Diversity, Ability, and Expertise in Epistemic Communities

Patrick Grim, Daniel J. Singer, Aaron Bramson, Bennett Holman, Sean McGeehan, and William J. Berger*

The Hong and Page 'diversity trumps ability' result has been used to argue for the more general claim that a diverse set of agents is epistemically superior to a comparable group of experts. Here we extend Hong and Page's model to landscapes of different degrees of randomness and demonstrate the sensitivity of the 'diversity trumps ability' result. This analysis offers a more nuanced picture of how diversity, ability, and expertise may relate. Although models of this sort can indeed be suggestive for diversity policies, we advise against interpreting such results overly broadly.

A number of computational models of group inquiry have garnered widespread attention (e.g., Zollman 2007; Grim 2009; Alexander 2013). Within a model of this type, Hong and Page (2004) demonstrate a 'diversity trumps ability' (DTA) result: for a group of agents confronting a task, "a randomly selected collection of problem solvers outperforms a collection of the best individual problem solvers" (Page 2007, 162). Because this result suggests that an organization is epistemically better off by recruiting a diverse set of candidates instead of just selecting the best individual performers, the Hong-Page result has been taken to have direct implications for policy. For example, claims that the Hong-Page result provides evidence that increasing underrepresented groups will result in disciplinary gains have been made in computer science (Cheryan et al. 2013), biomedical engineering (Chesler et al. 2010), biomedical science (Gibbs et al. 2014; Pickett et al. 2015), and science, technology, engineering, and mathematics (STEM) fields generally (Ferrini-Mundy 2013; Handley et al. 2015). The results have been cited in support of specific scientific practices, such as developing programs to erad-

Received April 2017; revised February 2018.

* To contact the authors, please write to: Bennett Holman, Room 430, Veritas B Hall, Yonsei University, 85 Songdogwahak-ro Yeonsu-gu, Incheon 21983, South Korea; e-mail: bholman@yonsei.ac.kr.

Philosophy of Science, 86 (January 2019) pp. 98–123. 0031-8248/2019/8601-0005$10.00
Copyright 2019 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/701070 Published online by Cambridge University Press

icate tuberculosis (Quissell and Walt 2015). The work has been presented to NASA, cited by the US Geological Survey, is one of four works cited in support of positive expected institutional effects of UCLA's (2014) diversity requirement, and has recently been appealed to in support of promoting diversity in the armed forces in a brief submitted to the Supreme Court (*Fisher v. University of Texas at Austin*, 579 U.S. ___ [2016]; Thompson 2014). A number of philosophers have also taken the result to apply to the value of epistemic diversity in scientific communities (e.g., Martini 2014; Stegenga 2016; Bright 2017).

In reviews and applications of their work, it is perhaps natural that the Hong-Page results about ability are taken as part of a larger case against expertise (Landemore 2013; Gunn 2014; Weymark 2015). Princeton University Press's book blurb characterizes Page's *The Difference* as revealing "how groups that display a range of perspectives outperform groups of like-minded experts" (https://press.princeton.edu/titles/8757.html). Elizabeth Anderson characterizes Hong and Page as showing "that diverse collections of nonexperts do a better job than experts in solving many problems," supporting the claim that "democracy, which allows everyone to have a hand in collective problem solving is epistemically superior to technocracy, or rule by experts" (2006, 12).

We regard the interpretational employment of the Hong-Page model as a general argument for diversity against expertise as a mistake. In what follows we replicate and then build on their work in order to evaluate claims about diversity and ability using a richer notion of expertise than in the original: in our view, genuine expertise requires being able to perform well on many problems of the same type, not just on a single problem.

There is an important difference between the careful way that the Hong-Page result is discussed by those who are sensitive to the details of the model and by others who have interpreted and applied it in a broad range of social contexts. Mayo-Wilson, Zollman, and Danks (2011) responsibly cite the result in support of the "independence thesis"—the claim that the properties of an epistemic community can differ from the properties of their agents. Here the claim is simply that the best epistemic groups are not necessarily composed of the individually highest-performing agents. Such a claim is right in line with the original conclusion that, under the right conditions, "a diverse group can often outperform a group of the best" (Hong and Page 2004, 16386). Other modelers have also been careful to retain the qualified claim that diversity *can* trump ability (Zollman 2011, 2013; Bright 2017).

