than just how the members in the jury think about the alternatives put to them, and that it also reveals how difficult it was for them to reach their judgement. This is relevant because, as suggested before, we should prefer a jury that is highly competent with respect to the case in question over a jury that does only marginally better than tossing a coin.

The discussion of the present section shows, by means of a classical statistical analysis, that we can in fact learn something about the competence of the jurors from the jury vote. In this section, we only learn something negative, namely, that the supposition that the competence of the jurors in the two juries of the example are equal cannot reasonably be maintained. This is most easily understood by looking at Figure 1, which is further explained in the note attached to the figure. In the following sections, we set up a model in which we learn much more than just that.

3.2. *A confidence interval for juror competence*

We now make this intuitive problem with the Condorcet formula mathematically precise by constructing a so-called confidence interval for juror competence that depends on both majority and jury size. As in the previous section, we assume that H^1 is true and that the competence parameter is q . Each juror votes independently and with identical probability, so that the number of votes n_1 has a binomial probability distribution. Its expectation is $E[n_1] = nq$, and the standard deviation is $SD[n_1] = \sqrt{nq(1-q)}$. So the mean and standard deviation of the majority Δ are

$$\begin{aligned} E[\Delta] &= E[n_1] - E[n_0] = 2E[n_1] - n = n(2q - 1), \\ SD[\Delta] &= 2\sqrt{nq(1 - q)}. \end{aligned}$$

For any given competence q and jury size n , we have a probability of roughly 95 percent that Δ lies within the specific bounds of two standard deviations around the mean, $E[\Delta] - 2SD[\Delta] < \Delta < E[\Delta] + 2SD[\Delta]$.

On this basis, we can construct a confidence interval for q . Suppose Δ lies at the edge of the interval indicated above. Then we would have one or other of the following:

$$\Delta = n(2q - 1) \pm 4\sqrt{nq(1 - q)},$$

which can be solved for q as a function of Δ , yielding the two roots

$$q_{\min}, q_{\max} = \frac{1}{2(n + 4)} \left(n + 4 + \Delta \pm \sqrt{n + 4 - \frac{\Delta^2}{n}} \right).$$

By way of interpretation, we say that any competence q inside the interval $[q_{\min}, q_{\max}]$ entails that the given majority Δ and jury size n are not so improbable that they are a cause for worry. If the $q_{\min}$ of a jury characterized by n and Δ is greater than the $q_{\max}$ of a jury with n' and Δ'

$$q_{\min}(n, \Delta) > q_{\max}(n', \Delta'),$$

then we can say that something very improbable has occurred: at least one of the two juries has in that case voted oddly. Moreover, this should give us cause to reconsider the assumptions of the statistical model at issue. Specifically, such a result would invite us to reconsider the assumption that the competences of the jurors from the two juries are the same.

3.3. Application to the example

Now let us have a second look at the example provided in Section 2. In Figure 1, the extremal values of the competence (that is $q_{\min}$ and $q_{\max}$, respectively) are plotted against the majority Δ for the two juries of 10 and of 100 members. For $n = 10$ and $\Delta = 10$, it turns out that $q_{\min}(10, 10) = 0.786$, while for $n = 100$ and $\Delta = 12$, we find $q_{\max}(100, 12) = 0.606$. With the given votes, we must therefore conclude that something highly improbable did occur. Since $q_{\max}(100, 50) = 0.783$ and $q_{\max}(100, 52) = 0.792$, we need a majority of at least 76 against 24 in the jury of 100 to feel that there is no cause for worry.

Now, perhaps we simply know the numerical value of juror competence. Alternatively, perhaps we know that all jurors have equal competence without knowing its value, in which case we might say that the competence of the jury is larger than is suggested by the larger jury, or that the competence of the jurors is much smaller than is suggested by the smaller jury, or possibly both. However, in any such case the result of List is applicable, and we must simply conclude that we have witnessed a freak accident.

Figure 1. Note: This is a graph of the bandwidth of reasonable values of the competence q against the size of the jury majority Δ for jury sizes of $n = 10$ and $n = 100$. For each jury separately, the areas within the ellipses indicate which values of the competence can be considered reasonable, depending on the size of the majority, that is, the absolute difference between the number of voters for the two alternatives. Along the vertical line at $\Delta = 10$, we find that the reasonable values for the competence of jurors from the jury with size 10 are all above 0.783, whereas the reasonable values for the competence of the jurors from the jury with size 100 are all below 0.606. In other words, we cannot reasonably suppose that jurors from the two juries have equal competence.

Alternatively, we might conjecture that the two juries have different values for juror competence. In order to sort this out, we can formulate the hypothesis that the jurors from the smaller jury are in fact more competent, and perform a statistical test on this. But the idea that the jury vote tells us something about the competence of the jurors can also be taken a step further. We can say that the vote of the jury indicates something about the competence of the jury directly, and this knowledge may be used with advantage in the choice between jury verdicts. Specifically, the unanimous vote of the jury of 10 should perhaps weigh more heavily, despite the rule of Equation (3), simply because the unanimity suggests that the jurors are competent.

In other words, the suggestion here is that a jury vote reflects more than just the truth or falsity of the hypothesis voted over. It also conveys information on how easy it is for jurors to vote correctly. A close call in the jury, such as the small majority of 12 in the jury of 100 members, indicates that the jurors find it hard to tell whether Jack murdered Jill, while the unanimous vote of the small jury seems to suggest that the jurors find Jack's guilt fairly clear. The main contribution of this article is in making precise what the size of the majority tells us about the competence of the jurors, and what the consequences are for assessing the jury vote.

4. Alternative explanations

Before embarking on this programme, however, we briefly consider some alternative explanations for the jury votes considered above.

4.1. *Asymmetry or heterogeneity of competence*

List (2004) shows that if we do not make the assumption of symmetrical competence, $q_0 = q_1$, but instead let these competences vary independently, the posterior odds do depend on the jury size. Depending on which of the two competences is larger, a larger jury with an equal absolute margin will have smaller or larger posterior odds for the hypothesis. By choosing the competences in the right way, this effect may even cause the posterior odds to lean towards the minority vote. However, the model with differing, but fixed competences q_0 and q_1 fails to capture the intuitions on jury verdicts voiced above. It introduces a dependence on jury size of an entirely different nature, one that does not strike us as getting at the root of the difficulty.

It may seem possible to provide a partial explanation for the above jury votes by including the heterogeneity of the jurors in the model (see, for example, Berend and Paroush, 1998; Dietrich, 2008; Ladha, 1993). Intuitively, if we assume a certain spread in the competence of jurors, as expressed in an assumed probability distribution over the competence parameter, then the vote from a larger jury will have added value. Just as in the case of the fixed, single competence, the mistakes of jury members are washed out by larger numbers. But in addition, the uncertainty stemming from randomly sampling the jurors is also diminished. Because of this added effect, it is to be expected that both the absolute majority and the jury size play a role in the eventual trust we put in the jury vote.

