Votes and Talk:

Sorrows and Success in Representational Hierarchy

Patrick Grim, Daniel J. Singer, Aaron Bramson, William J. Berger, Jiin Jung, & Scott Page

a version forthcoming in *Episteme* as
"Representation in Models of Epistemic Democracy"

Abstract

Epistemic justifications for democracy have been offered in terms of two different aspects of decision-making: voting and deliberation, or 'votes' and 'talk.' The Condorcet Jury Theorem is appealed to as a justification in terms votes, and the Hong-Page "Diversity Trumps Ability" result is appealed to as a justification in terms of deliberation. Both of these, however, are most plausibly construed as models of direct democracy, with full and direct participation across the population. In this paper, we explore how these results hold up if we vary the model so as to reflect the more familiar democratic structure of a representative hierarchy. We first recount extant analytic work that shows that representation inevitably weakens the voting results of the Condorcet Jury Theorem, but we question the ability of that result to shine light on real representative systems. We then show that, when we move from votes to talk, as modeled in Hong-Page, representation holds its own and even has a slight edge.

Introduction

Epistemic democracy theorists argue that there are instrumental reasons to prefer democratic systems over their alternatives. In particular, the claimed virtues of democracy are epistemic: When a correct answer exists among two alternatives, generally speaking, it will be the more likely choice of a democratic majority, these theorists claim. This constitutes an epistemic justification for democracy – one that is distinct from any of its ethical or social virtues.

Roughly, democracies aggregate information in two ways: they vote among extant alternatives or formulate policies through deliberation. Helen Landemore characterizes these two forms of democratic information pooling as 'counting' and 'talking' (Landemore 2013), while Elizabeth Anderson speaks of 'votes' and 'talk' (Anderson 2006). These two democratic procedures are modeled in two formal arguments that have been used to argue for the epistemic virtues of democracy: the Condorcet Jury Theorem (Marquis de Condorcet 1785/1995) and the Hong-Page diversity result (Hong and Page 2004), respectively. A number of arguments appeal to the Condorcet result as an epistemic justification for democracy (Cohen 1986; Gaus 1997; Grofman & Feld 1988; Miller 1986; Ladha 1992, Estlund 1994, 2008; List and Goodin 2001). More recent work has built on the Hong-Page result (Page 2007, 2011; Landemore and Elster 2012; Landemore and Page 2015). Several recent sources cite both Condorcet and Hong-Page as epistemic justifications for democracy (Anderson 2006, Landemore 2013, Schwartzberg 2015).

Epistemic justifications for democracy in terms of either votes or talk, however, tend to assume full and direct democratic participation across the population.[1] Such a 'pure' democratic form is a distant abstraction from almost all contemporary instantiations. Present-day political democracy is almost invariably *representative* democracy, in which democratic participation is filtered through political institutions with varying levels of hierarchical representation (Dahl 1956; Lijphart 1999; Bednar 2009).

In this paper, we investigate whether the epistemic justifications offered for democracy carry over to or are compromised by hierarchical representation. Our analysis builds on longstanding theoretical results and reports new findings. We show that justifications of democracy in terms of votes and talk don't fare equally well in a representative framework. When it comes to votes, as modeled by the Condorcet Theorem, representative democracy does worse than its purer simple-majority relative. But a representative hierarchy does not compromise the epistemic virtues of *talk* as it is modeled in the Hong-Page result. For a range of variations, the virtues of diverse discussion hold up within a representational structure. We even show that, in one way of implementing representation, having a representational structure actually amplifies the epistemic virtues of democracy.

Our results dovetail nicely with research on ‘deliberative democracy’ and representative electoral voting. We vindicate, for example, Rousseau’s classic critique of representative democracy as an inferior instrumental substitute for direct participation (Rousseau 1762, 1978) by showing the weakening of the Condorcet Jury Theorem in representational hierarchies. But our results also support appeals for various forms of non-electoral deliberation—public spheres, citizen forums, and ‘mini-publics’ (Habermas 1996; Fishkin 1991, 1995, 2009; Fung 2006; Warren & Pearse 2008; Niemeyer 2011; Grönlund, Bächtiger & Setälä 2013)—by showing the robustness of the epistemic virtues of democratic talk in representational structures.

In a first section we outline the sorrows of representative voting in the Condorcet result, with formal treatment and a new and more direct proof of a result from Boland (1989) in an appendix. We question the ability of that result to shine light on real representative systems though, since when it comes to votes, we cannot reasonably expect the representatives to faithfully respect the competencies of the represented. In the next sections, we use computer simulations to demonstrate the resiliency of a representative deliberation by sharing results of representational systems being added to the Hong-Page model.

What we offer are formal results in agent-based models in which abstract structures are offered as modeling representation and abstract dynamics are taken as modeling voting and deliberation. As is well known, the structure of ‘voting’ in the Condorcet theorem abstracts from real voting in many regards, with voter independence a major assumption. The voting in our models simply follows the Condorcet tradition. Our representational structures, similarly abstracted from reality, follow those in Boland (1989) and Boland, Proschan & Tong (1989), and they satisfy the four criteria set for representation in Urbinati and Warren (2008). Formats for group deliberation or discussion are borrowed directly from Hong and Page (2004). Our focus here is on a range of intriguing formal results. Application of formal results requires a wary eye as to whether the abstractions employed adequately capture relevant aspects of political reality, a task beyond the scope of our work here.

I. The Sorrows of Representative Hierarchy in the Condorcet Jury Theorem

The Condorcet Jury Theorem outlines simple conditions under which a collective will be more likely to make a correct decision than an individual. For any odd number of independent voters *n*, obviating the problem of ties, each with probability *p* of getting the question right, the probability of a correct vote of a majority is:

$$\sum_{k=(n+1)/2}^{n} \binom{n}{k} p^k (1-p)^{n-k}$$

What the Condorcet justification for democracy points out is that, as long as voters are more likely to be right than wrong (i.e. $p > 0.5$), the total probability of the majority being right quickly becomes stronger as the population size increases. Figure 1 shows the probability of majority correctness with increased number of voters and individual probability of correctness.

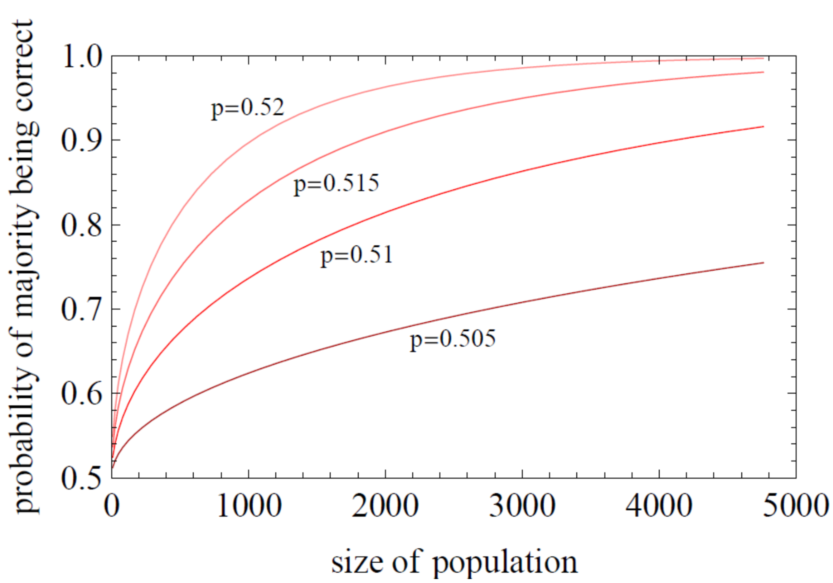

Fig. 1 Probability of a majority of different sizes being correct on a binary question with different probabilities of individual members being correct.

The Condorcet result relies on a number of objectionable assumptions, including that voters have identical probabilities of being correct and that their votes are probabilistically independent. Recent literature shows that both assumptions can be relaxed while maintaining the important result (see, e.g. Boland, Proschan & Tong (1989), Ladha (1992, 1995), Dietrich and Spiekermann (2013), and Boland (1989)). Here we'll focus on another assumption of those who use the Condorcet result to justified democratic institutions, namely the assumption that the result extends to representative democracies.

