B.J.Pol.S. **42,** 555–571 *Copyright © Cambridge University Press, 2011*
doi:10.1017/S000712341100041X
First published online 12 December 2011

Courts of Many Minds

KAI SPIEKERMANN AND ROBERT E. GOODIN*

In *A Constitution of Many Minds* Cass Sunstein argues that the three major approaches to constitutional interpretation – Traditionalism, Populism and Cosmopolitanism – all rely on some variation of a 'many-minds' argument. Here we assess each of these claims through the lens of the Condorcet Jury Theorem. In regard to the first two approaches we explore the implications of sequential influence among courts (past and foreign, respectively). In regard to the Populist approach, we consider the influence of opinion leaders.

Cass Sunstein explores three alternative approaches to constitutional interpretation in his important recent book, *A Constitution of Many Minds*:

1. Traditionalists insist that if members of a society have long accepted a certain practice, courts should be reluctant to disturb that practice.
2. Populists believe that if most people accept a certain fact or value, judges should show a degree of humility – and respect their view in the face of reasonable doubt.
3. Cosmopolitans believe that if many nations, or many democratic nations, reject a practice, or accept a practice, the US Supreme Court should pay respectful attention.[1]

All three approaches rest, as Sunstein sees it, on a common premise – the 'many-minds' argument of his book's title. The thought is that, 'if many people think something, their view is entitled to consideration and respect'[2] – not just as a matter of courtesy, but because the more of them there are the more likely they are to be right. 'The structure of the central argument is identical in all three contexts,' in that all three rest on the same formal foundations: the Condorcet Jury Theorem (CJT).[3] In so far as the CJT is the mechanism underlying all three approaches to constitutional interpretation, that same theorem should provide a formal basis for adjudicating among them. In his subsequent discussion of those issues, Sunstein himself abjures formalism in favour of more context-sensitive lawyer-style discussions. The aim of this article is to provide more formal assessments of the epistemic power of those approaches, through suitable elaborations and extensions of the CJT apparatus and related models. At the end of his book Sunstein asks, 'Is it possible to compare and to rank the three kinds of many minds argument?'[4] His conclusion is: 'For the United States, I have suggested that traditions are likely to

* London School of Economics (email: k.spiekermann@lse.ac.uk); Australian National University and University of Essex, respectively. This article was presented in 2010 at the Conference on Collective Knowledge and Epistemic Trust (Greifswald), the Annual Meeting of APSA, Washington, D.C., and the Faculty Workshop of Harvard Law School. The authors would like to thank those audiences for many insightful comments. They are grateful to three of the *Journal*'s anonymous referees who provided instructive comments. Franz Dietrich kindly helped with some notational issues.

[1] Cass R. Sunstein, *A Constitution of Many Minds: Why the Founding Document Doesn't Mean What It Meant Before* (Princeton, N.J: Princeton University Press, 2009), pp. ix–x.
[2] Sunstein, *A Constitution of Many Minds*, pp. ix–x.
[3] Sunstein, *A Constitution of Many Mind*, pp. x, 8–10.
[4] Sunstein, *A Constitution of Many Minds*, p. 214.

https://doi.org/10.1017/S000712341100041X Published online by Cambridge University Press

provide the strongest basis for constitutional law, and that international practices provide the weakest. Public convictions are in the middle.' A brief reminder of the basic structure of the CJT apparatus will be provided in the next section. Following that, we discuss informational cascades, a problem likely to plague Traditionalism, Sunstein's preferred approach.[5] In the next two sections, we analyse Populism and the likely fewer problems that arise when individuals are influenced by opinion leaders.

THE CONDORCET JURY THEOREM AND ITS DEMANDING ASSUMPTIONS

Sunstein's 'many-minds argument' relies on the claim that many minds are more likely to be right than one (or just a few) mind. The most famous technical result to show the potential of the many-minds argument is the CJT. The original version of the CJT applies when a group of jurors decides between two alternatives with majority decision. The following assumptions are made:

- *Competence*. Each juror votes for the correct alternative with probability $p > 0.5$.[6]
- *Independence*. The votes of the jurors are statistically independent, given the true state of the world regarding the correct alternative.

THE CONDORCET JURY THEOREM

If Competence and Independence are met, the following hold:

Asymptotic claim. The probability of a majority of jurors being correct converges to 1 as the number of jurors tends to infinity.
Non-Asymptotic claim. The majority of a larger group of jurors is more likely to be correct than the majority of a smaller group of jurors (provided the number of jurors is odd).

The CJT falls into two parts, the asymptotic and non-asymptotic claim. Various generalizations and refinements of the CJT have been proposed, and there has been an extensive discussion on the plausibility of the competence and the independence assumptions. For instance, Bernard Grofman *et al*. report that the CJT still holds with heterogeneous individual competence as long as the average competence is fixed and the distribution of competences is symmetric; Franz Dietrich proves that the asymptotic part holds as long as the average competence is above 0.5.[7] Another extension is offered by Christian List and Robert Goodin, who show that a version of the CJT can be applied to settings with more than two alternatives.[8]

Most relevant for our debate are studies as to how much the independence assumption can be relaxed while still maintaining the asymptotic or non-asymptotic part of the CJT. Several formal approaches to relaxing the independence assumption are conceivable.[9]

[5] Everything said there could also be applied with minor adjustments to the Cosmopolitan approach, which we do not here discuss separately in consequence.

[6] This subsumes the assumption of sincerity, which is often stated separately.

[7] Bernard Grofman, Guillermo Owen and Scott L. Feld, 'Thirteen Theorems in Search of the Truth', *Theory and Decision*, 15 (1983), 261–78; Franz Dietrich, 'The Premises of Condorcet's Jury Theorem Are Not Simultaneously Justified', *Episteme*, 58 (2008), 56–73.

[8] Christian List and Robert E. Goodin, 'Epistemic Democracy: Generalizing the Condorcet Jury Theorem', *Journal of Political Philosophy*, 9 (2001), 277–306.

[9] The influence of opinion leaders was first studied formally by Philip J. Boland, Frank Proschan and Y. L. Tong, 'Modelling Dependence in Simple and Indirect Majority Systems', *Journal of Applied Probability*,

https://doi.org/10.1017/S000712341100041X Published online by Cambridge University Press

Apart from the question of whether Competence and Independence hold in reality and what follows theoretically if they are relaxed, there is also the more fundamental issue as to whether we can simultaneously know both assumptions to be justified. Dietrich has recently shown that this is impossible:[10] if one knows that the votes are independent, one cannot know the competence of the jurors, and if one knows the jurors are competent one cannot justify the assumptions that their votes are independent.[11] Dietrich proposes to develop new jury theorems with more realistic premises.[12] In this article, we do not attempt to offer a new jury theorem, but we show various ways in which Independence can be undermined in practice and how results based on dependent votes differ from the (often rather fantastic) results of the classical CJT. This suggests that Sunstein's casual reliance on the CJT causes more problems for him than he may have imagined and that his own preference for Traditionalism is in need of some important qualifications. Populism may be a more attractive alternative, but it also faces potentially serious problems.