The exact effects of this on the confidence we have in a jury are certainly worth exploring, and in our concluding remarks we shall indicate how that might be done. However, for present purposes, it is more important to emphasize that, whatever the modelling details, the inclusion of competence heterogeneity is not going to accommodate the intuitions voiced in Section 3. At first glance, there may be a striking similarity between the probability distribution over competences that expresses this heterogeneity and the distribution over competences that we shall be using in this article. However, the latter distribution expresses our ignorance over the value of the competence of a homogeneous jury, and this is very different from the real variability of competences within a heterogeneous jury. Accordingly, in this article we adapt the distribution over competences in the light of the jury vote, whereas a fixed distribution is assumed in the models that try to take account of heterogeneity. In this respect, the model of the present article differs from almost all earlier models.

4.2. *Uncertainty concerning competence*

Goodin and Estlund (2004) are an exception. They notice a dependence between the jury vote and the competence very similar to what we noted in Section 3, and they argue that the competence of the jurors can be determined from the jury vote itself. However, the way they determine the competence is markedly different from ours.

We describe the proposal of Goodin and Estlund in some detail. As before, it is assumed that the jury members choose between two alternatives, H^0 and H^1 , that a total number n_0 choose H^0 while a number n_1 choose H^1 , and that $n_1 > n_0$. According to Goodin and Estlund (2004), the competence of jurors can be estimated by

$$\hat{q}_j = \frac{n_j}{n}.$$

That is, the proportion of voters that choose a particular option is an estimate of how competent the voters are on that particular alternative. Note, however, that the proportion can be read in two different ways. It may be that H^1 is the correct alternative and that voters are competent, choosing H^1 with a majority so that $q_1 > 1/2$. But it may also be that H^0 is the correct alternative and that the voters are incompetent, choosing H^0 with a minority so that $q_0 < 1/2$.

After providing these estimates for the competences on the basis of the vote, Goodin and Estlund (2004) discuss the support that a jury lends to the alternatives. Importantly, they show sensitivity to the fact that we do not know the competence in advance, and they take into consideration that the proportion of voters indicates this competence. This goes in the same direction as we are proposing in the following. However, to our mind, the method of estimation is far too simplistic.

The problem with their method is that from the jury vote we cannot conclude with certainty that the jury has one of the two estimated competences. Estimating the competence in the above way may lead to a good approximation of the support given by a jury, especially when the jury is large. But a more accurate expression of that support must also incorporate the epistemic uncertainty that surrounds the competence. An increase in jury size will lead to more accurate estimations, but in the model of Goodin and Estlund this fact cannot be taken into account. The analysis of the present article fares much better in this respect.

4.3. Coherent voting

We might consider an extension of the model of jury decisions in an entirely different direction, namely, by taking into account the interdependence of jurors. Arguably, a small jury has a completely different group dynamic than a larger jury, and the jury verdict may well reflect how the jurors have interacted. For example, it may be that jurors adjust their views to coincide with those of jurors sitting in close proximity to them. In a jury of 100, the jurors may then still be treated as approximately independent. But in a jury of 10, all jurors are in close proximity to each other. Therefore, a unanimous vote in a small jury may very well be the result of mindless groupthink rather than of high juror competence. Such failures of independence will generally throw the results of a jury vote into a different perspective.

We may also think that coherence is truth-conducive, and hence that the coherence of the jurors in the smaller jury is indicative of the veracity of the jury verdict. This idea is at the basis of the discussion that Bovens and Hartmann (2003: Section 3.6) give of the Condorcet formula. They also note its counter-intuitive consequences, adapt the model to include a positive correlation between the votes, and then show that in this model a smaller unanimous jury lends more credibility to the jury verdict than a larger jury whose verdict is divided, even when the absolute margins in both juries are equal. Moreover, by adapting the parameters in the model they can vary the degree to which the coherence of jurors adds to the credibility of the verdict.

The coherence model of Bovens and Hartmann provides a successful explication of some such intuitions concerning jury votes. It sensibly drops the assumption of the independence of the jurors, and employs the truth-conduciveness of the coherence of votes to avoid the counter-intuitive consequence of the Condorcet formula. A drawback of this solution is that it relies on particular parameter values that must be filled in at the start. Given these parameters, we can deduce the dependence of the posterior odds on the jury size, but this dependence is in a sense put in by hand. Yet, this drawback does not mean that we should discard the coherence model completely.

We do not motivate the model found in the following sections by a claim that it captures our intuitions on jury votes better. Rather, it captures another intuition about jury votes, differing from those captured in the coherence model of Bovens and Hartmann (2003), but similar to those voiced by Goodin and Estlund (2004). It is the idea that the competence of jurors can be partially revealed by the jury vote, as was indicated by the classical statistical analysis of Section 3. We think that this idea is of interest in its own right, and that it merits a more extensive treatment than was given to it by Goodin and Estlund.

5. Jury vote with unknown competence

In the following, we present a model in which the jury vote is indicative of how competent the jurors are concerning the hypothesis at hand. We retain the assumption that the jurors vote independently and concentrate on relaxing the assumption of single-valued juror competences. To do this, we employ Bayesian statistical inference. We first compute a posterior probability assignment over the competences q_0 and q_1 for H^0 and H^1 respectively, based on the given jury vote and a prior probability over competences and hypotheses. This inference determines how the jury vote informs us of the competence: we may derive an expectation value for the juror competences from it. More importantly, we compute the probability that a jury vote gives to the hypothesis voted over.

5.1. Learning juror competence

Before we make the model and its consequences formally explicit, we give an informal characterization of it. Readers who have no interest in the mathematical details may skip Subsection 5.2.

In Section 3 we argued that a jury vote might reflect more than just the truth or falsity of the hypothesis voted over: the vote also conveys information on how easy it is for jurors to vote correctly. A close call in the jury, such as the small majority of 12 in the jury of 100 members, might indicate that the jurors find it hard to tell whether Jack murdered Jill, while the unanimous vote of the small jury of 10 seems to suggest that the jurors find Jack's guilt all too clear. The model suggested in this article aims to make precise what the relative size of the majority tells us about the competence of the jurors and what the consequences are for assessing the jury vote.

We consider two alternatives and a jury voting between them, and we assume that one of them is correct. For each of the alternatives, we assume the voters have some

competence, meaning that there is some fixed chance that, if that alternative is correct, the voter chooses in favour that alternative. We assume that jury members are at least as good at choosing correctly as the toss of a fair coin. The crucial difference with earlier models in the literature on the Condorcet Jury Theorem is that at the start of the jury vote we are ignorant of juror competence: every value of the competence between $\frac{1}{2}$ and 1 is deemed possible and is assigned some probability. We do not assume some specific value for the competence.