Boland (1989) demonstrates that the Condorcet result is significantly weakened when applied in the simplest way in hierarchical representation. We offer a significantly more direct proof of that result in the appendix. To understand what's going on in that result, we'll consider a toy example with 9 individuals in a representational system. If we assume that each individual has a probability of correctness of 0.6, the probability that the majority of them is right, in a direct vote, is 0.7334. Now suppose that the 9 were divided in two 3 groups of 3, and each small group first votes to determine the vote of their representatives on a second level. Those representatives simply reflect the majority decision of their constituents, voting again as a group of 3. What is the probability that a majority of 3 representatives for our 9 voters will get the answer right? To calculate the answer, we first consider what the probability of success of the small groups is. Since there are only 3 people in the small groups, again using probability of 0.6 for each individual, the probability of a correct majority decision in a small group is 0.648.[2] The 3 representatives then each have a 0.648 probability, meaning that the probability of a majority of the three representatives being right is 0.7155, which is lower than the probability of the direct vote of 0.7334. Hence, representative hierarchy compromises the epistemic virtues that might be claimed for democracy.

This result can be generalized: with *n* voters, a direct majority vote always has a higher probability of producing the correct outcome than a representative hierarchy in which the voters are partitioned into g (odd) subsets each of which has an odd (and not necessarily equal) number of voters.[3] For the specific case in which there are *g* groups of *g* people each, Figure 2 shows comparative probabilities of a correct decision from a representational structure and for a simple majority decision of all $n = g \times g$ individuals involved.[4]

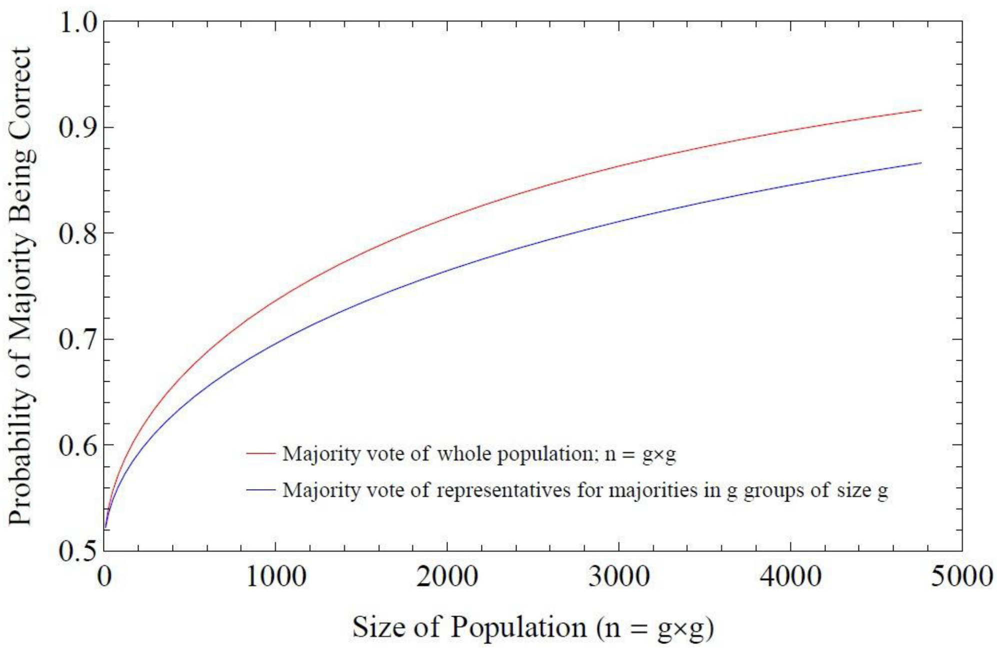

Fig. 2 Probability of a correct final vote in a majority vote of the whole population (red) and a representative vote as in figure 2 (blue): a majority vote among g representatives for groups of size g, each voting the majority vote of their constituency.

As population size goes to infinity, probabilities for both representational and simple majority decisions go to 1. But for any finite *n*, the probability of a correct decision with a straight democratic vote will be higher than for a decision within the comparative representative structure (Boland 1989 and appendix). What this result shows is that representative structures, when conceived of in the most simple way, diminish the overall probability that the institution gets the question right. This ultimately doesn't undermine arguments for democratic institutions based on the Condorcet result though, as the important characteristic, that adding more voters to the system significantly increases the probability that the right result will prevail, still obtains.

Though the Condorcet argument for democracy is weakened on that way of adding in representation, we're concerned that the formal model, constructed in that way, doesn't shine much light on actual representative democracies. The main problem is that it treats the the representatives' probabilities of getting the question right as simply the probability of the majority of their constituents getting it right. But it would be quite odd in a contemporary representative democracy for a representative to poll their constituents on any question they face. Rather, real representatives use their own judgments about what the best answer is, the result of which may or may not reflect what the majority of their constituents would have chosen. So to best model representative democracies, the probabilities of the representatives should be fixed in some other way.

If the probabilities of the representatives are not directly fixed by the probabilities of their constituents, the result described above no longer holds. Consider a simple case in which the representatives' probabilities of correctness are divided into two groups (perhaps representative of political parties), and assume that the distribution of those probabilities is a function of the probabilities of the electorate. Then, depending on the particular details of the case, the representative democracy might do better than the direct democracy, but it might also do worse. For example, if we treat the representatives as more likely to get it right than the average individual (it is their job, after all), then the representatives might do better than a direct vote: If the population has 600 “red” citizens and 400 “blue” citizens, with respective probabilities of correctness of 0.53 and 0.51, then in a direct vote, the total population has a probability of approximately 0.91 of getting it right. If that population is represented by 10 representatives, 6 red and 4 blue, with respective probabilities 0.9 and 0.6, then the representatives have a roughly 0.96 chance of getting it right.[5] Whether the representatives do better is highly dependent on the particular numbers though. If the probabilities of the red representatives are decreased a small amount to 0.8, the result no longer obtains. Further, the outcome is also highly contingent on the size of the representative body: if there are 15 representatives (9 red and 6 blue), rather than 10 (keeping the reduced 0.8 probability for the red representatives), the representative body again comes out ahead at 0.969.

What all of this shows is that any potential effect on the Condorcet argument of applying it to representative democracies rather than direct democracies is highly dependent on the contingent empirical facts of how the representative body is set up and how the representatives’ probabilities of correctness compare to that of those they represent. So, the *a priori* methods used by those who wish to defend democracy with the vote-based Condorcet arguments do not appear to extend to real representative democracies (except, of course, in the unrealistic way we’ve described above). Perhaps what’s surprising though is that this sorrow of epistemic democracy when it comes to voted-based arguments vanishes when we consider the talk-based arguments. So we turn to exploring that result now.

II. Democratic Talk and the Hong-Page Result

As we discussed above, democracies pool information by voting, and in real democracies, that works via representation. Real democracies are also characterized by wide and diverse *discussion*.. Real group discussion and deliberation are clearly complex social processes (Stasser

& Dietz-Uhler, 2001; Tindale, Nadler, Krebel & Davis 2001; Kohn, Paulus & Choi 2011; Levine & Smith 2013; Witte & Davis, 2014; Levine & Tindale 2015). Lu Hong and Scott Page (2004) offer a way to formalize democratic discussion. Though the Hong and Page model plausibly model is highly abstract, it does incorporate features of group interaction that go well beyond mere voting. On that basis it has been embraced in later work as a model for 'talk' as opposed to 'votes' in epistemic democracy (Anderson 2006; Landemore 2013, Schwartzberg 2015).

What Hong and Page claim to show with their model is that the aggregate wisdom from discussion in a diverse group will consistently exceed that from discussion in a group composed solely of those with the highest individual expertise. This has become known as the 'diversity trumps ability' result. There are two forms of the result. Hong and Page offer a mathematical proof for the limiting case of an infinite population, in which the diverse group performs best with probability one (Hong & Page 2004, pp. 16387 ff). That proof is used to explain what happens in more realistic cases of finite population in which the diverse group performs better with high probability. For finite populations, the result is shown to obtain in simulations rather than in formal proof. The same is true of our work here.

Hong and Page start out with an epistemic terrain or problem space that is a line that is n integers long $\{1,2,\ldots,n\}$. The points are arranged in a circle, moving from 1, 2, … to n-1 to n to 1, 2 again. Each of the spots on the terrain is mapped to a real number drawn from a uniform distribution on the interval [1, 100], which can be thought of as scores representing how well that 'epistemic spot' answers some question. Starting from any given point, agents try to find highest points on the terrain by using their assigned heuristics defined as ordered sets of k numbers, each between 1 and l.