TRADITIONALISM AND CASCADES

The original CJT set-up assumes that voters are making decisions either simultaneously with one another, or in ignorance of or in indifference to what other voters have done. The Traditionalist approach to constitutional interpretation envisages something very different. There, courts are making their decisions sequentially, not simultaneously. Furthermore, if they are Traditionalists, subsequent courts make their decisions not only in knowledge of but also in deference to earlier courts' decisions. That changes things dramatically.

In our model of Traditionalism we assume that different judges have to decide on the same dichotomous question at different points in time. Each judge has an independent and symmetric private signal of equal reliability as to which of the two alternatives is the correct one. We also assume that this signal is more likely to point to the correct alternative than the incorrect alternative, analogous to the competence assumption in the CJT. Suppose that judges care only about getting their decision right in the current case.[13]

(F'note continued)

26 (1989), 81–8. Krishna Ladha has worked extensively on the effect of correlated votes, where correlation is due to shared information, particularly in Krishna K. Ladha, 'The Condorcet Jury Theorem, Free Speech, and Correlated Votes', *American Journal of Political Science*, 36 (1992), 617–34; Krishna K. Ladha, 'Condorcet's Jury Theorem in Light of De Finetti's Theorem', *Social Choice and Welfare*, 10 (1993), 69–85; Krishna K. Ladha, 'Information Pooling through Majority-Rule Voting: Condorcet's Jury Theorem with Correlated Votes', *Journal of Economic Behavior & Organization*, 26 (1995), 353–72. This idea was extended and developed into a new jury theorem by Franz Dietrich and Christian List, 'A Model of Jury Decisions where All Jurors Have the Same Evidence', *Synthese*, 142 (2004), 175–202, assuming that all jurors have the same evidence. Serguei Kaniovski, 'Aggregation of Correlated Votes and Condorcet's Jury Theorem', *Theory and Decision*, 69 (2010), 453–68, has recently shown that a (pairwise) correlation table is not sufficient to determine the existing dependence relations between jurors and that different ways to instantiate the same pairwise correlations lead to different results.

[10] Dietrich, 'The Premises of Condorcet's Jury Theorem Are Not Simultaneously Justified'.

[11] With the exception of highly construed examples where it is known that both assumptions hold in virtue of the construction.

[12] A general account of the causal bases of dependence leading to a new jury theorem is presented in Franz Dietrich and Kai Spiekermann, 'Epistemic Democracy with Defensible Premises' (unpublished manuscript, 2010, available at http://personal.lse.ac.uk/spiekerk/papers/DietrichSpiekermann-EpistemicDemocracy.pdf).

[13] Among other things, they are not concerned with how their current decision might influence future judgements.

https://doi.org/10.1017/S000712341100041X Published online by Cambridge University Press

Then each judge has to form his or her decision by considering her own private signal and the history of votes. The history of votes is mutual knowledge.

It is well known that such set-ups can cause the problem of informational cascades. There can be histories of votes that constitute such strong evidence in favour of one alternative that all judges will always follow the evidence of the history and never vote according to their own private signal. If that happens, an informational cascade has begun. Judges will have stopped learning from their own signals and will blindly follow the judgement the historic voting record suggests. Informational cascades are a problem because the informational base on which all future judgements are grounded can be very thin.[14]

The literature on informational cascades has grown rapidly and has by now resulted in all sorts of technical refinements that we will not address in this article. The seminal contribution is provided by Sushil Bikhchandani, David Hirshleifer and Ivo Welch (henceforth: BHW).[15] A very simple set-up, roughly in line with the model introduced by BHW, suffices to clarify the problem with the many-minds argument based on sequential judgements.[16] In our treatment, unlike BHW and others, we will not always model judges as fully Bayesian rational. In particular, we want to maintain the possibility that judges can be irrationally overconfident about their own private signal, or that they vote according to their private signal as a matter of principle.[17]

This presumption of 'bounded rationality' makes room for assumptions that may ultimately be more realistic, both in general and particularly for the case at hand, than full Bayesian rationality. We know from experimental psychology that, while people generally update their beliefs in the direction indicated by Bayes's formula, they do so far more slowly than Bayesian rationality requires. 'A convenient first approximation to the data would say that it takes anywhere from two to five observations to do [what by Bayes's formula should be] one observation's worth of work in inducing a subject to change his opinions,' one researcher reports.[18] With regard to judges, evidence suggests that they are much more likely to hold on to their own view than to defer to majorities. One landmark

[14] This is what Adrian Vermeule, *Law and the Limits of Reason* (Oxford: Oxford University Press, 2009), pp. 5–6, see also pp. 75–7, calls the 'Burkean Paradox': 'Where actors defer to the information of past others, as the Burkean position would have them do, the result is a low-value ''information cascade'' rather than collective wisdom.'

[15] Sushil Bikhchandani, David Hirshleifer and Ivo Welch, 'A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades', *Journal of Political Economy*, 100 (1992), 992–1026. Similarly, Abhijit V. Banerjee, 'A Simple Model of Herd Behavior', *Quarterly Journal of Economics*, 107 (1992), 797–817; and Lones Smith and Peter Sorensen, 'Pathological Outcomes of Observational Learning', *Econometrica*, 68 (2000), 371–98; a useful introduction is Christophe P. Chamley, *Rational Herds: Economic Models of Social Learning* (Cambridge: Cambridge University Press, 2004).

[16] Note the contrast between the literature on cascades and on sequential voting games, as analysed in a seminal paper by E. Dekel and M. Piccione, 'Sequential Voting Procedures in Symmetric Binary Elections', *Journal of Political Economy*, 108 (2000), 34–55. The cascade literature suggests that the utility of voters is derived purely from their own court's decision in the case at hand, while voting games relate individual utility to the overall outcome of all actions by all agents.

[17] Of course, this could be done in a Bayesian framework by changing the utility functions of the judges. But the Bayesian treatment comes with some algebraic costs and little gain for the purposes of this article.