The use of a probability assignment over possible values of the competence allows us to express the impact of a particular jury vote on our estimations of juror competence. In fact, it allows us much more detail in what we learn from the jury vote than if we merely produce a best estimate, as proposed by Goodin and Estlund (2004). The model that is introduced below presents us with a probability assignment over possible competences after the jury vote, telling us what value is most probable for the competence, and by the shape of the probability assignment over competences, it also tells us how reliable that estimate is. Moreover, based on this probability assignment over the possible values of juror competence we can determine the support that the jury vote lends to the two alternatives voted over. In the end, we provide a probability for the two alternatives that takes into account both the jury vote itself and the fact that the jury vote tells us something about how competent the jurors are.

5.2. Mathematical model

Calculating the support that a jury lends to a hypothesis is a tricky business. In the foregoing, we simply had the partition of two hypotheses, H^0 and H^1 . But since the competence parameter is unknown, we must split these hypotheses up into continuous ranges of hypotheses, H_0 and H_1 . The hypotheses H^0 and H^1 are composite, each consisting of a range of more detailed statistical hypotheses, H_j , parameterized by the specific competences q_j . These competences are given real values, so the expressions $p(H_j)$ should therefore be regarded as probability densities rather than being themselves probabilities. The hypotheses H_{q_j} have the likelihoods

$$p(V_j^i | H_{q_j} \cap V_j^{i'}) = q_j,$$

for $j = 0, 1$. The probabilities of the aggregate hypotheses H^0 and H^1 are

$$p(H^0) = \int_0^1 p(H_{q_0}) dq_0,$$

$$p(H^1) = \int_0^1 p(H_{q_1}) dq_1,$$

The general idea of what follows is that we first compute the posterior probability over the more detailed hypotheses H_{q_j} on the basis of the jury vote, and then calculate the posteriors for the composite hypotheses by working out these integrals.

In the following, we assume for both q_0 and q_1 that the prior is equal and uniform over the interval $(\frac{1}{2}, 1]$, meaning that $p(H_{q_j}) = 1$ for $\frac{1}{2} < q_j \leq 1$ and $p(H_{q_j}) = 0$ for $0 \leq q_j \leq \frac{1}{2}$.

Thus the only prior assumption is that the jury members are not incompetent, but aside from that the prior density is flat, as an expression of the fact that, *prima facie*, we consider each competence value in the interval $[\frac{1}{2}, 1]$ equally probable. The above considerations entail

$$p(H^0) = \int_{\frac{1}{2}}^1 1 dq_0 = \frac{1}{2} = \int_{\frac{1}{2}}^1 1 dq_1 = p(H^1)$$

For reasons of simplicity we will not deviate from the uniformity assumption in what follows. The main results of this article do not hinge on this assumption.

It is convenient to reduce the number of parameters in this statistical model to a single one by a suitable substitution of the parameters over the domain. Note that in the above setup the use of the two parameters q_0 and q_1 does not mean that the statistical model is two-dimensional. The likelihoods involve q_0 if H^0 is true and q_1 if H^1 is true, but these are mutually exclusive hypotheses, so there is no overlap in which the likelihoods involve both parameters. Because of this we can employ a single range of hypotheses, H_r , with the parameter domain $r \in [0, 1]$, which is formally equivalent to the combination of q_0 in H^0 and q_1 in H^1 .

Let us make this formal equivalence precise.³ First, within the domain $r \in [0, \frac{1}{2})$, and setting $q_0 = 1 - r$, we have the following equalities:

$$\begin{aligned} p(V_i^0 | H_{q_0}) &= q_0 = 1 - r = p(V_i^0 | H_r) \\ p(V_i^1 | H_{q_0}) &= 1 - q_0 = r = p(V_i^1 | H_r). \end{aligned} \quad (4)$$

Similarly, in the domain $r \in (\frac{1}{2}, 1]$, and setting $q_1 = r$, we have the following equalities:

$$\begin{aligned} p(V_i^0 | H_{q_1}) &= 1 - q_1 = 1 - r = p(V_i^0 | H_r) \\ p(V_i^1 | H_{q_1}) &= q_1 = r = p(V_i^1 | H_r). \end{aligned} \quad (5)$$

Put into words, there is a formal equivalence between the likelihoods of the hypotheses H_r for $r < \frac{1}{2}$ and those of H_{q_0} for $q_0 > \frac{1}{2}$. Similarly, there is an equivalence between the likelihoods of the hypotheses H_r for $r > \frac{1}{2}$ and those of H_{q_1} for $q_1 > \frac{1}{2}$.

Now consider the resulting likelihoods for the hypotheses H_r . From the right-hand side of Equations (4) and (5) we can see that over the entire domain $r \in [0, 1]$, the hypotheses H_r have the likelihoods

$$\begin{aligned} p(V_i^0 | H_r) &= 1 - r, \\ p(V_i^1 | H_r) &= r. \end{aligned}$$

By updating the separate hypotheses H_r according to these likelihoods, we are effectively updating the hypotheses H_{q_0} and H_{q_1} for each of the values $q_0 \in (\frac{1}{2}, 1]$ and $q_1 \in (\frac{1}{2}, 1]$.

Next we consider the priors over the hypotheses H_{q_0} and H_{q_1} . Recall that we assumed a uniform prior probability distribution over both of them. But we can rewrite these priors in terms of priors over the hypotheses H_r in the respective domains $r \in [0, \frac{1}{2})$ and $r \in (\frac{1}{2}, 1]$ as follows.

$$p(H^0) = \int_{1/2}^1 p(H_{q_0}) d_{q_0} = \int_0^{1/2} p(H_r) dr,$$

$$p(H^1) = \int_{1/2}^1 p(H_{q_1}) d_{q_1} = \int_{1/2}^1 p(H_r) dr.$$

Hence, the uniform priors over the hypotheses H_{q_0} and H_{q_1} translate into a single uniform prior over the hypotheses H_r with $r \in [0, 1]$. We can interpret the probability of H_r with $r < 1/2$ as the probability of H_{q_0} by the translation $q_0 = 1 - r$ and, similarly, we can interpret the probability of H_r with $r > 1/2$ as the probability of H_{q_1} by the translation $q_1 = r$. We note, as an aside, that it is attractive to start out with uniform priors over the hypotheses H_{q_j} , or at least with priors that combine into a Beta distribution over H_r . Priors over H_{q_j} that have a different shape do not necessarily lead to posterior distributions that can be expressed in terms of a canonical function.