An example is helpful: Each agent has a heuristic, which is a set of k integers. Elements of the heuristic are used in sequence. In a case in which $n = 2000$, $k = 3$ and $l = 12$, for example, an agent's heuristic might be (3, 8, 2). Starting at some arbitrary point on the circle, that agent will first evaluate the value of a point three spots to the right from its origin, comparing it with the value at its current position. If the value at this spot is higher it will move to the new position (otherwise it will stay put). It will then try its second heuristic number, evaluating the spot 8 away. If it too is higher, it will move to that point and then try its third heuristic number 2. The agent keeps checking points to the right using it heuristic numbers in turn, moving to higher points, until it reaches a stopping point at a local maximum from which none of its heuristic numbers can guide it to a point with a higher value. See Figure 3 for another example.

We can measure the overall expertise of an agent in this model in terms of how well the agent can be expected to do in finding the maximums on the terrain. Let $\varphi(i)$ denote the stopping point for a heuristic φ from an initial starting point i, with $V[\varphi(i)]$ the achieved value at that stopping point. The overall performance of an agent with heuristic φ is the expected value $E[V; \varphi]$ of the stopping points for each of the n possible starting positions i. If we assume that all of the starting points are equally likely, that simply amounts to the average of the $V[\varphi(i)]$ for all i. Given any specific epistemic terrain, agents individuated by heuristic can be ranked in terms of overall performance. Page and Hong's 'best-performing' agents on a terrain are those with the highest expected values: those with the highest average value reached from the n points on that specific terrain.

The Hong Page result compares the expected performance of collaborative groups of agents exploring a terrain together. Their comparison is between expected values for (a) a group of 'best-performing' agents on a given terrain—the 9 with the highest expected values on that terrain, for example—and (b) a diverse group of 9 agents assigned random heuristics. The way that agents act together in groups is a variation on the pattern of individual exploration outlined above. The procedure is a sequential relay among, in this example, the 9 group members. Starting from a given point, the first agent finds the highest point she can get to (as outlined above). Once she has reached that personal stopping point, the baton is passed to the next member in the group. The second member of the group starts from the stopping point of the first member in order to see whether her heuristic can reach a point with a higher value from there. The relay procedure continues in sequence through the 9 participants until a point is reached from which none of the 9 agents can find a point with a higher value.

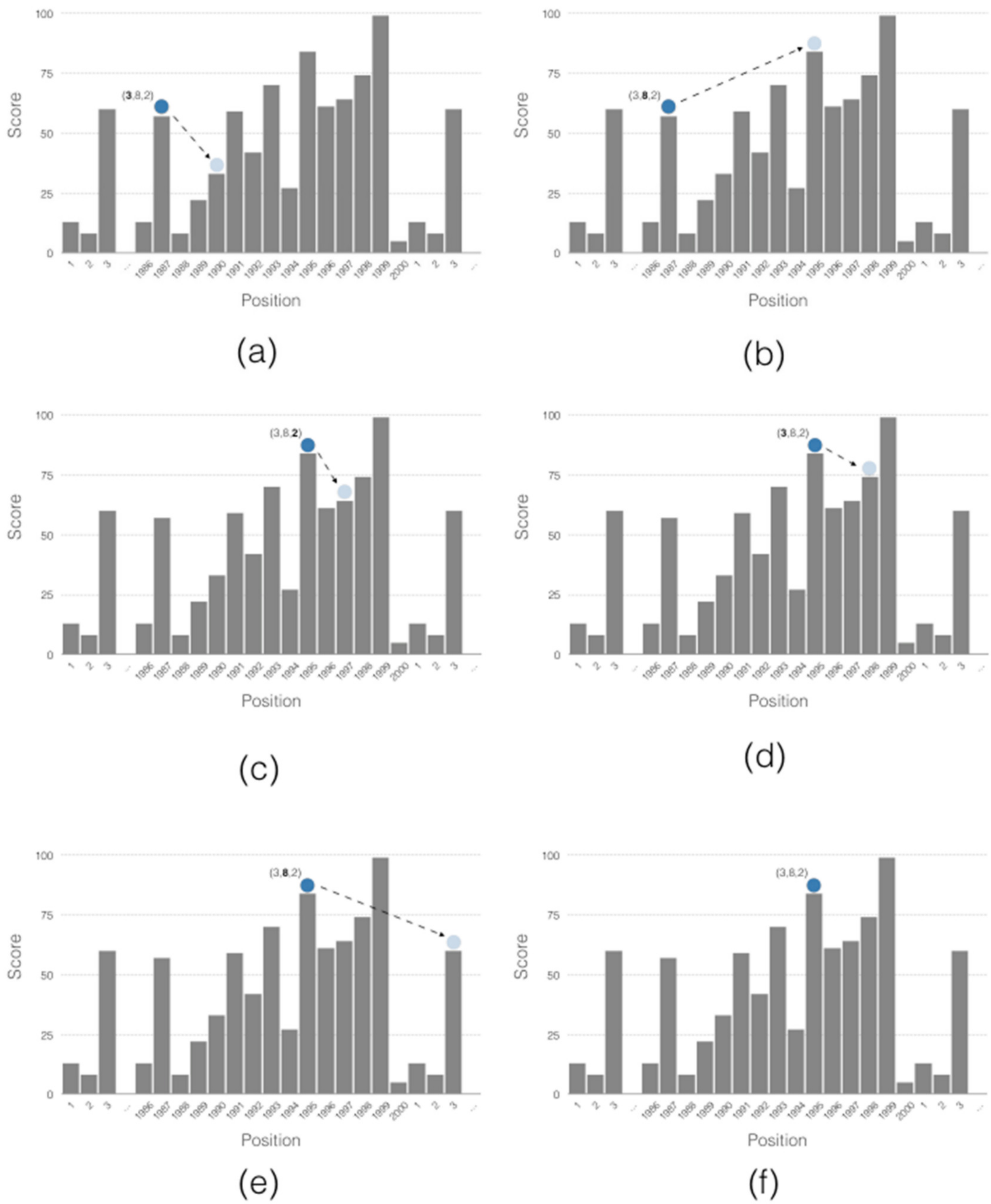

Fig. 3 In this example, the agent begins at position 1987 with the heuristic (3, 8, 2). The agent first checks the value at location 1990 (1987 + 3), but finding it lower than that of its current position it remains at position 1987. Moving to its next heuristic number 8, the agent finds location 1995 (1987 + 8) to have a higher value and so moves to the new location. From there the agent deploys its heuristic number 2, then 3, followed by 8. In checking positions 1997, 1998, and 3 (1995 + 2, 1995 + 3, and 1995 + 8 mod 2000, respectively), the agent finds no spot with a greater value than the one it currently occupies, and therefore stops.

The overall group performance can be measured like individual performance. Let $g(i)$ denote the relay stopping point for a group g from an initial starting point i, with $V[g(i)]$ the achieved value at that stopping point. The performance for a group g is the expected value $E[V; g]$ of the stopping points for each of the possible group starting positions i. If we assume that all of the starting points are equally likely again, that amounts to the average of the $V[g(i)]$ for all i.

What Hong and Page establish is that a group of agents with random heuristics usually does better in this relay group problem-solving format than a group of the individually best-performing agents. They term this the 'Diversity Trumps Ability' result, and it is typically taken to show the virtues of democratic discussions over those that include only the individual experts.

The Hong-Page result has been challenged with regard to both how unexpected it should be and for the relevance of the mathematical theorem offered in explanation (Thompson 2014). By the theorem, diversity trumps ability with probability 1, given stark conditions regarding group and population size, problem difficulty, and group diversity. When those conditions are relaxed, the probability is no longer 1. That said, replications show the result to be extremely robust: the epistemic success of a group of random heuristics proves consistently superior to that of a group of those which individually score the best.

III. Resilience of Representational Democracy in the Hong-Page Model

Like above, we ask here how the Hong-Page model can be expanded to apply to representational hierarchies, like the kinds discussed by Anderson (2006), who discusses diverse citizens working in smaller groups. This structure is also touted in advocacies of citizen forums and 'mini-publics' (Fishkin 1991, 1995, 2009; Urbinati & Warren 2008). But how do the democratic virtues of talk hold up in that kind of representative structure? Here we consider two different idealized instantiations of a deliberative hierarchy. The first uses a relay format based on the original dynamics of the Hong-Page model. The second uses an alternative 'tournament' format.