[18] Ward Edwards, 'Conservatism in Human Information Processing', in Benjamin Kleinmuntz, ed., *Formal Representation of Human Judgement* (New York: Wiley, 1968); reprinted in Daniel Kahneman, Paul Slovic and Amos Tversky, eds, *Judgement under Uncertainty: Heuristics and Biases* (Cambridge: Cambridge University Press, 1982), pp. 359–69.

https://doi.org/10.1017/S000712341100041X Published online by Cambridge University Press

study of the US Supreme Court reports that justices vote in line with their own preferences (operationalized as 'the same way they voted on the last such cases') rather than with precedent (operationalized as 'the way the majority voted in the past case') in an astonishing 90.8 per cent of cases.[19]

Modelling Court Decisions and Cascades

At each point in time t ($t = 0, 1, 2, \ldots$), there is a court of k judges. The current judges decide simultaneously between two alternatives. The judges on the court change each time period, so that no judge votes on an issue twice. A judge's decision is labelled by time in the superscript and by the label of the judge at that time ($1, \ldots, k$), which appears as a subscript. Thus, the decision of judge i at time t is v_i^t. There is a state of the world such that either $\theta = 1$ or $\theta = 0$ is factually true, and this state does not change over time. Each judge i at time t receives a private signal $S_i^t \in \{0, 1\}$ about the state of the world. For simplicity, we assume that these signals are of equal quality for all judges, that they are better than random and that the two alternatives are treated symmetrically, such that:

$$Pr(S_i^t = 1|\theta = 1) = Pr(S_i^t = 0|\theta = 0) = p > 0.5,$$

for all $t = 1, 2, \ldots$ and $i = 1, 2, \ldots, k$. For each judge, the decision as to which alternative he votes for is based on his own private signal, and the history of previous judgements, h_{t-1}. In other words, a judge's judgement function maps the history h_{t-1} of all previous judgements and the judge's private signal S_i^t onto a decision to vote for 0 or 1:

$$v_i^t : (h_{t-1}, S_i^t) \rightarrow \{0, 1\}$$

Decision functions can take different shapes. For illustration, consider the classic starting example presented by BHW.[20] There, courts are of size $k = 1$. The first judge will vote in line with her private signal. All subsequent judges know the complete history and their own private signal. They calculate their degree of belief by Bayesian updating. As soon as the absolute margin of votes is 2 or greater, each judge will vote for the opinion with more support regardless of their own signal because the support from two or more other judges outweighs the (potentially contrary) evidence of one's own private signal. Therefore, a cascade starts to run very quickly, and the cascade will often settle for the wrong alternative. The implicit decision function in the basic BHW model is: take all historic votes and add your own private signal as another vote; then back the majority winner.[21]

[19] Jeffrey A. Segal and Harold J. Spaeth, 'The Influence of Stare Decisis on the Votes of United States Supreme Court Justices', *American Journal of Political Science*, 40 (1996), 971–1003.

[20] In the BHW model, one individual decides at each point in time (i.e. $k = 1$). The state of the world is fixed randomly at the start with equal probability, and all judges begin with the same prior of $\pi(\theta = 1) = \pi(\theta = 0) = 0.5$. The utility function of judges is defined such that a correct decision yields payoff 1, an incorrect decision 0:

$$u(v, \theta) = \begin{cases} 1 & \text{if } v = \theta \\ 0 & \text{otherwise.} \end{cases}$$

The judges maximize their utility, and here this means they vote for the alternative that they consider more likely to be correct.

[21] Different modelling choices are conceivable to break ties, especially randomizing or assuming that judges have marginally more confidence in their own signal than in other judgements.

https://doi.org/10.1017/S000712341100041X Published online by Cambridge University Press

This decision rule is Bayesian rational under the BHW set-up. It is also quite intuitive: since all judges are equally competent, the alternative with more private signals in favour of it is more likely to be correct and, therefore, judges should vote for it. But there are other decision functions that are likely to occur in reality. For example, a judge could always vote in line with his own signal. Or a judge might succumb to the historic majority only if it reaches a certain higher threshold. Or judges could have a limited memory or a limited willingness to incorporate decisions that lie far in the past. One could model this within a Bayesian framework by assuming that such judges have mistaken beliefs about the distribution of competence in a society, or one could assume that these judges have utility functions that represent not only an interest in getting the decision right but also in voting according to some procedural conditions. But a simpler and often more plausible way to model these different decision functions is to assume that these judges simply follow heuristics that are not Bayesian rational. Judges may partly be driven by normative considerations or by the desire for expressive voting. Thus, we will work with simple, not necessarily Bayesian rational, decision functions in the remainder of this article.

What happens if we allow for multi-member courts with $k > 1$? As is well known from the game-theoretic literature on votes in multi-member panels and courts, this question in principle allows for subtle and sophisticated strategic analyses.[22] Again, we propose to simplify matters. We assume that judges care only about getting their own vote in the case at hand right, and take other observed votes to be informative, i. e. in line with private signals. A more careful analysis of the strategic subtleties has to wait for another day. Yet again, in practice, given that judges have limited insight into the preferences and beliefs of their predecessors and peers, this assumption of non-strategic behaviour may not be unreasonable, as sophisticated strategic considerations would require a much richer knowledge of the decision environment than we would expect in real-world settings.

With these preliminaries out of the way, we can state how we model the judges' decisions (pseudo code to illustrate the way the model has been programmed is provided in the supplementary material §1). We consider a very simple class of decision functions for judges, which can be described by just two parameters: the length of their memory, m, and the weight they give to their own private signal, w.[23] The length of memory determines how many previous decisions are taken into account. If $m = 0$, the judge only takes his own signal into account. The weight of their own signal is a natural number that specifies, in effect, how many votes the judge allocates to himself when considering his private signal and all previous decisions he considers. For instance, in a one-member court, if w is large ($>m$) the judge will vote on the basis of his or her own private signal alone. If $w = 1$ and $m = 2$, then the judge will vote against his or her own private signal if and only if both previous judges have voted against it. If $w = 2$ and $m = 3$, then the judge

[22] David Austen-Smith and Jeffrey S. Banks, 'Information Aggregation, Rationality, and the Condorcet Jury Theorem', *American Political Science Review*, 90 (1996), 34–45; Timothy Feddersen and Wolfgang Pesendorfer, 'Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting', *American Political Science Review*, 92 (1998), 23–35.