But the result has also been cited in support of a much stronger claim: that diversity is generally (or even necessarily) epistemically beneficial. Nunn (2012) relies heavily on the result to argue that the medical community would be better off if it moved away from evidence-based medicine and incorporated a plurality of "medical models" (e.g., narrative medicine, evolutionary

https://doi.org/10.1086/701070 Published online by Cambridge University Press

medicine, and complexity medicine). "It is not an a priori assumption or mere hand-waving optimism to say that people working with many models and their associated methods do better than those working with only a few models and methods. Rather, it is a claim that . . . is grounded in the modeling experiments and theorems of Hong and Page" (Nunn 2012, 976). Our results show that Hong and Page's result does not support strong conclusions like these.

We are not the first to critique the large literature on the Hong and Page result. Thompson (2014) gives an influential critique of their original paper, but responses to Thompson's critique show that one part of it is easy to work around (Page 2015) and that its other parts do not significantly undermine how the result is used by political theorists (Kuehn 2017). One of Thompson's main claims is debunked by showing that there are diverse groups that regularly outperform random groups, a point discussed further below. Pace Thompson, our work supports Hong and Page's original result about the performance of groups of high-performing agents versus diverse groups. But by extending their model to more plausibly include groups of genuine experts, rather than mere high performers, we show that many popular conclusions about the implications of the Hong and Page result for expertise are unwarranted.

In section 1, we introduce and replicate the Hong-Page result, emphasizing its strength within its original parameters. In section 2, however, we argue against interpreting 'best performing agents' in the Hong-Page model as 'experts'. In section 3, we expand the Hong-Page model by adding a parameter that 'smooths' the landscapes to shine light on a specific feature of expertise: portability of ability. In that modified model, measuring expertise by transferability of ability from one problem to others of a similar smoothness, it turns out that groups of experts typically outperform more diverse groups. Within those parameters 'ability trumps diversity'.

In sections 4 and 5, we show that whether diversity or expertise triumphs is affected by other features as well, such as what method groups use to work together and how large a set of problem-solving methods is available. Even within the highly abstract models considered by Hong and Page, although diversity does trump ability in some cases, ability trumps diversity in others. Given that both diversity and ability have their place, unqualified assertions of a general triumph of diversity over ability are unwarranted. In section 6 we survey implications of our results for model construction and model interpretation regarding diversity and expertise.

1. The Hong-Page Result. Hong and Page offer several variations of a formal model of a group working together to solve a problem (Hong and Page 2004; Page 2007, 2011). In these models, agents use heuristics to explore an epistemic landscape. The DTA result is that epistemic outcomes for groups

https://doi.org/10.1086/701070 Published online by Cambridge University Press

of randomly chosen individuals will consistently exceed the performance of groups composed solely of the best performing individuals.[1] In Hong and Page's original model, epistemic exploration proceeds along a looped landscape of 2,000 points (so, e.g., 10 points to the right of 1,995 is point 5). For each of the 2,000 points of the landscape a height is assigned as a random integer between 1 and 100; higher points are interpreted as better answers to the question. Individual agents are identified by a *heuristic*, modeled as an ordered set of k numbers, each of which is a number between 1 and l. We begin, following the Hong and Page original, with ordered sets of three numbers ($k = 3$) between 1 and 12 ($l = 12$). With these parameters, there are 1,320 possible agents defined by distinct heuristics (respecting order but avoiding duplication).

Individuals use their heuristics as follows. An agent starts at, say, location 112 of the 2,000-point landscape, which carries a value (height) of 80. The agent then applies the first number of its heuristic by asking: Does the point that many steps to the right offer a higher value? If so, it moves to that point. If not, it stays put. In either case, it then applies the second number of its heuristic. Does that offer a point with a higher value? If so, it moves to that point. It then uses the third number in the same way. Once the third number has been tried, it starts over with the first number. An individual stops only when none of its numbers can reach a higher point; that is, it has reached its local maximum via applying the cycled heuristic from the initial point of 112. In exploring the landscape in this way, there is a determinant value that the agent reaches starting at each of the 2,000 points. The average of those constitutes an individual's score, which we use to rank our 1,320 agents. The nine 'best' individuals will be those with the nine highest scores.