In a representational hierarchy for relay discussion, a relay-based deliberation (like the one in Hong and Page's original model, described above) is instantiated in two levels in the manner of figure 4. Starting from an initial point on the terrain, the first subgroup uses the relay format to find the highest point they can achieve from that point, which can be seen as reflecting the pooled wisdom of that subgroup. The first of the three representatives reports that stopping point to the second subgroup, which advances the deliberation by *starting* at that stopping point. The second subgroup finds its stopping point, which again reflects the group's combined wisdom. That stopping point is the starting point for the third group, and the rotation continues through the three representatives, stopping at a final point from which none of the represented deliberations within the subgroups is able to achieve a higher value. The performance for the whole group is the average final value achieved from all 2000 points.

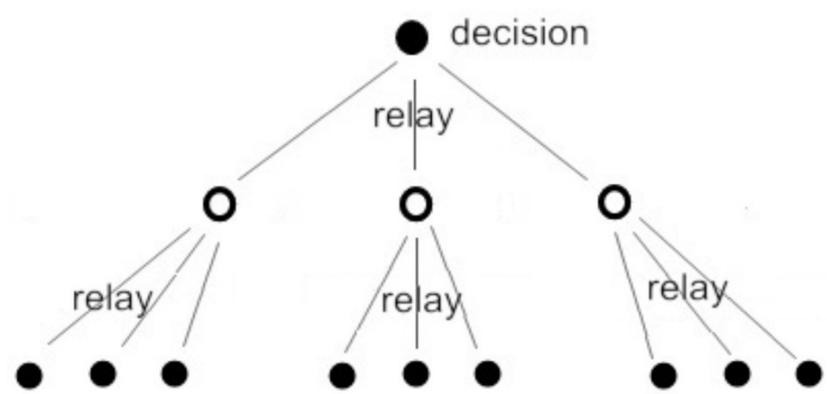

Fig. 4 Representational discussion using sequential relay dynamics

In the representational structure used in testing the Condorcet result, representatives simply used majority decisions of their sub-groups in a second round of voting. In this representational structure for Page-Hong, the representatives can be thought of as using the relay deliberation of their sub-groups as their ‘heuristic’ in a second round of relay deliberation. Each subordinate group comes to the result it believes best for each point and sends its delegate with its solution to the higher body. In representational deliberation on the model of Hong-Page, the delegates from each group meet and use a second ‘relay’ deliberation in seeking still more optimal results. Like in the Condorcet parallel, a representative’s position is conditioned by the lower tiered result, though ultimate policy is not tied to the initial belief of any one of the constituents regarding the optimal solution to the problem at hand.

Using this model of representation, we compare relay-based discussions among all 9 members and relay-based discussions among three representatives of three groups of 3. In this model of discussion, in contrast to the Condorcet model of voting, representation holds its own. Results are virtually tied. Over a typical 1000 runs, the average scores for open and representative discussions among 9 randomly-chosen individuals are 94.89775 and 94.89646 respectively—a difference of 0.0014%. Moreover, representation actually does better in terms of the absolute number of cases, which indicates just how close a tossup this relay schema is. Representation beat open discussion in 516 of 1000 cases; open discussion beat representation in 463. The left graph in Fig 6 shows nearly identical probability density functions and therefore a lack of any significant difference in terms of the Kolmogorov-Smirnov two-sample test. We get similar results when we compare an open discussion of the 9 ‘best’ heuristics with a representational hierarchy for the same individuals. Averages come in at 92.7688825 and 92.7493605 respectively, a difference of 0.02%.

Results for relay representation therefore confirm the basic findings of Hong and Page. In both open relay discussions and representational relay discussions, groups of randomly-chosen participants perform approximately 2% better than groups of the individual top performers.

IV. Representation's Edge in a Hong-Page Variation

In a second instantiation of representational hierarchy, we model discussion as a simultaneous tournament rather than a sequential relay. Hong and Page claim that this variation gives essentially the same results. With a representational hierarchy, that turns out not to be true.

In this tournament format for group deliberation, we eliminate the 'single file' or 'each in turn' assumption of the relay. For discussion among a group of 9, all 9 agents use their heuristics simultaneously, each reporting the stopping point with their highest value. The stopping point with the highest value achieved by *any* of them is then treated by all members as a new starting point. At each stage all agents in the group simultaneously leapfrog to the highest point achieved by any of them, without having to pass a baton in sequence. Although this decisional format has a different dynamic and can be expected to hit different points in the process, the final stopping point will again be a local maximum from which none of the agents' heuristics can reach a higher point. The performance for the group is the average final value achieved from all of the 2000 starting points.

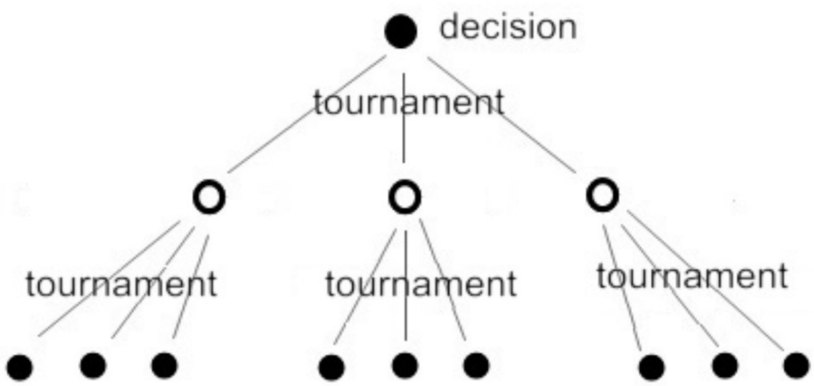

Fig. 5 Representational discussion using simultaneous tournament dynamics

For representation we again use the simple example of 3 groups of 3. Each small group engages in a tournament deliberation and find their highest group value for each of the 2000 points in the epistemic terrain. The result of that deliberation is handed to a representative, who uses that result of his group's pooled wisdom as a guide in tournament deliberation with other representatives on the second level. On the second level, for each point on the terrain, all representatives simultaneously report the result achieved in their groups' deliberation. The point at which the highest value is reported then becomes a new starting point. Representatives then report the highest value each of their groups has reported from *that* point, and so on. The performance for the group again is the average final value achieved from all of the 2000 starting points.

All that has been changed in this instantiation of representation is the structure of discussion on each level: a simultaneous tournament in this case as opposed to a sequential relay in the previous section. Here, however, we have different results: this second representational structure not only preserves the virtues of democratic discussion but amplifies them. Over 1000 runs, the average value for a group of 9 random agents with tournament discussion among representatives of groups of 3 comes in higher than the average for those same 9 in open tournament discussion. The difference in each case is slight: typical averages are 96.588187 and 96.450588 respectively, a difference of 0.14%. What is most striking though is that representation's lead, however slight, is evident *in all 1000 of 1000 runs*. In a tournament structure, though not by much, representation robustly beats out direct democracy.

Figure 6 shows the outcome and significance results for relay discussion (left), in which representation does much the same as open discussion, and for tournament discussion (right), in which representation shows an edge. A Kolmogorov-Smirnov two-sample test[6] confirms that the distribution of relay outcomes are nearly identical, but the outcome distribution for the tournaments are significantly different. Except for the fact that representation holds its own in tournament as well as relay discussion, the advantages for tournament representation are so small that any attempt to read off real practical implications is in vein. The theoretically intriguing fact remains that tournament representation shows higher scores so robustly though.

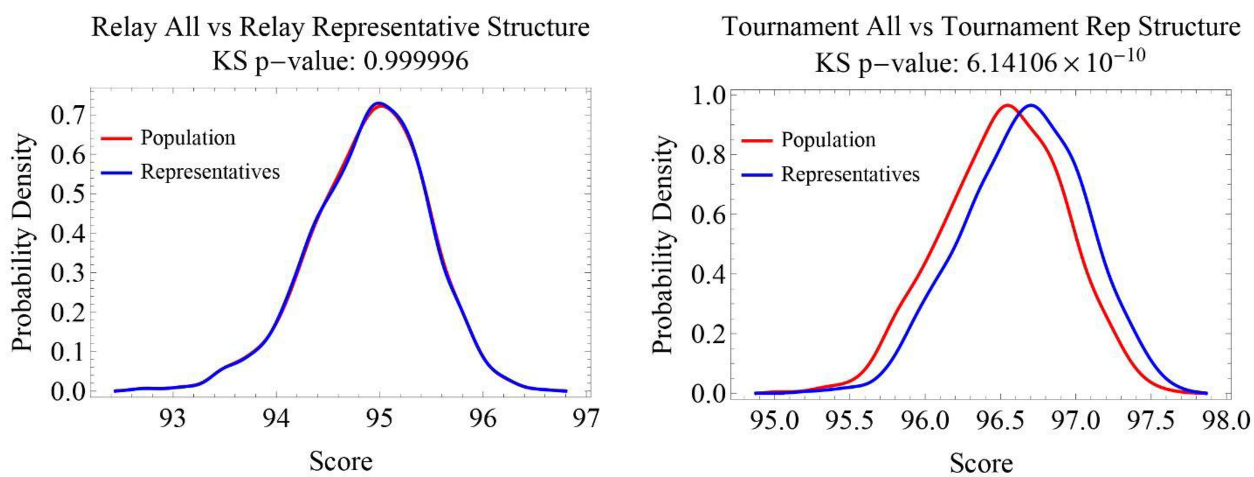

Fig. 6 The empirical probability density functions of the highest points reached by the population and representatives in the relay (left) and tournament (right) experiments. Results of the Kolmogorov-Smirnov two-sample test are shown above each.