[23] We have already reported psychological evidence that empirically w may well be larger than 1. For similar evidence that m may be limited, see, for example, Yaacov Trope, 'Inferences of Personal Characteristics on the Basis of Information Retrieved from One's Memory', *Journal of Personality & Social Psychology*, 36 (1978), 93–106; reprinted in Kahneman, Slovic and Tversky, *Judgement under Uncertainty: Heuristics and Biases*.

https://doi.org/10.1017/S000712341100041X Published online by Cambridge University Press

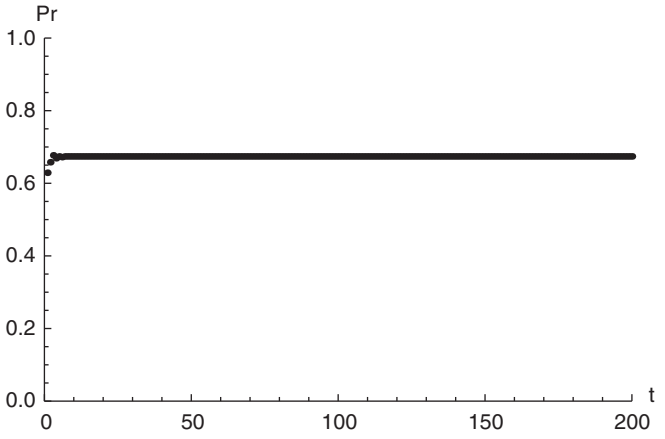


Fig. 1. Average group competence dependent on time in 1,000 simulations with k = 9, and judges with m = 20 and w = 1

will vote against his or her own private signal if and only if all three previous judges have voted against it.[24] More generally, judges decide as follows:

1. Count all previous votes for the two alternatives that are in the reach of memory, m.
2. Add w votes to the votes for the alternative indicated by private signal, S_i^t.
3. Vote for the alternative with more votes. If tied, vote according to own signal S_i^t.[25]

Homogeneous Judges

We have run computer simulations to determine the epistemic competence of courts voting in sequence. The courts decide by majority vote, and (unless stated otherwise) we assume courts consist of nine judges ($k = 9$). We usually consider 200 time periods. Since the private signals judges receive are stochastic, one needs to determine the average group competence across many runs. Our results are based on the averaging of 1,000 runs each. We assume a reliability of individual signals of $p = 0.55$.

In the first scenario, all judges have maximal memory and assign a weight of 1 to their own private signal. Figure 1 shows the average competence the courts have at each point in time. The first court (without any history available) has a competence in line with the Condorcet Jury Theorem with $p = 0.55$ and $n = 9$, which leads to a group competence of 0.621. The group competence rises for the subsequent courts because these courts can draw on previous decisions. However, cascades arise very quickly because all later judges will vote with the historic majority as soon as an absolute margin of 2 or more arises. After a certain point, an informational cascade of the same sort that Sunstein worries

[24] In the case of multi-member courts, these calculations would have to be expressed as margins of all previous votes. See also supplementary material §1.

[25] This is one way of formalizing Vermeule's suggestion that 'individual judges might adopt an intermediate approach, according to which they give some but not complete deference to the views of the past, and correlatively think for themselves to some degree or in some circumstances.' See Vermeule, *Law and the Limits of Reason*, p. 76.

https://doi.org/10.1017/S000712341100041X Published online by Cambridge University Press

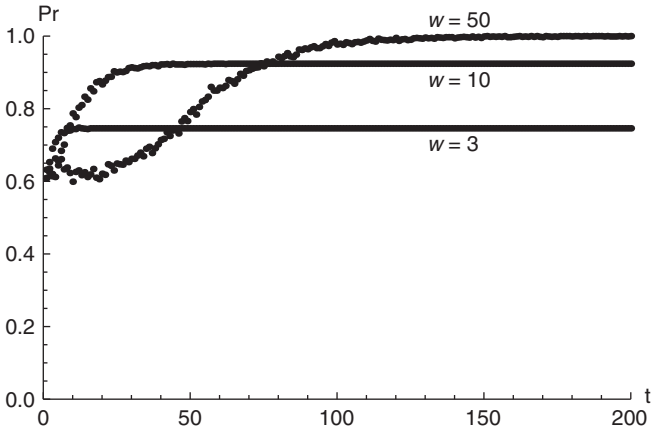


Fig. 2. Average group competence dependent on time in 1,000 simulations with k = 9, and judges with m = 200 and different levels of w

about with the Populist approach sets in.[26] Then, instead of 'many-minds' we merely have 'many-mimics', which confer no epistemic advantage (Figure 1).

The emergence of cascades is delayed if the judges assign a higher weight to their own signal. Figure 2 shows group competence for weights 3, 10 and 50, while all other parameters remain unchanged. Two effects can be observed. On the one hand, increasing the weight increases the group competence to which the courts converge after sufficient time. On the other, a higher weight delays the convergence and indeed the improvement in group competence (both effects being greater the greater the weight). Both effects are due to the fact that cascades arise later, and decisions are therefore based on more private signals, which improves the epistemic performance for the later votes. Judges who are very self-confident about the quality of their own signal will persist in revealing their own signals and avoid cascades for longer. This increases the epistemic performance of later courts. It is this kind of situation in which Sunstein's many-minds argument gets more traction. Ironically, these settings require that the judges are 'stubborn' about their own votes and slow to follow the many-minds of others.

These results suggest that the many-minds argument has to be treated with care when the voting is sequential, which is the case for both Traditionalism and (to a certain extent) for Cosmopolitanism. If judges are quite responsive to the opinions of their predecessors, they can quickly trigger cascades, compromising the capacity for many minds to enhance group competence. If they are more likely to reveal their own private signal in their decision because they are less responsive to previous judgements, they are not using the previous decisions to improve their own vote, but later judgements can benefit because early cascades are prevented and the information from more independent assessors is taken into account.

The upshot of this analysis is clear. The only way in which courts composed of homogeneous judges will be able to achieve any substantial epistemic advantage over courts that pay too much attention to their predecessors is by judges attaching very little

[26] Sunstein, *A Constitution of Many Minds*, pp. 103–6, 171–3.

https://doi.org/10.1017/S000712341100041X Published online by Cambridge University Press

importance to the judgements of previous judges, relative to their own. That is to say, judges would have stubbornly to stick with their own views in the face of a very substantial body of traditional evidence in the opposite direction. In short, in this scenario, achieving the epistemic power of the many-minds would require judges largely to resist tradition rather than bowing to it.

Judges Who Distinguish between Informative and Cascade Votes

So far we have assumed that judges can observe only the votes of their predecessors, but not the reasons why their predecessors voted as they did. This led to a dilemma. A judge has only one vote with which to perform two different tasks: on the one hand, reveal his private signal and, on the other, aggregate the votes that have taken place previously.