The substitution above is useful because we have thereby disposed of a parameter, replacing q_0 and q_1 by the single parameter r . Moreover, we can model the impact of the jury vote on the combined uniform probability assignments over $q_0 \in (1/2, 1]$ and $q_1 \in (1/2, 1]$ by modelling its impact on the uniform probability assignment over $r \in [0, 1]$.

We shall condition this distribution on the jury vote $V_{n\Delta}$, characterized by the number of votes n_0 for H^0 and n_1 for H^1 or, equivalently, by the size of the jury $n = n_1 + n_0$ and the majority $\Delta = n_1 - n_0$. Then the posterior probability distribution over H_r results in a well-known form for the posterior distribution, the Beta distribution

$$p(H_r | V_{n\Delta}) = \frac{(n+1)!}{n_0!n_1!} r^{n_1} (1-r)^{n_0},$$

with $r \in [0, 1]$. For $r > 1/2$ we are thereby indirectly specifying the posterior probability distribution over the hypotheses H_{q_1} according to the transformation $q_1 = r$, while for $r < 1/2$ we are indirectly specifying the posterior for the hypotheses H_{q_0} , using the transformation $q_0 = 1 - r$.

From this expression we can derive the posterior probability of the hypotheses H^0 and H^1 :

$$p(H^0 | V_{n\Delta}) = \frac{(n+1)!}{n_0!n_1!} \int_0^{1/2} r^{n_1} (1-r)^{n_0} dr = 1 - p(H^1 | V_{n\Delta}). \quad (6)$$

This can be written in terms of the jury size n and the majority Δ using $n_0 = (n-\Delta)/2$ and $n_1 = (n+\Delta)/2$. The expectation values for the competences of the jurors, finally, are given by the following normalized integrals:

$$E(q_0) = \frac{1}{p(H^0 | V_{n\Delta})} \int_0^{1/2} r^{n_1} (1-r)^{n_0+1} dr,$$

$$E(q_1) = \frac{1}{p(H^1 | V_{n\Delta})} \int_{1/2}^1 r^{n_1+1} (1-r)^{n_0} dr.$$

If $n_1 > n_0$, then we will have that $E[q_1] > E[q_0]$, because on the assumption that H^0 is true, a majority for H^1 is more likely if the competence q_0 is low.

Before we investigate Equation (6) in the next section, we want to address a possible criticism of the derivation of the posteriors. It may be objected that the jury vote is used twice: once for the determination of the posterior over competences and then again for the determination of a posterior for the hypotheses based on some expected competence. But in the model above there is no such double usage. We only employ the vote to determine a probability distribution over the parameter r , which summarizes the two competences q_0 and q_1 . All the other probability assignments are derived from this distribution without using the data again.

6. Calculating the posterior probability

In the preceding section we derived a probability assignment for the hypotheses concerning Jack's guilt, under the assumption of the independence of jurors, but without assuming a uniform value for the competences q_0 and q_1 . The assignment is an expression in which both the size of the majority and the jury size play a role. In contrast, in Section 2 we presented the result of List (2004) that, on the assumption of any particular competence $q_0 = q_1 = q$, the probability of the hypotheses only depends on the majority.

In this section, we investigate the consequences of Equation (6) both analytically and numerically. The implications of the present model are in accordance with our intuitions on the relation between jury votes and the hypothesis voted over: if, in a large jury, we have a close call between the votes for the two alternatives, this is taken as a sign that it is hard to decide between the hypotheses, that is, that the competence of the jurors is therefore probably low, and accordingly it is taken as a reason to put less trust in how the jury has voted than the trust suggested by the absolute margin alone. We thereby provide a more accurate model of an aspect of jury voting that was initially noticed by Goodin and Estlund (2004).

6.1. Analytic results

We first give an analytical characterization of how the probability for the hypotheses depends on jury size and absolute margin. Significantly, we retain an important consequence of the Condorcet formula, as discussed in Section 2. On the assumption that $n_1 > n_0$, or $\Delta > 0$, we have the pairwise inequality

$$p(H_r|V_{n\Delta}) > p(H_{1-r}|V_{n\Delta})$$

for all $r \in (\frac{1}{2}, 1]$. Hence, we also have that

$$\int_{\frac{1}{2}}^1 p(H_r|V_{n\Delta})dr > \int_{\frac{1}{2}}^1 p(H_{1-r}|V_{n\Delta})dr.$$

Via the translation $1 - r = q_0$ and $r = q_1$ within $r \in (\frac{1}{2}, 1]$, we thus obtain the inequality $p(H^1|V_{n\Delta}) > p(H^0|V_{n\Delta})$ on condition that $\Delta > 0$.

Further, let us look at the so-called marginal likelihoods of the hypotheses H_j on the two votes $V_{n+1}^0 \cap V_{n+2}^1$. These two votes effectively enlarge the jury while keeping the absolute margin Δ fixed. In Appendix 1, it is shown that if $\Delta = n_1 - n_0 > 0$, we have

$$p(V_{n+1}^0 \cap V_{n+2}^1 | H^0 \cap V_{n\Delta}) > p(V_{n+1}^0 \cap V_{n+2}^1 | H^1 \cap V_{n\Delta}). \quad (7)$$

Since we also have

$$\frac{p(H^1 | V_{n+2,\Delta})}{p(H^0 | V_{n+2,\Delta})} = \frac{p(V_{n+1}^0 \cap V_{n+2}^1 | H^1 \cap V_{n\Delta})}{p(V_{n+1}^0 \cap V_{n+2}^1 | H^0 \cap V_{n\Delta})} \times \frac{p(H^1 | V_{n\Delta})}{p(H^0 | V_{n\Delta})},$$

the inequality of Equation (7) entails that the odds of Jack's guilt decrease monotonically as the jury size n is increased, if we hold the absolute margin $\Delta > 0$ fixed. This is in accordance with the intuitions voiced in the preceding sections, in particular that the jury size affects the probability of the hypothesis.

Second, we investigate the limiting behaviour in order to arrive at a generalization of Condorcet's Theorem for posterior odds. The mere fact that $p(H^0 | V_{n\Delta})$ increases with the jury size for fixed $\Delta > 0$ does not yet determine the limiting value for $p(H^0 | V_{n\Delta})$ as n goes to infinity. However, as shown in Appendix 2, if the absolute margin Δ is held constant in the limit, we find the asymptotic behaviour

$$\lim_{n \rightarrow \infty} p(H^0 | V_{n\Delta}) = \frac{1}{2}.$$

It is further shown that if the fractional majority, $f = \Delta/n$, is held constant in the limit, we have instead

$$\lim_{n \rightarrow \infty} p(H^0 | V_{n, nf}) = 0.$$

Finally, we show that for the latter limiting behaviour, it is not necessary for the majority to increase linearly with n . It is enough if Δ increases more quickly than $\sqrt{n}$ to ensure that $p(H^0 | V_{n\Delta})$ tends to zero.