	Relay	Tournament
Randoms	516/1000 0.0014	1000/1000 0.1376*
'Bests'	424/1000 -0.0293*	999/1000 0.0879*

Table 1: Results indicate the number of runs in which the representative structure outperformed the direct set up. The top number indicates the number of runs in which representation beats direct, while the bottom number indicates the average margin between the two instantiations. The star indicates significance at alpha < 0.001. Note, however, that the result in the top left is not significant even at alpha = 0.05

What explains the epistemic superiority of tournament representational systems? It cannot be those groups do better than non-representational groups from every point on a random landscape. We'll show that we can construct landscapes in which either will perform best from a given point.

First, consider a way in which a representational group can outperform an open group. The small groups in the representational structure can sometimes get to spots that the single larger group cannot because the larger group will get stuck on a local high point that one of the small groups, lacking a heuristic in the larger group, would not get stuck on. The small group can get past that local high point that the larger group is stuck on, managing to find a higher peak beyond. Here is an example to clarify:

Consider a case in which we simplify individual heuristics to be single numbers (rather than triplets) of integers 1 through 9. Let the whole group be on agent with each of the possible heuristic values. Let the terrain have a local peak at spot 7 (value 96), a higher peak at spot 20 (value 99), and values 85, 89, and 92 at spots 5, 10, and 15. All other values are 50 (Figure 7).

Starting from point 0, the group of 9 will get stuck on 7 and won't be able to reach anything higher beyond that. But a small group with a 5 but no 7 will find the higher value at spot 20 in four hops to increasingly higher points.

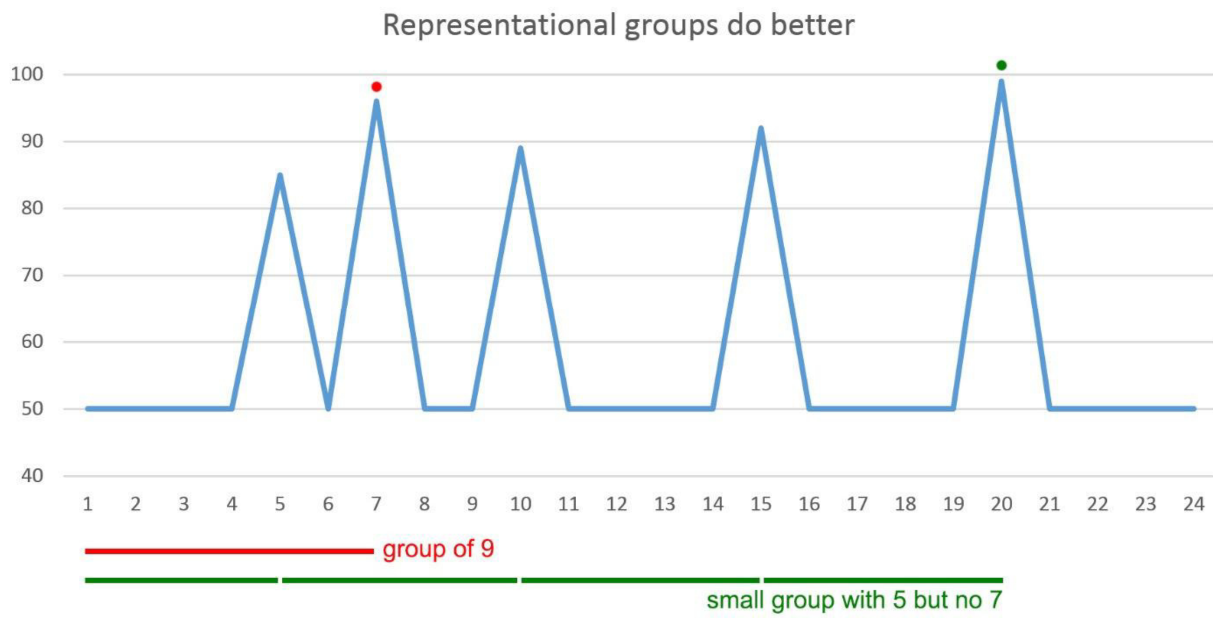

Fig. 7 A landscape on which a small group does better than the full group of 9

So that kind of set up might explain why representation groups do better. But there will also be cases in which the group of 9 does better than the three groups of 3. Let our groups be group A, 3 agents with heuristics {1}, {3}, and {5}; group B with {2} {4} and {6}; and group C with {7}, {8}, and {9}. Let our terrain be a value of 86 at spot 5, 84 at spot 6, 90 at spot 7, 82 at spots 8, 10, and 12, a high of 99 at spot 13, 94 at 15 and 50 for all other points (Figure 8). In that case, starting from point 0, the group of 9 will hit 90 with heuristic 7, jumping from there to 99 at spot 13 with a heuristic 6. The smaller groups will not do so well. Group A [1 3 5] will go to 86 at spot 5, but neither 1, 3, nor 5 will take it higher. Group B [2 4 6] will go to 84 at spot 6, but neither an additional 2, nor 4, nor 6 will take it higher. Group C [7 8 9] will go to 90 at spot

7. From there it will go to 94 at spot 15 using heuristic 8, missing spot 13 in the process and with nothing but 50's from there on out.

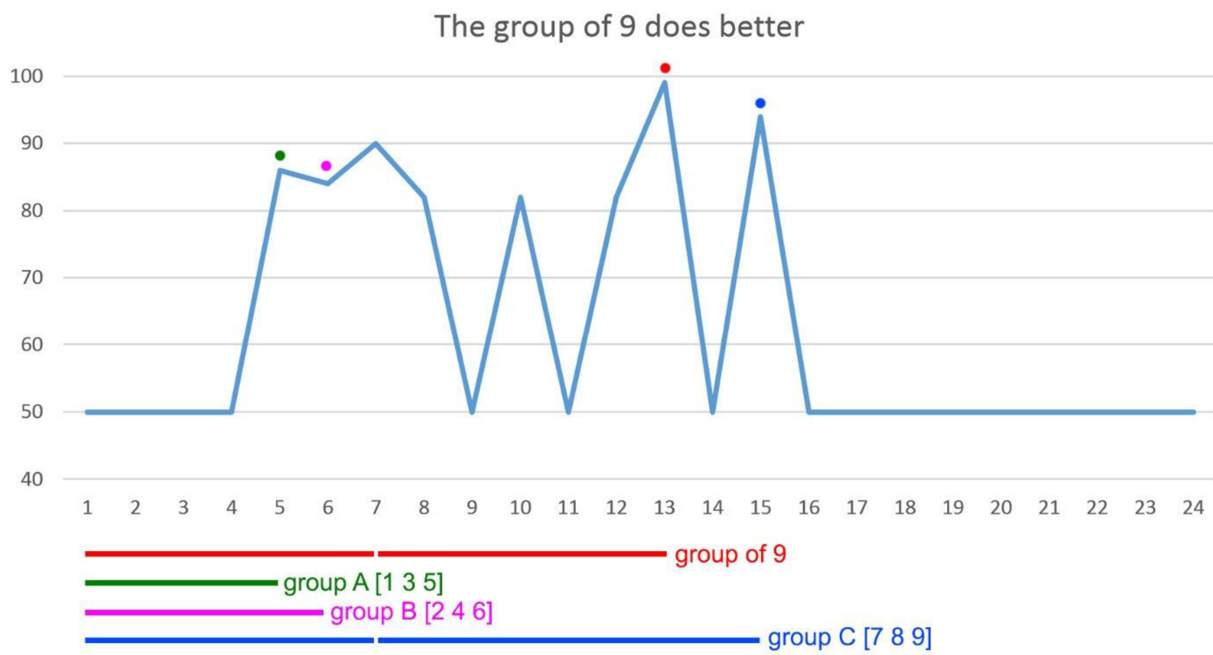

Fig. 8 A landscape on which the group of 9 does better than smaller groups

So, for individual starting points on random terrains, it is possible for either whole groups of 9 to do better than 3 representational groups of 3 or for the contrary to hold. The frequency with which these appear is not the same, however. It is far more common for representational groups to do better than the full group (across 100,000 starting points, representational groups do better in 3278; the full group does better in only 183).