The situation could be improved if these two tasks were separated through a division of labour. Using this approach, some judges reveal their private signals, while others vote to aggregate the signals that have been revealed so far. However, for this division of labour to work it is necessary that the aggregators distinguish between informative votes and aggregated (cascade) votes. Judges serving as aggregators must know whether any given predecessor voted the way she did because she was following her own signal, or because she was following the majority of past decisions. It is not unrealistic to assume judges would know this, however: since judges can read the opinions of previous judges, they can determine whether their predecessors have voted with the traditional majority or whether they have voted in line with their own independent reasoning. A neat distinction may be difficult in practice, but judges should at least have some indication of who the independent voices among their predecessors were.

For this set-up, we hypothesize a heterogeneous court. Some judges on that court always reveal their private signals. Other judges on that court (indeed, we hypothesize, a majority of judges on the court) make their votes as in our first set-up by pooling their private signal with the votes of judges on previous panels.[27] But – and this is the crucial difference between this set-up and the last – we assume that, in so pooling, current judges take account of only the votes of previous judges who voted on the basis of their own private signals. That is to say, these 'discriminating' judges vote on the basis of previous judges' votes only when those votes are truly 'informative' and not merely the product of an informational cascade.

Our simulation is based on courts consisting of four informative judges who always vote purely in line with their own signal, and five judges who give a weight of 1 to their own signal and consider previous votes, but only those of the informative judges. Figure 3 reveals that these heterogeneous courts perform remarkably well after some time. The set-up with heterogeneous judges who can discriminate between informative and uninformative votes improves the epistemic performance of the group over time because the judges practise the described division of labour. The informative judges provide evidence, the other judges aggregate that evidence.

As revealed by Figure 3, the probability that such a court will reach the correct decision does not plateau. Instead, it continues to increase the more previous courts there have been that have to be taken into account. The group competence approaches 1 rather slowly. After taking into account fifty previous courts, the probability of the majority of the current court reaching the correct decision is only around $\Pr \approx 0.90$, and after 100 is

[27] Vermeule anticipates this part of our model, but not the next, when writing, 'Perhaps some judges in the stream of precedent or tradition have contributed independently, while some have not'. See Vermeule, *Law and the Limits of Reason*, p. 76.

https://doi.org/10.1017/S000712341100041X Published online by Cambridge University Press

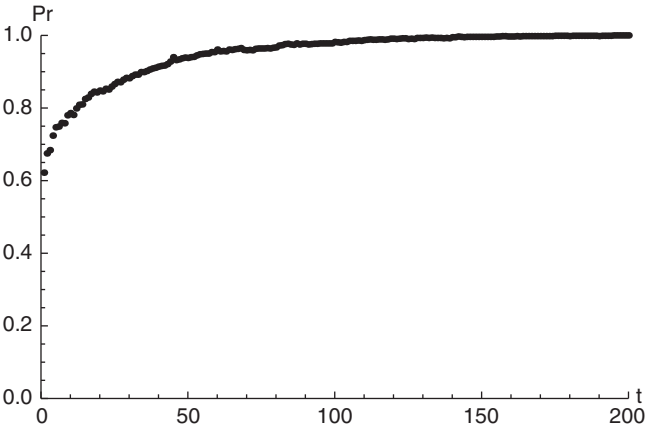


Fig. 3. Average group competence dependent on time in 1,000 simulations with k = 9 and two types of judges: four informative voters with either m = 0 or w > 1,800, and five discriminating judges who only aggregate previous informative votes with m = 200 and w = 1

still only around Pr ≈ 0.95. By the time 200 previous courts have been taken into account, however, a correct decision is virtually certain. Furthermore, with courts that are heterogeneous in this way, the probability of a correct decision is a much more rapidly increasing function of the number of previous courts than it is with homogeneous courts with very stubborn judges (e.g. $w = 50$ in Figure 2).

Here, then, is a second way the Traditionalist argument might work. Judges on heterogeneous courts can improve their chances of reaching correct decisions by taking into account the decisions of previous judges, provided they do so in this very particular way. But note well the irony. Traditionalist courts of this sort benefit epistemically only from judges taking account of the votes of previous judges who were not themselves Traditionalists and who voted purely on the basis their own private signal rather than on the basis of the history of votes before them.

This result and the previous one can be interpreted in different ways. One can take them as starting points for an empirical analysis of the functioning of courts: To what extent do judges consider previous judgements? And do they distinguish between colleagues who vote with the tradition and those who do not? Our findings clearly indicate the importance of further research on those questions by students of judicial politics. The last result can also be taken as a starting point for a normative argument in favour of diversity in courts. We have seen that the epistemic performance of a court is poor if all judges primarily follow the judgements of their predecessors. But the epistemic performance is also quite poor if they do not consider past results at all. A mix is needed that takes past results into account without suppressing the use of independent judgements entirely. Diverse courts are likely to be better placed in that regard.

Traditionalism and Cosmopolitanism Share the Same Problem

So far we have explored the issue of informational cascades purely in relation to Traditionalist courts. Notice that the same problems arise with Cosmopolitan courts as well, however, in so far as decisions there too being typically (if not necessarily exclusively) sequential in form. Just as there is no reason to suppose that your own court is the very first

https://doi.org/10.1017/S000712341100041X Published online by Cambridge University Press

one to take a Traditionalist stance towards its interpretive task, so too there is no reason to suppose that your own court is the very first one to take a Cosmopolitan stance towards its interpretive task. Yet, if the foreign courts from which Cosmopolitan judges borrow themselves have simply borrowed from other, earlier foreign courts, then once again we could easily have a case not of 'many-minds' but merely of 'many-mimics'. *In extremis*, all the foreign courts from which you are borrowing might themselves have borrowed (either directly or at several removes) from one and the same Ur-court that set the very first precedent that then got picked up in all subsequent decisions across all the different jurisdictions.

There are, of course, various other problems in implementing a Cosmopolitan approach to constitutional interpretation. One among them, obviously, is determining which foreign jurisdictions are good comparators to your own.[28] To all those problems, we add another: both Traditionalism and Cosmopolitanism are in any case seriously compromised by the risk of informational cascades that is endemic to both, in so far as both involve sequential decision processes.

POPULISM AND OPINION LEADERS

In *A Constitution of Many Minds*, Sunstein suggests that one way to apply the many-minds argument in legal theory is to embrace Populism, which is to say, follow the judgements of the majority of the population when deciding on fundamental legal principles and values. Once again, the formal framework Sunstein relies on, without spelling out the details, is the CJT.[29] If the CJT was applicable without qualifications, the majority of the population should be almost infallible as long as each individual is at least somewhat more competent than a coin toss. However, there are good reasons to believe that the CJT is not that applicable.