As a model for discussion within a group, Hong and Page employ a sequential ('relay') activation of the agents. Consider a group of nine participants. Starting from a given point, the first agent uses her heuristic to find the highest point within her reach. Once she has found her maximum reachable height she "passes the baton" to the next agent who begins where she left off. He then searches for a higher maximum by employing his heuristics until his search is exhausted, at which point he passes the baton to the third agent, and so forth, until all nine agents have exhausted their searches. At that point the baton is again passed back to the first agent on the list, and the agents are activated one by one in the same order. The final decision for the group will be the local maximum from which none of the agents can find a higher point. The discussion can be thought of as a conversational relay that proceeds in orderly fashion around a circular table. The score for the group

1. Although accompanied by a mathematical theorem intended to offer partial understanding, the main result comes from simulations rather than a formal proof. As such, we too focus on computational experiments.

https://doi.org/10.1086/701070 Published online by Cambridge University Press

will be the average height achieved using each of the 2,000 locations as starting points.

What Hong and Page compare are the results of a modeled discussion of this form for (*a*) a group composed of individuals with heuristics drawn at random from the heuristics pool at large and (*b*) a group composed of those with the 'best' heuristic sets—those with the highest individual scores. The DTA result states that the random group consistently does better. In our reproduction of the Hong-Page result we compared the scores of (*a*) nine random agents and (*b*) the nine agents with the highest individual scores. We found the average of the maximal heights reached by the nine 'best' individuals over 1,000 different random landscapes to be 92.53 (median 92.67). This is compared to 94.82 (median 94.83) for a group of nine random individuals, indicating an improvement of roughly 2%. We found a higher score achieved by random agents in 97.6% of the 1,000 runs.

We take the primary support for the claim that 'diversity trumps ability' in Hong and Page (2004) to be the 'computational experiment' that is central to their piece. They supplement that simulation with a mathematical theorem offered as an explanation of "the logic behind the simulation results" (16386). When a necessary condition regarding tie-breaking is added (Thompson 2014; Page 2015; Kuehn 2017), what the theorem shows is that given strict conditions regarding group and population size, and specific definitions of problem difficulty and group diversity, diversity trumps ability with probability 1. In both their 'computational experiment' and our replications, the strict conditions required for the theorem are significantly weakened. Here the probability falls below 1, but our replications above indicate that the central result is still very strongly supported.

2. Interpreting Hong and Page: Best Performance, Ability, and Expertise. The Hong-Page result is extremely suggestive, and it has been offered as support for a number of strong conclusions already mentioned. In both their original work and in later applications, Hong and Page allude to diversity as a value in affirmative action (Hong and Page 2004; Page 2007). They also draw conclusions regarding business and research teams: "When selecting a problem-solving team from a diverse population of intelligent agents, a team of randomly selected agents outperforms a team comprised of the best-performing agents" (Hong and Page 2004, 16385). It is to their credit, we think, that Hong and Page tend not to use the term 'experts'. In reviews and applications of their work, however, their results are taken as part of a larger case against expertise (Anderson 2006; Landemore 2013; Gunn 2014; Weymark 2015). As outlined below, however, there is an important characteristic of expertise that is not captured in the original model. Results differ in significant and interesting ways when the model is expanded to better reflect that aspect of expertise.

https://doi.org/10.1086/701070 Published online by Cambridge University Press

Following Hong and Page (2004; as well as Page 2007, 2011), we think of a landscape as representing a specific problem or question; for example, which of this year's cars has the best gas mileage (Page 2007), or what is the best combination of drugs to treat a particular illness? Importantly, Hong and Page model these as completely random landscapes, where there is no correlation between the heights of any positions on the landscapes and any others. Because of that, different landscapes yield 'best performing individuals' with very different heuristics, and an individual who is 'best performing' on one random landscape is likely to do extremely poorly on another. No matter how linked two problems or the techniques for solving them may be (e.g., calculating gas mileage not merely for 2016 models but for 2017 and 2018 as well), by modeling those problems as distinct random landscapes, best performing heuristics cannot be expected to carry over from one problem to its close relatives. This spells trouble for interpreting the DTA result as applying to experts.