What explains this difference? One clue is the number of points sampled by each structure. Taken over 100 runs, the average number of distinct points sampled from each of 2000 points using a tournament among 9 agents is just over 26. The average number of distinct points sampled using a representative structure of 3 groups of 3 agents is a significantly higher at

27.5. So, representational groups do slightly better because they sample a slightly larger number of points on the terrain.

Although we have not emphasized it here, it is worth noting that the model of tournament discussion achieves higher epistemic scores than relay discussion across the board, regardless of whether we are dealing with individually best-performing individuals or random participants and regardless of whether the discussion configuration is representative or whole-group.

Lastly, it is instructive to point out that the original Hong-Page diversity result is maintained by our representative structure. Above, we compared the performance of groups of both random and best-performing agents separately in direct versus representative structures. The Hong-Page diversity result compares groups of random agents and groups of 'the best.' That result is maintained in representational structures: in more than 97% of our runs we find that a random collection of agents still outperforms a group of the best-performing individuals. Diversity remains important in the representational structure, and is enhanced by the tournament instantiation.

	Relay	Tournament
Direct	973/1000 2.2329*	973/1000 1.3907*
Representation	973/1000 2.1702*	977/1000 1.4404*

Table 2: Results indicate the number of runs in which a random group of nine outperformed a group of the 9 individually best-performing agents. Top numbers indicate the number of runs in which representation beats direct, while the bottom number indicates the average margin between the two instantiations. The star indicates significance at alpha < 0.001.

V. Robustness

With simulation results like those offered here, the robustness of the results under model variations key to understanding how broadly the result applies. In all of the results above, we use a terrain created by picking a random number between 1 and 100 for each of 2000 spots. We might instead pick a normal distribution for our terrain, with for example a mean of 50 and a standard deviation of 10. Results will be a predictable transform of those above, though they will also show a bit more 'daylight' as a result. Using a normally-distributed terrain, results for open and representational tournament discussion come in at 69.77 and 69.95 respectively rather than 96.51 and 96.65, a percentage difference of 0.25% rather than 0.14%.

Throughout, we use a model in which the three heuristic values for each agent were drawn from a pool of 12. What happens if we expand that pool, so that heuristics are composed of three numbers from a pool of 24, for example, or 36? Since the original Hong-Page effects can be understood in large part in terms of redundancy (Hong & Page 2004, Thomson 2014), we should expect this to make a difference. Average results across 1000 runs for pools of 12, 24 and 36 are shown in Figure 9.

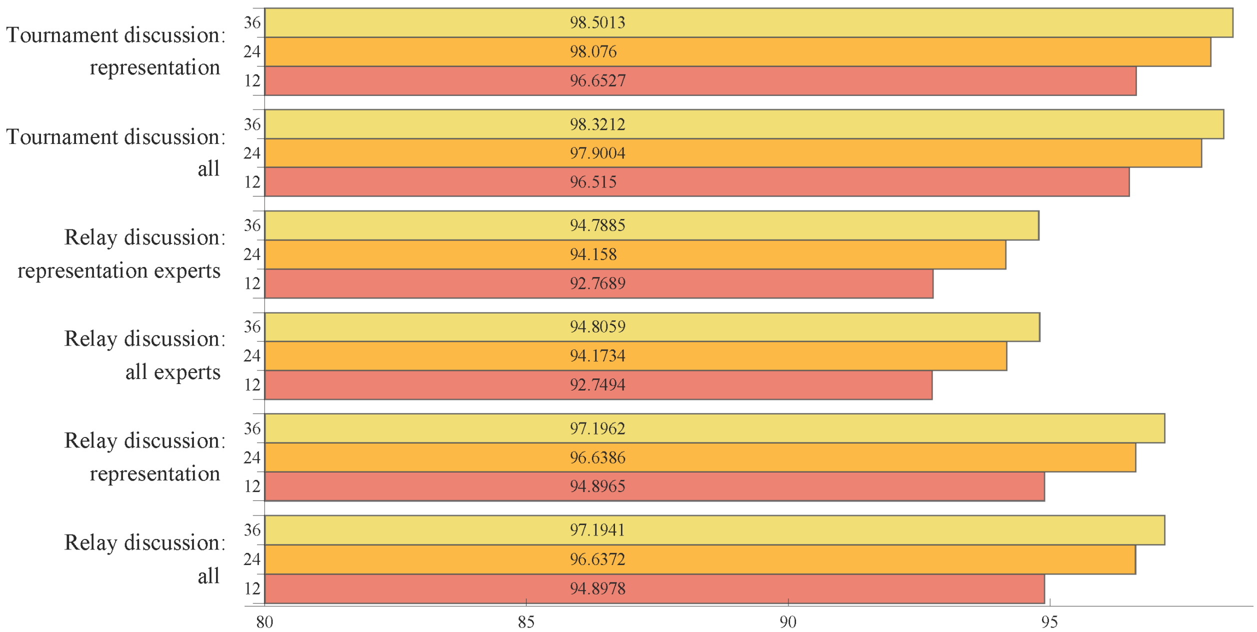

Fig. 9 The effect of enlarging the pool from which heuristics values are drawn, averages over 1000 runs.

For cases of relay representation in heuristic pools of different sizes, there is no consistency in whether the representational structure does better or worse than full participation. The same proves true in the case of best-performing individuals in full or representational relay discussion. Whatever the size of pool heuristics, a representational structure for tournament discussion does better than full discussion among the population, though the margin of superiority declines as the pool of heuristics is increased. The central fact highlighted above, however, holds throughout: In *every* case in *every* run of 1000 the representational structure for tournament discussion comes in with a higher score.

Do results hold for larger numbers of larger groups? If we consider not 3 groups of 3 agents, but 5 groups of 5, results remain roughly comparable, with a particularly robust result for the triumph of representation in the case of tournament discussion. For relay discussion of the form considered in section III, a representational structure still does almost exactly the same as relay discussion among the population as a whole. The same holds for relay discussion among best-performing individuals. The central results that we have highlighted for tournament discussion prove robust in the expanded case: in tournament discussion representation does better (an average of 97.180 as opposed to 97.096). Here as before, the most significant fact may be that representational representation of tournament discussion once again did better in every one of the 1000 runs.

VI. Conclusion

As mentioned in the introduction, epistemic democrats claim that democratic institutions provide instrumental reasons to favor democratic institutions. According to these theorists, democracy produces better epistemic outcomes. Formal arguments in support of such a claim have been based largely on models of either voting (Condorcet) or diverse discussion (Hong-Page) in a ‘pure’ or ‘direct’ democracy (Cohen 1986; Estlund 1994; Anderson 2006; Landemore and Page 2015, Schwartzberg 2015). But direct democracy of such a form is far removed from almost all forms of political procedure in use today. Contemporary political democracy is almost invariably *representative* democracy, in which democratic participation is filtered through levels of hierarchical representation. Here, we focused on whether the epistemic arguments for democracy carry over to abstract models of hierarchical representation as well. The point of departure for the paper is to ask whether two theorems touted by epistemic democrats would hold up in a representative or hierarchical structure.

Our results indicate that the answer depends first and foremost on whether it is virtues of voting or discussion that are at issue. Both analysis and simulation show that the epistemic virtues of voting that appear in the Condorcet theorem are significantly weakened in a representational structure. As we suggested above, it is also unclear that theorem tells us much about real representative democracies that is independent of the contingent empirical facts of how those systems are set up. When it comes to 'votes,' inclusive democracy proves epistemically superior by the lights of the Condorcet theorem, just as Rousseau argued (Rousseau 1762/1978).

The story is different for the epistemic virtues of *discussion*. Within a relay deliberation like in the original Hong-Page model, epistemic results for representational structures prove virtually identical to results for full deliberation among all participants. So, in the case of talk rather than votes, representation holds its own. In a simultaneous tournament structure, in fact, epistemic scores for a representational structure slightly but persistently edge out those for the population as a whole.