In his discussion of Populism, Sunstein suggests that constitutional courts should check and potentially revise their judgements if they experience a 'public backlash'.[30] Thus, *'intense public opposition is a clue'* that the court got a decision wrong.[31] Of course, Sunstein also observes that the public can get things badly wrong: 'We have seen that if a systematic bias is present, the majority will not be right. If most people think that free trade is bad, even though it is (usually) good, governments will do badly if they follow the view of most people.'[32] Sunstein mentions the negative role of cascades, which undermine the independence of voters, and he discusses the problem of cascades induced by 'meaning entrepreneurs'.[33] Overall, Sunstein is quite sceptical of judges being able to distinguish

[28] There is an analogous problem, perhaps, in determining as a Traditionalist which past precedents are relevantly similar to the case before your current court (Vermeule, *Law and the Limits of Reason*, pp. 71–2).

[29] Sunstein, *A Constitution of Many Minds*, p. 136, chap. 7.

[30] Sunstein, *A Constitution of Many Minds*, p. 125.

[31] Sunstein, *A Constitution of Many Minds*, p. 165, Sunstein's emphasis. We are puzzled by Sunstein's proviso that Populism should only look for 'intense public opposition'. To see that this statement does not make much sense in a CJT context, it suffices to consider two examples: (1) assume competence and independence hold, but each single voter is not very competent ($p = 0.51$, for instance). In a large population, the majority is very likely to be right, but the result is also very likely to be tight (an expected 51 per cent versus 49 per cent); (2) assume the voters are heavily influenced by one opinion leader taking the incorrect stance. The incorrect stance wins by a landslide. These two examples show that the size of the majority alone should not make us more comfortable to accept an outcome. Without knowing that the independence assumption holds, the size of the majority does not tell us much.

[32] Sunstein, *A Constitution of Many Minds*, p. 169.

[33] Sunstein, *A Constitution of Many Minds*, p. 172.

https://doi.org/10.1017/S000712341100041X Published online by Cambridge University Press

between situations in which the CJT applies, and situations in which a violation of the independence or competence condition renders the CJT inapplicable.[34]

We agree with Sunstein's assessment that the independence condition is of the most severe concern.[35] In this section we explore the implications of violating the independence condition in one specific way: by introducing opinion leaders influencing the population. We find support for Sunstein's scepticism regarding Populism, in that even moderate levels of influence by one or a few opinion leaders can seriously distort results. However, in a more diverse society with multiple opinion leaders, this effect is mitigated. In this section, we start with one opinion leader and then move on to modelling the influence of multiple opinion leaders with and without correlation between them.

One Opinion Leader

Many different mechanisms to induce correlation between votes, and thereby violate the independence condition, are conceivable. We begin with a simple and politically highly relevant constellation: all voters are equally influenced by the stance of one opinion leader. This opinion leader could be a politically opinionated television broadcaster, a newspaper or an influential public figure. Such a scenario was first discussed formally by Boland *et al.*[36]

The opinion leader has competence $\hat{p}$ (we signify all variables regarding the opinion leader with a hat), which is the probability that the opinion leader adopts the correct position. The opinion leader does not vote, but influences his followers among the voters. Voters follow the opinion leader (i.e. adopt his position) with probability π. If a voter does not follow the opinion leader, she has competence p to vote for the correct alternative, similar to the standard CJT set-up. A positive probability-of-following π induces positive correlation between the position of the opinion leader and the vote of each follower,[37] and also among the votes of all the followers themselves, so that the independence assumption of the CJT no longer holds.

Figure 4 shows the effect of various levels of probability-of-following π on the group competence, i.e. the probability that a majority of voters is correct. We set $\hat{p} = 0.6$ and $p = 0.55$. The result for $\pi = 0$ shows the normal asymptotic CJT result – the group competence tends to 1 with increasing group size. With a slight opinion-leader influence of $\pi = 0.05$ the asymptotic result still holds, but the speed of the convergence is diminished. For higher levels of probability-of-following (Figure 4 displays results for $\pi = 0.1$ and 0.2) the group competence increases first as the group size increases, but then converges towards the competence of the opinion leader $\hat{p} = 0.6$.

The important upshot of this result is that even moderate levels of opinion leader influence can derail the CJT result. If the voters are $p = 0.55$ competent and at least one in

[34] Sunstein, *A Constitution of Many Minds*, p. 175.

[35] Most worries about the competence condition, as in David M. Estlund, *Democratic Authority: A Philosophical Framework* (Princeton, N.J.: Princeton University Press, 2008), pp. 225–8, are in our view best seen as really worries about independence. All that the CJT requires by way of competence, recall, is that voters be 'better than random'; and how could they be worse than random, except by some common influences that systematically affect many voters at once, thus violating the independence assumption?

[36] Boland, Proschan and Tong, 'Modelling Dependence in Simple and Indirect Majority Systems'.

[37] However, the probability-of-following is not usually identical with the correlation coefficient, as we discuss in the supplementary material §2. In this regard we develop a different technical treatment from that of Boland *et al*. In other respects our set-up is very similar to theirs.

https://doi.org/10.1017/S000712341100041X Published online by Cambridge University Press

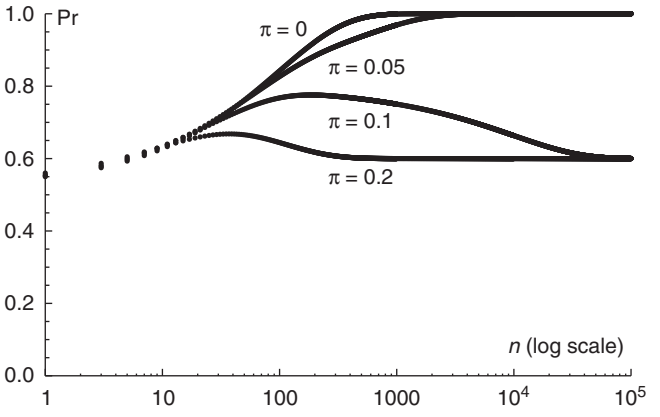


Fig. 4. The influence of one opinion leader on the group competence dependent on group size for various levels of probability-of-following, for odd n only

ten follows the opinion leader, the group competence converges towards the competence of the opinion leader, and not 1. The exact threshold where the CJT breaks can easily be determined (see supplementary material §2). The group competence converges to $\hat{p}$ rather than to 1 if:

$$\pi > \frac{p-0.5}{p},$$

and to $(\hat{p}+1)/2$ for $\pi = (p-0.5)/2$. For Figure 4 the threshold is $(0.55-0.5)/0.55 \approx 0.091$. The result also shows that the higher the voters' individual competence, the more robust are the CJT results, in the sense that a higher probability-of-following is needed to overturn them.