6.2. Numerical results

In addition to these qualitative results, we have also done some numerical calculations with the aid of Mathematica.

The integral in Equation (6) can be written as a known transcendental function, the so-called regularized incomplete Beta function. In Figure 2, we have plotted the relation between the probability $p(H^0 | V_{n\Delta})$ and the jury size $n = n_1 + n_0$ for various values of the size of the majority $\Delta = n_1 - n_0$. These calculations illuminate the case of the two juries considered at the beginning of this article. With a unanimous verdict of guilty in a jury of 10 ($\Delta = 10$ and $n = 10$), for example, we find the probability $p(H^1 | V_{10,10}) = 0.9995$. For a jury of 100 with a majority of 12 for guilty, on the other hand, we calculate a smaller probability that the jury verdict is correct, namely, $p(H^1 | V_{100,12}) = 0.8839$. The important point here is that the probability depends not only on the majority, as it did when we chose some fixed competence. It decreases as n is increased, and this effect may counter-balance a difference in the size of the majority.

Figure 2. Note: The plots show the probability of the alternative H^0 after incorporating the jury vote, $p(H^0|V_{n\Delta})$, as a function of the jury size, n , for a range of values of the absolute difference between the number of voters for the two alternatives ($\Delta = 2$, $\Delta = 5$, $\Delta = 10$, and $\Delta = 12$). In all these cases, the majority of the jury voted for alternative H^1 . The salient aspect of the plots is that for a fixed value of the absolute majority, the probability of the alternative supported by the minority, H^0 , increases with the size of the jury. The intuitive reason for this is that, for fixed majority size, a larger jury size signals a lower juror competence.

In Figure 3, we see the dependence of the probability $p(H^0|V_{n\Delta})$ on the majority size Δ spelled out in more detail. Note, first, that for $\Delta = 0$, this probability is one-half, as it should be. Furthermore, for any fixed jury size, the probability of the hypothesis H^0 decreases with increasing majority size. Finally, for fixed majority size Δ and increasing jury size n , we can see that the probability $p(H^0|V_{n\Delta})$ slowly increases towards a half again. For very large juries, as also suggested by Figure 2, a small majority does not carry much weight.

7. Concluding remarks

Now that we have obtained these results, what can we say of earlier results in the literature? Recall that the model of Goodin and Estlund (2004) is motivated by the same intuition as the model presented in this article. However, we claim that the present model is more convincing than the one that Goodin and Estlund provide, because it takes into account both the learning of juror competence and the uncertainty associated with this learning. That is, unlike Goodin and Estlund, our model accommodates the fact that juror competence may not be reflected exactly in the proportion of jurors. We think this is an improvement over their model.

The main result of List (2004) was that the probability of the hypotheses voted over only depends on the absolute margin Δ . Of course, this is still valid under the assumption of fixed symmetrical competence. But with the foregoing considerations in mind, we see that if we do not know the competences q_0 and q_1 , and if we decide to learn about these competences on the basis of the jury vote, then both the absolute majority and the jury size matter. The model presented in this article thus provides an alternative account of how the confidence we may have in a jury vote depends on the size of the jury as well

Figure 3. Note: This graph shows the probability of the alternative H^0 after incorporating the jury vote, $p(H^0 | V_{n\Delta})$, as a function of the jury size n and the size of the majority Δ . As in Figure 2, we see that for any majority size, the probability of H^0 increases.

as on the absolute majority, in emendation of the mechanism provided by Bovens and Hartmann (2003).

We conclude with some suggestions on how to develop the results of the present article. First, we think that it is important for the practical applicability of Condorcet-style results to relax the assumption of the theorems concerning the independence of jurors. As mentioned above, Bovens and Hartmann (2003) successfully model a jury vote with dependent jurors, and it will be interesting to see if that model can be combined with the model presented in this article. Another way to incorporate jury dynamics into the analysis is presented by adapting the prior, so as to make it less sensitive to almost or entirely unanimous voting. As indicated before, we might think that unanimity in small juries is due to mindless groupthink and not a sign of a high juror competence or of truth-conducive coherence. If so, we could correct for a possible overestimation of competence by choosing a prior over competences that is non-uniform, but peaked around a value of r close to half, but such a strategy would only be warranted if we were to have an independent motivation for such a prior.

Perhaps a more accurate way of modelling the interaction between group members is by dropping the assumption on the likelihood that is expressed in the left-hand equality of Equation (1) altogether. In the model presented in this article, jury votes $V_{n\Delta}$ are characterized by the numbers n and Δ only, because the likelihoods of the hypotheses H^j only depend on these numbers. But we might also partition the space of possible jury votes differently, according to other characteristics of the votes, and employ hypotheses that have more complicated likelihood functions over that space. While this will no doubt provide interesting new insights, we can scarcely hope to attain analytical results for a model with these more involved statistical hypotheses.

An entirely different line of research concerns the real variation of competences within the jury, as discussed earlier in this article. For example, Dietrich (2008) shows

that the classical Condorcet Jury Theorem still holds if we suppose that the competences of jurors vary, as long as their average competence is larger than a half. This raises the question whether we can also derive an expression for the posterior odds of the hypothesis on the assumption of a certain spread in the competences of the respective jurors. A suitable statistical setting for answering this question is hierarchical Bayesian modeling, in which we may suppose the juror relative competence q_{ij} to be drawn at random from a distribution of possible values for the competence. Again, analytical results may be very hard to come by, but software packages such as WinBUGS are well equipped to investigate such models numerically.

Finally, we expect that much can be gained by applying the present insights to the discussion over the coherence measures proposed in Bovens and Hartmann (2003), and continued in Haenni and Hartmann (2006). The reliability parameter employed there is formally similar to the competence parameter employed in the present article. It will be interesting to see if and how we can adapt our estimations of the reliability of measurement apparatuses or witnesses from their coherence. Because the well-known impossibility result of Bovens and Hartmann (2003) relies on variability in the reliability parameter and because in the present article we have shown how to adapt the probability assignment over the values of this parameter, we expect that casting their impossibility result in terms of the present findings will lead to interesting new insights.

Appendix I. Relative size of the marginal likelihoods

In this appendix, we prove that the marginal likelihood of the hypothesis H^0 for the combined votes $V_{n+1}^0 \cap V_{n+2}^1$, given an earlier jury vote $V_{n\Delta}$ for which $\Delta > 0$, is larger than the corresponding likelihood of the hypothesis H^1 . Mathematically, this may be expressed as

$$p(V_{n+1}^0 \cap V_{n+2}^1 | H^0 \cap V_{n\Delta}) > p(V_{n+1}^0 \cap V_{n+2}^1 | H^1 \cap V_{n\Delta}). \quad (8)$$

We will do so by writing out the marginal likelihoods in terms of the likelihoods of the statistical hypotheses H_r for $r < \frac{1}{2}$ and $r \geq \frac{1}{2}$, respectively.