These results lend formal support to arguments for federalism (e.g. Bednar 2009) and dicta like Justice Louis Brandies' claim that states in a federated system act as the "laboratory of democracy." Small groups search for optima independently, with those results handed on for a second-order deliberation. Under certain conditions this process proves as effective as one in which an entire population of agents seeks a best solution on the given terrain. The argument for hierarchical representation in terms of efficiency is well-established (Pitkin 1967; Dahl 1983).[7] With regard to epistemic benefits of talk, as opposed to votes, our results indicate that representation's virtues of efficiency need not represent an epistemic sacrifice.

Within the constraints of the abstract models used here, our results offer a formal argument for the value of deliberation grounded in first-order discussion within small distributed groups. As noted in introduction, there is a wide contemporary literature that argues for distributed forms of small-group public discussion (Fishkin 1991, 1995, 2009; Fung 2006;

Warren & Pearse 2008; Niemeyer 2011; Grönlund, Bächtiger & Setälä 2013). Our results may be taken as a formal vote of confidence that such a structure can hold its own.

Appendix: Limitations of the Condorcet Jury Theorem in Representational Structures[8]

We suppose there is a vote between binary exclusive options A and B, where each voter has a probability $p > 0.5$ of choosing the correct option. We compare two forms of aggregation. (1) In a direct majority, each voter casts a ballot for A or B; whichever receives the most votes is the outcome. (2) In a representative structure we assume the voters are partitioned into g groups of size m, where both g and m are odd. Voting takes place in two stages. In the first stage members of each group g vote using direct majority. In the second stage whichever outcome was chosen by the majority in each group is cast as the vote for that group in a second majority vote among group representatives. The option winning the second vote is the representative outcome.

With these conditions a direct majority vote is more likely than a representative structure to produce a correct outcome. This result is anticipated in Boland (1989). In this appendix we offer what we consider to be a more direct and accessible proof.

To provide insight for the general result we first consider the case of nine voters who are partitioned into three groups of size three. We can then make the following observations:

Observation 1. With nine voters, a direct majority produces the correct outcome if and only if five or more of the voters choose the correct option.

Observation 2. With nine voters partitioned into three groups of size three, the representational structure always produces the correct outcome whenever six or more voters choose the correct option. No matter how the six correct votes are assigned to the three groups of three, there will inevitably be a majority of groups with a majority of correct votes.

Observation 3. With nine voters partitioned into three sets of size three, the representational structure cannot produce the correct outcome if three or fewer voters make the correct choice. No matter how the three correct votes are assigned to the three groups of three, it is impossible to obtain a majority of groups with a majority of correct votes.

Given these observations, the interesting cases arise when either four or five voters make the correct choice:

Observation 4. With nine voters partitioned into three sets of size three, the representative voting structure can produce either outcome when exactly four or exactly five voters make the correct choice.

To see why this is true, consider two realizations of choices and partitions using c for the correct choice and i for incorrect.

Example 1: $\{(c, c, c), (c, i, i), (c, i, i)\} \rightarrow \text{Outcome} = i$

The first subset has three voters who make the correct choice and the other two subsets have exactly one voter who makes the correct choice. Thus with five voters making the correct choice the representative outcome is still incorrect.

Example 2: $\{(c, c, i), (c, c, i), (i, i, i)\} \rightarrow \text{Outcome} = c$

In both of the first two subsets exactly two of the voters make the correct choice, and in the third subset none do. Thus the partitions in Example 2 produce the correct outcome even though only four of the voters make the correct choice. These examples lead to Observation 5:

Observation 5. With nine voters partitioned into three sets of size three, there are distributions of votes such that the representative hierarchy produces the correct outcome when direct majority voting does not. There also exist distributions of votes in which the opposite holds: direct majority voting produces the correct outcome while representative hierarchy does not.

Neither form of aggregation dominates the other in the sense of always producing at least as good an outcome for any distribution of votes. What remains to be established is whether one institution is better *on average* given our assumptions regarding the individuals' probability of correctness. The following claim provides an answer to that question.

Claim 1. With nine voters, direct majority rule has a strictly higher probability of producing the correct outcome than a representative hierarchy in which the voters are partitioned into three sets of size three.

Proof. As before we use c to denote the correct outcome and i for the incorrect outcome. From above, we know that if six or more, or three or fewer, voters vote c then we get the same outcome for direct majority and representative hierarchy voting. Let $R(4, c)$ denote the set of all realizations in which exactly four of the voters vote c and the outcome of a representative hierarchy is c. Designate the sets $R(4, i)$, $R(5, c)$ and $R(5, i)$ similarly. The sets $R(4, i)$ and $R(5, c)$ produce the same outcome as direct majority; i.e., for any realization in $R(4, i)$, four of the voters vote c and the group outcome is i, the same as a direct majority vote with four c votes. The only cases in which a direct majority vote and representative democracy produce different outcomes are therefore $R(4, c)$ and $R(5, i)$. Within $R(4, c)$ a representative hierarchy produces the correct outcome and direct majority voting does not. Within $R(5, i)$ the opposite holds. To complete the proof, we rely on the following lemma regarding cardinality of the sets:

Lemma 1. $| R(5, i) | = | R(4, c) |$

The proof of the lemma is as follows: Consider any realization in $R(5, i)$. It will consist of five voters who vote c and four who vote i such that two of the three subsets in the partition have a majority of voters who vote i. The only way this can happen is if two of the subsets contain two voters who vote i, as in example 1:

Example 1: $\{(c, c, c), (c, i, i), (c, i, i)\} \rightarrow$ Outcome = i

Given this example, we can construct a mirror realization in which every c signal becomes an i and every i signal becomes a c.

Mirror of Example 1: $\{(i, i, i), (i, c, c), (i, c, c)\} \rightarrow$ Outcome = c

The lemma follows from two observations. First, any realization has a unique mirror realization. Second, the mirror realization to any realization in $R(5, i)$ lies in $R(4, c)$ and vice versa.

To complete the proof of Claim 1 we only need to show that any realization in $R(5, i)$ it has a higher probability than its mirror realization in $R(4, c)$. This follows directly from the fact that individuals vote c with a probability $p > 0.5$. The probability of any particular realization in $R(5, i)$ – the cases in which a direct majority produces a correct result while a representative hierarchy does not – is $p^5(1 - p)^4$. The probability of the mirror realization in $R(4, c)$ is $p^4(1 - p)^5$, which is always smaller than $p^5(1 - p)^4$ when $p > 0.5$, thus completing the proof.

General Result

We can now state a general result given n voters partitioned into g (odd) subsets.

Claim 2. With n voters, a direct majority vote has a strictly higher probability of producing the correct outcome than a representative hierarchy in which the voters are partitioned into g (odd) subsets each of which has an odd (and not necessarily equal) number of voters.

Proof. Let $R(v, x)$ denote the set of all voting realizations in which v voters choose c and the representative hierarchy produces the outcome x. Then let D_C denote the set of all $R(v, x)$'s in which direct majority voting produces a correct outcome while a representative hierarchy does not, and let D_I denote the set of $R(v, x)$'s where the opposite holds.

By construction:

$D_C = \{R(v, i) \,\&\, v > n/2\}$

$D_I = \{R(v, c) \ \& \ v < n/2\}$

For each set of $R(v, i)$'s in D_C define its partners as $R(n - v, c)$. Note that partners create a one to one mapping between sets in D_C and sets in D_I. Suppose that there exists at least one realization in $R(v, i)$; i.e., that $R(v, i)$ is nonempty. As above, define the mirror realization by changing every *c* vote to an *i* vote and every *i* vote to a *c*. It follows that the mirror realization of a realization in $R(v, i)$ belongs to its partner set $R(n - v, c)$. Similarly, every realization in $R(n - v, c)$ has a unique mirror realization in $R(v, i)$. Therefore, there are the same number of realizations in $R(v, i)$ as there are in $R(n - v, c)$. Also following the logic from above, the probability of any given realization in $R(v, i)$ equals $p^v(1 - p)^{n-v}$ which strictly exceeds the probability of its mirror realization with probability $p^{n-v}(1 - p)^v$ as long as $v > n/2$. This implies that each set in D_C has a higher probability of occurring than its corresponding set in D_I, which completes the proof.

References

Anderson, E. (2006). The epistemology of democracy. *Episteme*, 3(1-2), 8-22.

Bednar, J. (2009). *The Robust Federation: Principles of Design*. New York: Cambridge Univ. Press.