Two Opposed Opinion Leaders

The previous section suggests that the influence of one opinion leader, if sufficiently strong, can have deleterious consequences for the group competence. But assuming the presence of just one opinion leader is unrealistically pessimistic. A pluralistic society is more likely to have several opinion leaders, and one can hope that biases in different directions will be less bad than the uncountered influence of just one opinion leader.

We assume that each opinion leader j has a competence $\hat{p}_j$, which is the probability that she will support the correct position. Each voter is influenced by exactly one opinion leader, and each opinion leader has n_j potential followers. Within each group there is a probability, π_j, that voters within that group will adopt the view of the opinion leader.

We start by considering an extreme case: two opinion leaders with perfectly negative correlation between them. This is a model for 'hyper-partisan' politics where a society is influenced by two opinion leaders with diametrically opposed positions. Table 1 compares the results for different levels of the probability-of-following π and for two different partitions of the population.[38] In the first case we consider an almost equal split of 501 and 500 group members. In the second case, we give the first opinion leader 701 followers, and the second opinion leader 300 followers. As before, we assume the voters have

[38] Supplementary material §3 provides an analytical result for group competence in this setting.

https://doi.org/10.1017/S000712341100041X Published online by Cambridge University Press

TABLE 1 *Group Competence: Two Perfectly Negatively Correlated Opinion Leaders*

	Probability-of-following π						
Group partition	0	0.1	0.2	0.3	0.5	0.75	1
501, 500	0.999	0.998	0.995	0.990	0.967	0.886	0.600
701, 300	0.999	0.976	0.799	0.620	0.600	0.600	0.600

Note: The first opinion leader has competence $\hat{p}_1 = 0.6$, the other $\hat{p}_2 = 0.4$. If voters are not following the opinion leader they have competence 0.55.

competence $p = 0.55$ if they are not influenced by their opinion leader. The first opinion leader has competence $\hat{p}_1 = 0.6$. Since the two opinion leaders are perfectly negatively correlated, the second opinion leader has competence $\hat{p}_2 = 0.4$.

If the voters do not follow their opinion leaders at all ($\pi = 0$), the result is simply the CJT result for $p = 0.55$ and $n = 1{,}001$. If the voters always follow their respective opinion leader, the opinion leader commanding the majority (i.e. the first opinion leader) determines the vote, and the group competence boils down to his competence. However, for other positive but not perfect levels of opinion leader influence, we can make two observations:

1. For group partition (501, 500) the group competence is higher if the two opinion leaders are perfectly negatively correlated, compared to the result without correlation, particularly at higher values of π (compare the top row of Table 1 with the second row of Table 2).
2. The group competence is higher if the two perfectly negatively correlated opinion leaders have more similar-sized groups of followers (compare the top with the bottom row of Table 1).

While we do not offer a formal proof here, the mechanism behind this result is easy to grasp. Negative correlation leads to non-independent votes that 'cancel each other out', and it does so more completely where the groups are of similar size. For instance, when the partition is (501, 500) and the probability-of-following is 0.5, around 250 voters are expected to vote in one direction as followers of the first opinion leader, and since the two opinion leaders are 100 per cent negatively correlated, around 250 voters are expected to vote in the other direction as followers of the second opinion leader. These votes cancel each other out. The remaining voters (around 500) vote according to their private signal with competence 0.55. The result is, therefore, very likely to be correct, almost as likely as if the decision were taken purely by a group of 500 independent voters each having individual competence of 0.55.[39]

This cancelling effect is compromised, however, the more disparate the size of the groups. In the second row of Table 1, 701 voters are potential followers of the first opinion leader and 300 are potential followers of the second opinion leader. If the probability-of-following in each group is 0.5 once again, there will be around 350 people following the first opinion leader, and around 150 voting the opposite way, following the second opinion leader. That leaves around 500 independent voters voting according to

[39] For comparison: the CJT result for $n = 500$ and $p = 0.55$ is 0.986.

https://doi.org/10.1017/S000712341100041X Published online by Cambridge University Press

TABLE 2 *Group Competence and Multiple Opinion Leaders*

No. of opinion leaders	Group partition	Probability-of-following π						
		0*	0.1	0.2	0.3	0.5	0.75	1*
1*	1,001	0.999	0.750	0.600	0.600	0.600	0.600	0.600
2*	501, 500	0.999	0.899	0.838	0.835	0.824	0.784	0.600
3	333, 334, 334	0.999	0.95	0.84	0.70	0.65	0.65	0.648
5	1×201, 4×200	0.999	0.98	0.90	0.83	0.69	0.68	0.683
11	11×91	0.999	0.99	0.96	0.91	0.84	0.76	0.753

Note: All $\hat{p}_i = 0.6$ and $p = 0.55$. Rows and columns marked with * show analytical results; all other results are based on Monte Carlo simulations. All analytical results are rounded to three digits, all Monte Carlo estimates to two digits.

their private signal with individual competence of 0.55, once again. But in this case, it also leaves around 200 voters voting the way the first opinion leader indicates whose votes are not cancelled out by votes of people following the second opinion leader. This reduces the group competence, as these 200 uncancelled voters have a probability of 0.4 to vote incorrectly, in which case the independent voters are unlikely to overrule them.[40]

Uncorrelated Opinion Leaders

Another interesting setting with multiple opinion leaders arises if the opinion leaders are many, and they are mutually uncorrelated. Analytical results for multiple opinion leaders become increasingly unwieldy. For this reason we rely primarily on Monte Carlo numerical simulations to estimate the group competence in settings with multiple opinion leaders (see supplementary material §3).

Table 2 shows the results. We keep the total number of voters constant at $n = 1{,}001$, but change the number of opinion leaders (column 1) and the respective almost equal sizes of the groups of voters that they influence (column 2). The competence of all opinion leaders is assumed to be $\hat{p} = 0.6$, that of voters (if they do not follow the opinion leader) $p = 0.55$.

The results for just one opinion leader are as discussed above for that case. Moderate to high levels of probability-of-following will lead group competence to converge to that of the opinion leader, i.e. 0.6. It is also unsurprising that no influence of opinion leaders ($\pi = 0$) leads to the normal CJT result for 1,001 jurors, which is about 0.999. When the groups follow their respective opinion leaders with certainty ($\pi = 1$), the results can be determined with some simple combinatorial calculations, calculating the probability of a correct majority among the opinion leaders.