We first determine the likelihoods of the hypotheses H_r for the two votes $V_{n+1}^0 \cap V_{n+2}^1$:

$$p(V_{n+1}^0 \cap V_{n+2}^1 | H_r \cap V_{n\Delta}) = r(1 - r).$$

Recall that the hypotheses H^0 and H^1 are composed of the statistical hypotheses H_r . The likelihood of the hypothesis H^0 for the two votes $V_{n+1}^0 \cap V_{n+2}^1$ is

$$\begin{aligned} p(V_{n+1}^0 \cap V_{n+2}^1 | H^0 \cap V_{n\Delta}) &= \int_0^{\frac{1}{2}} p(H_r | H^0 \cap V_{n\Delta}) p(V_{n+1}^0 \cap V_{n+2}^1 | H_r \cap H^0 \cap V_{n\Delta}) dr \\ &= \frac{1}{p(H^0 | V_{n\Delta})} \frac{(n+1)!}{n_0! n_1!} \int_0^{\frac{1}{2}} r^{n_1} (1-r)^{n_0} r(1-r) dr. \end{aligned} \quad (9)$$

in which we have used the normalization $p(H^0 | V_{n\Delta})$. Following Equation (6), we have within $r \in [0, \frac{1}{2}]$ that

$$\begin{aligned}
 p(H_r|H^0 \cap V_{n\Delta}) &= \frac{p(H_r \cap H^0|V_{n\Delta})}{p(H^0|V_{n\Delta})} = \frac{p(H_r|V_{n\Delta})}{p(H^0|V_{n\Delta})} \\
 &= \frac{1}{p(H^0|V_{n\Delta})} \frac{(n+1)!}{n_0!n_1!} r^{n_1} (1-r)^{n_0}.
 \end{aligned} \tag{10}$$

The marginal likelihood of the hypothesis H^1 for the two votes is given by a similar expression, with the differences that the integration bounds are $\frac{1}{2}$ and 1 and that the normalization is $p(H^1|V_{n\Delta})$.

In order to compare the two marginal likelihoods, it will be convenient to write the likelihood of H_1 in terms of the same integration bounds, making use of the symmetry in the integral expression:

$$\begin{aligned}
 p(V_{n+1}^0 \cap V_{n+2}^1 | H^1 \cap V_{n\Delta}) &= \frac{1}{p(H^1|V_{n\Delta})} \frac{(n+1)!}{n_0!n_1!} \int_{\frac{1}{2}}^1 r^{n_1} (1-r)^{n_0} r(1-r) dr \\
 &= \frac{1}{p(H^1|V_{n\Delta})} \frac{(n+1)!}{n_0!n_1!} \int_0^{\frac{1}{2}} r^{n_0} (1-r)^{n_1} r(1-r) dr.
 \end{aligned} \tag{11}$$

We can now compare the two marginal likelihoods by comparing the functions appearing under the integration sign. Specifically, we will investigate the expression

$$\begin{aligned}
 &p(V_{n+1}^0 \cap V_{n+2}^1 | H^0 \cap V_{n\Delta}) - p(V_{n+1}^0 \cap V_{n+2}^1 | H^1 \cap V_{n\Delta}) \\
 &= \int_0^{\frac{1}{2}} (p(H_r|H^0 \cap V_{n\Delta}) - p(H_r|H^1 \cap V_{n\Delta})) r(1-r) dr \\
 &= \frac{(n+1)!}{n_0!n_1!} \int_0^{\frac{1}{2}} \left(\frac{r^{n_1} (1-r)^{n_0}}{p(H^0|V_{n\Delta})} - \frac{r^{n_0} (1-r)^{n_1}}{p(H^1|V_{n\Delta})} \right) r(1-r) dr,
 \end{aligned} \tag{12}$$

which represents the difference between the marginal likelihoods of Equations (9) and (11). If this function is positive, then the marginal likelihood of H^0 is larger than that of H^1 , which is what we have set out to prove.

The expression inside the integral of Equation (12) consists of two parts. We now make some observations on the part between brackets:

$$g(n_0, n_1, r) = \frac{r^{n_1} (1-r)^{n_0}}{p(H^0|V_{n\Delta})} - \frac{r^{n_0} (1-r)^{n_1}}{p(H^1|V_{n\Delta})}.$$

First, because of the normalizations, $p(H^j|V_{n\Delta})$, we have

$$\int_0^{\frac{1}{2}} g(n_0, n_1, r) dr = 0. \tag{13}$$

Next, if we assume that $n_1 > n_0$, Equation (7) says that $p(H^1|V_{n\Delta}) > p(H^0|V_{n\Delta})$, so that we have

$$g(n_0, n_1, 1/2) = \left(\frac{1}{p(H^0|V_{n\Delta})} - \frac{1}{p(H^1|V_{n\Delta})} \right) \frac{1}{2^n} > 0. \tag{14}$$

Figure 4. Note: The figure plots the functions $p(H_r|H^j \cap V_{n\Delta})$ for $j = 0, 1$ against $r \in [0, \frac{1}{2}]$. The values of n and Δ are kept fixed. As expressed in Equation (16), the two areas in between the two curves are equal.

Furthermore, the equation $g(n_0, n_1, r) = 0$ has two solutions in r . One is $r = 0$; the other is

$$r^* = \frac{c}{1+c} \quad \text{with } c = \left(\frac{p(H^0|V_{n\Delta})}{p(H^1|V_{n\Delta})} \right)^{1/\Delta}. \quad (15)$$

Together with Equations (12) and (14), Equation (15) entails that in the domain $r \in [0, r^*]$ we have that $g(n_0, n_1, r) < 0$, while in $r \in [r^*, \frac{1}{2}]$ we have that $g(n_0, n_1, r) > 0$. Finally, with Equation (13) this entails that

$$-\int_0^{r^*} g(n_0, n_1, r) dr = \int_{r^*}^{\frac{1}{2}} g(n_0, n_1, r) dr. \quad (16)$$

In other words, the entire negative contribution to the integral of Equation (13) lies in $r < r^*$, while the entire positive contribution to it lies in $r > r^*$. All this is illustrated in Figure 4.