Table 1 shows the top nine heuristic sets in 10 different model runs on different random landscapes. Notice that there is frequent redundancy of heuristic numbers among the 'best performing' agents on each landscape. On the first landscape, for example, the numbers 4 and 12 appear in every one of the 'best performing' heuristic sets. The redundancy of the 'best performing' set is a major part of Hong and Page's own analysis of both the formal results and social implications: why hire five individuals with the same background if you will just hear the same message five times? But there is clearly something arbitrary about the numbers that show up as part of a successful heuristic. While the numbers 12 and 4 appear in all of the best performing heuristics for the first landscape in table 1, neither number appears in any of the best performing heuristics for the second or third landscapes.

TABLE 1. Heuristic Sets for the 'Best Performing' Agents on 10 Different Fully Random Landscapes

Landscape	Heuristic Sets
1	(12 4 5), (12 2 4), (12 5 4), (12 4 2), (5 12 4), (4 12 2), (6 12 4), (4 5 12), (12 4 6)
2	(5 7 6), (10 8 7), (8 7 10), (7 10 8), (7 5 6), (7 8 10), (11 10 8), (5 6 7), (10 11 8)
3	(1 10 3), (1 6 2), (1 3 10), (3 1 10), (6 2 1), (10 3 1), (10 1 3), (1 10 6), (7 5 3)
4	(11 4 1), (12 2 8), (11 2 12), (4 11 1), (11 1 4), (4 1 11), (12 11 2), (5 8 2), (8 12 2)
5	(6 1 2), (3 6 1), (6 1 3), (1 2 7), (3 6 2), (1 3 6), (2 6 7), (7 1 2), (1 2 6)
6	(4 8 7), (3 4 8), (4 8 3), (7 4 8), (4 3 8), (1 8 7), (3 8 4), (3 8 7), (8 7 2)
7	(3 12 1), (1 3 12), (12 1 3), (3 1 12), (8 3 12), (11 12 8), (1 8 12), (12 1 8), (12 3 1)
8	(2 6 11), (11 2 6), (6 11 2), (11 6 2), (6 2 11), (9 6 11), (2 11 6), (11 9 6), (11 6 9)
9	(8 7 2), (8 2 7), (2 7 8), (8 6 7), (6 8 7), (7 6 4), (6 7 8), (7 8 6), (2 8 7)
10	(2 8 3), (8 3 2), (12 11 3), (3 12 11), (12 3 11), (11 3 12), (2 3 8), (11 12 10), (12 11 10)

https://doi.org/10.1086/701070 Published online by Cambridge University Press

Table 2 shows the percentage of cases in which each of the 12 heuristic numbers appears among the three heuristic numbers of the nine 'best performing' agents on 100 random landscapes. Importantly, each heuristic number shows roughly equal representation across the random landscapes as a whole. What this shows is that, although certain heuristics perform best on individual random landscapes, the fact that a heuristic performs well on one random landscape tells us nothing about how it will perform on another random landscape. In the original Hong-Page model, the 'best performing' on a specific landscape might therefore be better thought of as the 'luckiest' on the landscape: those that happen to have heuristic sets attuned to that specific case.

Agents who get things right only by luck hardly qualify as experts. A minimum requirement of genuine expertise is that experts can be expected to perform well on new and different questions in their field of expertise. For example, experts at judging car fuel efficiency should be expected to do well not only on cars produced in 2016 but also on those produced in 2015 and 2017, since the same methods used to get a good estimate in one year should work to find a good estimate in others years. Much of the literature on relying on experts presupposes this minimal criterion (e.g., Hardwig 1985; Elga 2007). Goldman argues explicitly for it: expertise, he tells us, "includes a capacity or disposition to deploy or exploit [a] fund of information to form beliefs in true answers to new questions that may be posed in the domain" (2001, 91).