The principal finding in Table 2 is that, as a general rule, more opinion leaders tend to result in higher group competence.[41] To see the general tendency of increasing group competence with more opinion leaders, compare, for instance, the group competence arising from the set-up with three opinion leaders with the set-up with eleven opinion leaders. In Table 2, we see that eleven opinion leaders yield greater group competence for all positive levels of probability-of-following. This is unsurprising. What drives these

[40] How this constellation affects group competence in general depends on p, $\hat{p}$ and the group partition.
[41] The results for the setting with two opinion leaders represent an anomaly that we will discuss below.

https://doi.org/10.1017/S000712341100041X Published online by Cambridge University Press

results is the tendency for opinion leaders to pull in different directions and thus 'cancel' each other out according to the law of large numbers, and for more opinion leaders to provide more independent points of judgement.

In the most extreme case, when there are as many opinion leaders as voters, and the opinion leaders are not more competent than a coin toss ($p = 0.5$), we expect the opinion leaders to be about equally split in their support for the two alternatives. Consequently, the votes caused by the opinion leaders will also tend to be equally split. Therefore, the votes caused by the opinion leaders tend to cancel, and the votes of those whose votes were independent of any opinion leader decide the result. As long as there are enough independent votes, the group competence will be high. Obviously, when the opinion leaders are better than random they tend to cause more correct votes, improving the results further.[42] However, for smaller numbers of opinion leaders, or if opinion leaders have differential numbers of followers, the cancellation effect is less reliable.

In addition, from the discussion above we know that the competence of a group with one opinion leader converges towards the competence of the opinion leader as the group size increases if $\pi > (p-0.5)/p$. In the limit ($\pi = 1$) there are just as many independent points of judgement as there are opinion leaders. Since more independent judgements are better than fewer, more opinion leaders lead to better results.

Cases with two opinion leaders present something of an anomaly. There, group competence is persistently high even for relatively high probabilities-of-following. That is purely because the majority will typically not be wrong unless both opinion leaders are wrong, which is quite rare. By contrast, in a group with three opinion leaders the majority will typically be wrong if either two or three of them are wrong, which happens more frequently. This is why the results with two opinion leaders yield higher levels of group competence even for relatively high probabilities-of-following.

Two striking conclusions emerge from this analysis. The first conclusion is that, in general, it is epistemically better to have more opinion leaders rather than fewer if we have to have opinion leaders at all. The second conclusion is that having just two opinion leaders with an almost equal number of followers is epistemically pretty good across most of the range of possible probabilities-of-following; and where the probability-of-following is relatively high it takes a moderately high number of opinion leaders to beat the epistemic performance of an electorate with just two opinion leaders and an almost equal number of followers.

POPULISM RECONSIDERED

Is Populism preferable to Traditionalism and Cosmopolitanism from an epistemic point of view? The answer to this question depends very much on the setting of the parameters. If voters were both competent and independent in their votes, then increasing the number of voters as much as possible is epistemically a good idea. But the independence assumption is not likely to hold. The influence of opinion leaders, in particular, undermines independence (other mechanisms, too, are conceivable and have been discussed).

With regard to opinion leaders, the message is mixed. On the negative side, having a single opinion leader who is even moderately influential can severely reduce the epistemic performance of a group. The same is true if several opinion leaders pull in the same

[42] This less extreme case is the more realistic one, of course: it is hard to think why people would follow opinion leaders unless they at least believed (perhaps wrongly of course) that the opinion leader's opinion was more likely to be correct than their own.

https://doi.org/10.1017/S000712341100041X Published online by Cambridge University Press

direction (that is, are positively correlated). On the positive side, having negatively correlated or a great many uncorrelated opinion leaders causes much smaller reductions in epistemic performance, even if they are quite influential on voters.[43] One reason this is true with uncorrelated leaders may be akin to the mechanism discussed among negatively correlated ones: the effects of uncorrelated opinion leaders (roughly) cancel. Given enough diversity among opinion leaders and voters who are not following their opinion leaders too blindly, relatively good epistemic outcomes can still be expected.

Another effect should be taken into account when evaluating the epistemic effects of opinion leaders – the surprisingly positive effect of polarization. If two opinion leaders are highly polarized (i.e. highly negatively correlated) and if they have about the same numbers of followers, then their influence tends to cancel each other out, and the remaining, independent votes tend to arrive at the correct conclusion. This suggests that political polarization, however undesirable in other respects, is good from an epistemic perspective, as long as there is a balance in the number of followers of each leader. However, if one opinion leader dominates and influences more voters, the probability of correct decisions deteriorates quickly.

One important practical challenge for Populism remains even if no problems of dependence arise: how do the judges know what the majority of voters think? Since it is largely unfeasible (certainly unusual) to have referendums on matters of legal doctrine, this 'epistemic bottleneck' is not a trivial problem.[44] Popular majorities can often be tight and hard to judge from any small and almost inevitably biased sample of people with whom any given judge is likely to interact. Even if a judge were able to draw an unbiased sample (i.e. not just talking to his friends, neighbours and colleagues), taking a small sample can substantially undercut the main mechanism by which Populism produces the good epistemic effects that it does, by drawing on a large population.

CONCLUSION

The independence assumption of the CJT has long been identified as a highly problematic supposition. Sunstein's use of the CJT framework to compare Traditionalism, Cosmopolitanism and Populism calls for an analysis of different mechanisms that undermine independence. In the case of Traditionalism and Cosmopolitanism, the sequential voting process causes problems because later judges will change their vote in the light of earlier opinions. In the case of Populism, a problem with independence arises particularly when opinion leaders influence voters, a scenario that is all too common in politics.

Populism is attractive, probably more attractive than Traditionalism, because it includes many voters and avoids the problem of sequential voting. Given the right constellation, opinion leaders do not too badly reduce the epistemic performance of the population. From this perspective of 'truth tracking', Sunstein's preference for Traditionalism is questionable. If one wants to argue for Traditionalism on epistemic grounds, one needs to provide arguments about how the problems arising from sequential voting can be avoided. The two canvassed here both essentially amount to abandoning Traditionalism in important respects.

[43] For another analysis of why negatively correlated opinion leaders might be good from an epistemic point of view, see Lu Hong and Scott Page, 'Interpreted and Generated Signals', *Journal of Economic Theory*, 144 (2009), 2174–96.

[44] Vermeule, *Law and the Limits of Reason*.

https://doi.org/10.1017/S000712341100041X Published online by Cambridge University Press