We make one further observation on the function $r(1-r)$, namely, that it is monotonically increasing in r over the domain $r \in [0, \frac{1}{2}]$. Now recall that in the domain $r \in [0, r^*]$, the contribution of the integral is entirely negative. The factor with which the function $g(n_0, n_1, r)$ is multiplied over this domain is, on average, strictly less than $r^*(1-r^*)$, and the contribution to the whole integral of Equation (12), therefore, has the following lower bound:

$$\int_0^{r^*} g(n_0, n_1, r) r(1-r) dr > r^*(1-r^*) \int_0^{r^*} g(n_0, n_1, r) dr. \quad (17)$$

In the domain $r \in [r^*, \frac{1}{2}]$, on the other hand, the integral is entirely positive, and the factor with which the function $g(n_0, n_1, r)$ is multiplied is, on average, strictly more than $r^*(1 - r^*)$, thus leading to a contribution with a lower bound

$$\int_{r^*}^{\frac{1}{2}} 2g(n_0, n_1, r)r(1 - r)dr > r^*(1 - r^*) \int_{r^*}^{\frac{1}{2}} 2g(n_0, n_1, r)dr. \quad (18)$$

Combining these two equations, we have a lower bound of the difference between the two marginal likelihoods covering the entire domain of r .

Hence we can determine the lower bound of the difference between the marginal likelihoods, as follows.

$$\begin{aligned} & p(V_{n+1}^0 \cap V_{n+2}^1 | H^0 \cap V_{n\Delta}) - p(V_{n+1}^0 \cap V_{n+2}^1 | H^1 \cap V_{n\Delta}) \\ &= \int_0^{\frac{1}{2}} g(n_0, n_1, r)r(1 - r)dr \\ &= \int_0^{r^*} g(n_0, n_1, r)r(1 - r)dr + \int_{r^*}^{\frac{1}{2}} g(n_0, n_1, r)r(1 - r)dr \\ &> r^*(1 - r^*) \left(\int_0^{r^*} g(n_0, n_1, r)dr + \int_{r^*}^{\frac{1}{2}} g(n_0, n_1, r)dr \right) = 0. \end{aligned}$$

The crucial step in this derivation is, of course, the inequality, which is based on the two lower bounds of Equations (17) and (18). Together they establish Equation (8).

Appendix 2. Limiting behaviour of the probabilities

The posterior probability for H^0 conditional on the jury vote $V_{n\Delta}$ can be written as a regularized, incomplete Beta function (see Abramowitz and Stegun, 1964: 263, formulae 6.6.1, 6.6.2). Specifically,

$$\begin{aligned} p(H^0 | V_{n\Delta}) &= \frac{(n+1)!}{n_0!n_1!} \int_0^{\frac{1}{2}} dr r^{n_1} (1 - r)^{n_0} \\ &= \frac{B_{\frac{1}{2}}(n_1 + 1, n_0 + 1)}{B(n_1 + 1, n_0 + 1)} \\ &= I_{\frac{1}{2}}(n_1 + 1, n_0 + 1). \end{aligned} \quad (19)$$

In this appendix, we exploit certain asymptotic properties of the regularized, incomplete Beta function to show that $p(H^0 | V_{n\Delta})$ tends to one-half as n tends to infinity at constant Δ , but to zero if the fractional majority, $f = \Delta/n$, is held constant in the limit. It is also shown that if Δ increases more quickly than $\sqrt{n}$, $p(H^0 | V_{n\Delta})$ still tends to zero in the limit of n to infinity.

Theorem 1

If $\Delta = n_1 - n_0 \geq 0$ is constant, then $I_{\frac{1}{2}}(n_1 + 1; n_0 + 1)$ tends to $\frac{1}{2}$ in the limit that $n = n_1 + n_0$ tends to infinity.

Proof. On changing the integration variable from r to $t = (1 - 2r)^2$, we find

$$I_{\frac{1}{2}}(n_1 + 1; n_0 + 1) = 2^{-n-2} \frac{(n+1)!}{n_0!n_1!} \int_0^1 \frac{dt}{\sqrt{t}} (1 - \sqrt{t})^{n_1} (1 + \sqrt{t})^{n_0}. \quad (20)$$

This expression can be rewritten as

$$I_{\frac{1}{2}}(n_1 + 1; n_0 + 1) = 2^{-n-2} \frac{(n+1)!}{n_0!n_1!} \int_0^1 \frac{dt}{\sqrt{t}} (1 - \sqrt{t})^{n_0} (1 + \sqrt{t})^{\Delta}. \quad (21)$$

The last factor in the integrand can be expanded as the finite binomial series

$$(1 - \sqrt{t})^{\Delta} = \sum_{m=0}^{\Delta} \frac{\Delta!}{m!(\Delta - m)!} (-1)^m t^{\frac{m}{2}},$$

and this allows the evaluation of the integral term for term:

$$I_{\frac{1}{2}}(n_1 + 1; n_0 + 1) = 2^{-n-2} \frac{(n+1)!}{n_0!n_1!} \sum_{m=0}^{\Delta} \frac{\Delta!}{m!(\Delta - m)!} (-1)^m \frac{\Gamma(\frac{m+1}{2})\Gamma(n_0 + 1)}{\Gamma(n_0 + \frac{m+3}{2})}. \quad (22)$$

The Stirling expansion, namely

$$\Gamma(n) = (n-1)! = \sqrt{2\pi n} n^{n-\frac{1}{2}} e^{-n} \left[1 + O\left(\frac{1}{n}\right) \right]$$

is now used to give the asymptotic expressions

$$\begin{aligned} \frac{(n+1)!}{n_0!n_1!} &\sim 2^{n+1} \sqrt{\frac{n}{2\pi}} \\ \frac{\Gamma(n_0 + 1)}{\Gamma(n_0 + \frac{m+3}{2})} &\sim n_0^{-\frac{m+1}{2}} = \left(\frac{2}{n - \Delta} \right)^{\frac{m+1}{2}}. \end{aligned}$$

On inserting these forms into Equation (22), we find that $I_{\frac{1}{2}}(n_1 + 1; n_0 + 1)$ is asymptotically equivalent to

$$\frac{1}{2\sqrt{\pi}} \sqrt{\frac{n}{n - \Delta}} \left[\left[\Gamma\left(\frac{1}{2}\right) \right] - \Delta \Gamma(1) \sqrt{\frac{2}{n - \Delta}} + \cdots + (-1)^{\Delta} \Gamma\left(\frac{\Delta + 1}{2}\right) \left(\frac{2}{n - \Delta}\right)^{\frac{\Delta}{2}} \right]$$

All the terms in the square brackets, except for the first one, vanish in the limit of large n , and there is only a finite number of these terms. So only the first term survives, and since $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, we have thereby proved that $I_{\frac{1}{2}}(n_1 + 1; n_0 + 1)$ tends to $\frac{1}{2}$ in the limit as n tends to infinity.