Beisbart, C. & L. Bovens (2013). Minimizing the threat of a positive majority deficit in a two-tier voting system with equipopulous units. *Public Choice* 145 (1-2), 75-94.

Boland, Philip J. (1989) Majority systems and the Condorcet jury theorem. *Journal of the Royal Statistical Society. Series D* (*The Statistician*), 38, 181-189.

Boland, Philip J., Frank Proschan & Y. L. Tong (1989). Modeling dependence in simple and indirect majority systems," *Journal of Applied Probability*, 26, 81-88.

Condorcet, M. J. A. N. d. C., Marquis de. (1785/1995). An essay on the application of analysis to the probability of decisions rendered by a plurality of votes (fifth part). In I. McLean & A. B. Urken (Trans. & Ed.) *Classics of Social Choice*. Ann Arbor: University of Michigan Press, pp. 91-112.

Cohen, Joshua. 1986. An epistemic conception of democracy. *Ethics,* 97, 26–38.

Dahl, Robert A. (1956). *A Preface to Democratic Theory*. Chicago: University of Chicago Press.

Dahl, Robert A. (1983). *Dilemmas of Pluralist Democracy: Autonomy versus Control*. New Haven, CT: Yale University Press.

Estlund, David (1994). Opinion leaders, independence, and Condorcet's jury theorem. *Theory and Decision*, 36, 31-162.

Estlund, David (2008). *Democratic Authority: A Philosophical Framework*. Princeton, NJ: Princeton University Press.

Fishkin, J. S. (1991). *Democracy and Deliberation: New Directions for Democratic Reform*. New Haven, CT: Yale Univ. Press.

Fishkin, J. S. (1995). *The Voice of the People: Public Opinion and Democracy*. New Haven, CT: Yale Univ. Press.

Fishkin, J. S. (2009). *When the People Speak: Deliberative Democracy and Public Consultation*. New York: Oxford University Press.

Fung, A. (2006). Democratizing the policy process. In R. Goodin, M. Moran & M. Rien, eds., *The Oxford Handbook of Public Policy.* New York: Oxford Univ. Press, pp. 669-685.

Gaus, G. (1997). Does democracy reveal the voice of the people? Four takes on Rousseau. *Australasian Journal of Philosophy*, 75, 141-162.

Goodin, Robert E. & Kai Spiekermann (2012). Epistemic aspects of representative government. *European Political Science Review*, 4:3, 303-325.

Grofman, B., & S. L. Feld (1988). Rousseau's general will: A Condorcetian perspective. *American Political Science Review*, 82, 567-576.

Grönlund, K., A. Bächtiger & M. Setälä (2013). Eds., *Deliberative Mini-Publics: Practices, Promises, Pitfalls*. New York: Columbia University Press.

Habermas, J. (1996). *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. Trans. T. Berger. Cambridge, MA: MIT Press.

Hong, Lu & Scott E. Page (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences, 101*, 16385-16389.

Hong, Lu & Scott E. Page (2009). Interpreted and generated signals. *Journal of Economic Theory*, 144, 2174-2196.

Kohn, N. W, P. B. Paulus & Y. Choi (2011). Building on the ideas of others: An examination of the idea combination process. *Journal of Experimental Social Psychology*, 47, 554-561.

Ladha, K. K. (1992). The Condorcet jury theorem, free speech, and correlated votes. *American Journal of Political Science*, 36, 617-634.

Ladha, K. K. (1995). Information pooling through majority-rule voting: Condorcet's jury theorem with correlated votes. *Journal of Economic Behavior and Organization*, *26*, 353-375.

Landemore, Helene (2013). *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton, NJ: Princeton University Press.

Landemore, H., & Elster, J. (Eds.). (2012). *Collective Wisdom: Principles and Mechanisms*. New York: Cambridge University Press.

Landemore, Hélène and Scott E. Page (2015). Deliberation and disagreement: Problem solving, prediction, and positive dissensus. *Philosophy, Politics, and Economics,* 14(3), 229-254.

Levine, J. M., & E. R. Smith (2013). Group cognition: Collective information search and distribution. In D. E. Carlston & D. E. Carlston, eds., *The Oxford Handbook of Social Cognition*. New York: Oxford Univ. Press, pp. 616-633.

Levine, J. M., & R. S. Tindale (2015). Social influence in groups. In M. Mikulincer, P. R. Shaver, J. F. Dovidio, & J. A. Simpson, eds., *APA Handbook of Personality and Social Psychology, Volume 2: Group Processes*. Washington, DC: American Psychological Association, pp. 3-34.

Lijphart, A. (2012). *Patterns of Democracy: Government Forms and Performance in Thirty-six Countries*. Yale University Press.

List, C., & Goodin, R. E. (2001). Epistemic democracy: generalizing the Condorcet jury theorem. *Journal of Political Philosophy*, *9*(3), 277-306.

Miller, N. R. (1986). Information, electorates, and democracy: Some extensions and interpretations of the Condorcet Jury Theorem. In B. Grofman & G. Owens, eds., *Information Pooling and Group decision Making: Proceedings of the Second University of California, Irvine Conference on Political Economy*. JAI Press, Greenwich, pp. 173-192.

Niemeyer, Simon. (2011). The emancipatory effect of deliberation: Empirical lessons from mini-publics. *Politics & Society,* 39, 103-140.

Page, Scott E. (2007). *The Difference*. Princeton, NJ: Princeton University Press.

Page, Scott E. (2011). *Diversity and Complexity*. Princeton, NJ: Princeton University Press.

Pitkin, Hanna F. (1967). *The Concept of Representation*. Berkeley, CA: University of California Press.

Rousseau, J-J. (1762, 1978). *The Social Contract*. Trans. J. Masters, R. Masters. New York: St. Martins.

Schwartzberg, Melissa (2015). Epistemic democracy and its challenges. *Annual Review of Political Science*, 18(1), 187-203.

Stasser, G. and Dietz-Uhler, B. (2001). Collective choice, judgment, and problem solving. In M. A. Hogg & R. S Tindale, eds., *Blackwell Handbook of Social Psychology: Group Processes*. Oxford: Blackwell, pp. 31-55.

Thompson, Abigail (2014). Does diversity trump ability?: An example of the misuse of mathematics in the social sciences. *Notices of the American Mathematical Society*, 61, 1024-1030.

Tindale, R. S., J. Nadler, A. Krebel, & J. H. Davis (2001). Procedural mechanisms and jury behavior. In M. A. Hogg & R. S Tindale, eds., *Blackwell Handbook of Social Psychology: Group Processes*. Oxford: Blackwell, pp 574-602.

Urbinati, Nadia & Mark E. Warren (2008). The Concept of Representation in Contemporary Democratic Theory. *Annual Review of Political Science*, 11, 387-412.

Warren, Mark E. & H. Pearse (2008). Eds, *Designing Deliberative Democracy: The British Columbia Citizens' Assembly*. New York: Cambridge Univ. Press.

Witte, Erich & James H. Davis (2014). Eds, *Understanding Group Behavior: Consensual Action by Small Groups*, vol. 1. New York: Psychology Press.

---

[1] For a notable exception see Landemore (2013).

[2] Goodin & Spiekermann (2012) offer some related calculations for differently structured representative bodies.

[3] The formula for probabilities in the general case of g groups of m individuals, with both g and m odd, is:

$$r = \sum_{k=(g+1)/2}^{g} \binom{g}{k} q^k (1-q)^{g-k} \quad \text{where} \quad q = \sum_{k=(m+1)/2}^{m} \binom{m}{k} p^k (1-p)^{m-k}$$

(q is the probability of a group getting the correct answer, itself calculated for each group in terms of individual probabilities p.) For even numbers, the result holds with either a calculation altered to break ties or requiring a strict (m/2) + 1 correct votes.

[4] Although we don't pursue the degree of difference with different numbers of groups and members, it is worthy of note that Beisbart and Bovens (2013) find mean majority deficit results to be greatest in those cases in which $g = m$.

[5] These numbers were obtained by simulation rather than by analysis, so they are only given approximately.

[6] The Kolmogorov-Smirnov two-sample test was chosen because it is a non-parametric test to determine whether two samples might be pulled from the same distribution and our observed outcomes did not match any parametric distribution.

[7] Landemore (2013) says bluntly "all-inclusiveness is simply not feasible, and … representative democracy so far remains the only option for our mass societies…" (p. 90).

[8] Primary author on the appendix is Scott Page.