Theorem 2

If $\Delta \sim n^{\beta}$ for large n , with $\frac{1}{2} < \beta \leq 1$, then $I_{\frac{1}{2}}(n_1 + 1; n_0 + 1)$ tends to zero in the limit.

Proof. Recall that we can rewrite Equation (19) as Equation (20). The latter can also be rewritten as

$$I_{\frac{1}{2}}(n_1 + 1; n_0 + 1) = 2^{-n-2} \frac{(n+1)!}{n_0!n_1!} \int_0^1 \frac{dt}{\sqrt{t}} (1-t)^{n_1} (1+\sqrt{t})^{-\Delta}, \quad (23)$$

and we now split the integral into two pieces, corresponding to $0 < t < \varepsilon^2$ and $\varepsilon^2 < t < 1$, where ε will be specified in a moment. Clearly,

$$\int_0^{\varepsilon^2} \frac{dt}{\sqrt{t}} (1-t)^{n_1} (1+\sqrt{t})^{-\Delta} < \int_0^{\varepsilon^2} \frac{dt}{\sqrt{t}} = 2\varepsilon$$

whereas

$$\begin{aligned} \int_{\varepsilon^2}^1 \frac{dt}{\sqrt{t}} (1-t)^{n_1} (1+\sqrt{t})^{-\Delta} &< (1+\varepsilon)^{-\Delta} \int_0^1 \frac{dt}{\sqrt{t}} (1-t)^{n_1} \\ &= (1+\varepsilon)^{-\Delta} \frac{\Gamma(\frac{1}{2})\Gamma(n_1+1)}{\Gamma(n_1+\frac{3}{2})}. \end{aligned}$$

We insert the last two inequalities into Equation (23) and also employ the Stirling expansion, as in the proof of Theorem 1. This yields

$$I_{\frac{1}{2}}(n_1 + 1; n_0 + 1) < \sqrt{\frac{n}{2\pi}} \varepsilon + \sqrt{\frac{n}{n+\Delta}} (1+\varepsilon)^{-\Delta}.$$

Now choose $\varepsilon = n^{-\alpha}$ and put $\Delta \sim n^{\beta}$, thereby obtaining

$$I_{\frac{1}{2}}(n_1 + 1; n_0 + 1) < \frac{1}{\sqrt{2\pi}} n^{\frac{1}{2}-\alpha} + (1+n^{-\alpha})^{-n^{\beta}}.$$

Now $(1+n^{-\alpha})^{n^{\beta}}$ tends to e (Euler's number) in the limit of large n , so we obtain

$$I_{\frac{1}{2}}(n_1 + 1; n_0 + 1) < \frac{1}{\sqrt{2\pi}} n^{\frac{1}{2}-\alpha} + \exp[-n^{\beta-\alpha}].$$

asymptotically. For any $\alpha > \frac{1}{2}$, the first term above vanishes asymptotically, and for any $\beta > \alpha$, so does the second term. Hence, for any $\frac{1}{2} < \beta \leq 1$, we have shown that $I_{\frac{1}{2}}(n_1 + 1; n_0 + 1)$ tends to zero in the limit as n tends to infinity.

Corollary

If $f = \Delta/n$ is constant, then $I_{\frac{1}{2}}(n_1 + 1; n_0 + 1)$ tends to zero in the limit as $n = n_1 + n_0$ tends to infinity.

Proof. This follows immediately by taking the special case $\beta = 1$ in Theorem 2.

Notes

We would like to thank audiences at the BSPS 2008 conference in St Andrews and the FMP 2008 conference in Tilburg for helpful discussion. This research was carried out as part of a project funded by the Dutch Organization of Scientific Research (NWO VENI-grant no. 275-20-013). We also thank the Spanish Ministry of Science and Innovation (research project FFI2008-1169) for generous support.

1. The exact condition is obtained by solving f_1 for $\log L > 0$. This yields the solution that $f_1 > \frac{1}{1+x}$, in which $x = \log \frac{q_1}{1-q_0} / \log \frac{q_0}{1-q_1}$. We are not aware of any earlier formulation of this result, but since it is fairly straightforward we will not provide a more detailed proof of the condition.
2. Note that for fixed competences, the probability that Δ remains equal to a finite constant as the jury size n increases becomes smaller and smaller and tends to zero in the limit. So, even though the Condorcet formula is consistent with the Condorcet Jury Theorem, its domain of application becomes smaller and smaller and shrinks to nothing.
3. In the following, we leave out the value $r = \frac{1}{2}$. Since this is a possibility with measure zero, its omission has no consequences.

References

- Abramowitz M and Stegun IA (1964) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover.
- Berend D and Paroush J (1998) When is Condorcet's Jury Theorem valid? *Social Choice and Welfare* 15: 481–8.
- Bovens L and Hartmann S (2003) *Bayesian Epistemology*. Oxford: Oxford University Press.
- Dietrich F (2008) The premises of Condorcet's Jury Theorem are not simultaneously justified. *Episteme—a Journal of Social Epistemology* 5(1): 56–73.
- Goodin RE and Estlund D (2004) The persuasiveness of democratic majorities. *Politics, Philosophy and Economics* 3(2): 131–42.
- Haenni R and Hartmann S (2006) Modeling partially reliable information sources: A general approach based on Dempster-Shafer theory. *Information Fusion* 7: 361–79.
- Ladha KK (1992) The Condorcet Jury Theorem, free speech, and correlated votes. *American Journal of Political Science* 36: 617–34.
- Ladha KK (1993) Condorcet's Jury Theorem in light of De Finetti's Theorem. *Social Choice and Welfare* 10: 69–85.
- List C (2004) On the significance of the absolute margin. *British Journal for the Philosophy of Science* 55: 521–44.
- List C (2008) The epistemology of special majority voting: Why the proportion is special only in special conditions. *Social Choice and Welfare* (forthcoming).
- List C and Goodin R (2001) Epistemic democracy: Generalising the Condorcet Jury Theorem. *Journal of Political Philosophy* 9: 277–306.

About the authors

Jan-Willem Romeijn is an Assistant Professor in the Faculty of Philosophy at the University of Groningen, where he teaches the philosophy of science. His research interests include statistical

inference, formal and social epistemology, and interpretations of probability. Jan-Willem Romeijn, Oude Boteringestraat 52, 9712 GL, Groningen, The Netherlands [email: j.w.romeijn@rug.nl]

David Atkinson is Emeritus Professor of Theoretical Physics. Since his retirement he has been active in philosophy, publishing articles on thought experiments, the interpretation of quantum mechanics, infinitely many colliding balls, and probabilistic regresses in epistemology. David Atkinson, Pelsterstraat 12, 9711 KL, Groningen, The Netherlands [email: d.atkinson@rug.nl]