The same point applies to one natural conception of 'ability'. While there may be a weak notion of ability whereby one has ability when one can succeed at just a single instance of a task, a more natural conception of ability treats the needed success as counterfactually robust or transportable: one has an ability to Φ only if one is likely to succeed at that task under a range of conditions (although, of course, the breadth of that range is context dependent). Someone who has the ability to judge the quality of livestock should be able to give us reliable results across multiple herds, for example. An abil-

TABLE 2. Percentage of Cases in Which Each Value Appears among the Three Heuristic Numbers of the Nine 'Best Performers' on 100 Random Landscapes

Heuristic No.	%	Heuristic No.	%
1	22.0	7	29.7
2	20.0	8	25.6
3	24.5	9	29.1
4	20.4	10	25.8
5	25.5	11	29.2
6	21.3	12	26.9

https://doi.org/10.1086/701070 Published online by Cambridge University Press

ity to predict the weather requires being disposed toward accuracy in more than one case.

We suggest that Hong and Page's best performing agents on random landscapes do not meet these requirements for expertise and ought not to be considered to have a general ability in the transportable sense. Below, we use the term 'ability' as Hong and Page do, signaling high performance alone, whether transportable or not. We restrict the term 'expert' to refer to only those whose ability is transportable. In that terminology, our replication of Hong and Page's simulations shows that their formal result is secure on random landscapes: there, diversity trumps ability. That does not entail, we will show, that diversity trumps ability on other landscapes or that diversity trumps expertise.

Below, we introduce variations in which landscapes possess regularity or informational content beyond that of purely random landscapes. We consider "smoothed" landscapes with correlation between values at successive points. We then track the robustness of the Hong-Page result across such differences in epistemic landscapes, finding areas in which the DTA results appears to be reversed.

In line with the discussion above, we then model expertise as the transportability of success from one landscape of a given smoothness to another. 'Expertise' appears when there are individuals whose performance on one landscape can predict similar performance on another of a similar character. The results of the next section outline a case in which this sense of expertise appears only when diverse groups lose their advantage over groups of high-performing agents.

3. Ability over Diversity on Smoother Epistemic Landscapes. On the purely random landscape used in the original Hong-Page model, nearby points on the landscape are uncorrelated. Although locations closer to each other on a landscape are more likely to be within the reach of one of the heuristic numbers of an agent, the assigned values or 'answers' at proximate points may or may not have similar heights. If we introduce correlation between the heights of nearby points, we create 'smoother' landscapes. In this section, we explore the robustness of the DTA result on such smoother landscapes.

The interpretation of these smoother landscapes is fairly natural if we conceive of the heuristics as investigatory strategies. Some landscapes allow for strategies that can 'hill climb', using heuristics that improve incrementally from one answer to a superior one at a neighboring point. For these problems a proper methodology allows a sense of incremental progress. Other problems offer essentially no advantage to hill climbing: a move to a nearby solution is just as likely to yield progress as moving to a completely remote part of the landscape. It is this second kind of question, which we might call "strategy

https://doi.org/10.1086/701070 Published online by Cambridge University Press

resistant," that is best represented by the random landscapes of Hong and Page's original model. Here we expand the original Hong and Page model by introducing a parameter for the correlation of location and height (smoothness) and vary this parameter in order to explore the relative success of groups of diverse versus best performing agents on the wider class of problems represented by smoother landscapes.[2]

One way of smoothing a random landscape is to interpolate values between a number of randomly set points. For example, instead of assigning a random value to each of 2,000 points, we could set the height of roughly every third location and fill in the value of the intermediate points to create straight lines between the randomly set spots. The result would be a landscape that is less random and less rugged than the original. For a still smoother landscape, we might assign random values to every fifth point, or every tenth point, and fill in the gaps by drawing descending or ascending lines between the assigned points.[3]

To construct landscapes like this, we assign a random height value to point 1. For a *smoothing factor* of x, we pick a random integer between 1 and $2x$ and assign a random height value to that point. The locations of our assigned points therefore have an average distance of x.[4] Points between the assigned ones are positioned on a line of ascending or descending values between them (rounding the heights to integer values). Example epistemic landscapes with smoothing factors of 0, 5, 10, and 20 are shown up to 200 positions in figure 1 to illustrate the effect.

We now ask how robust the DTA result is with increasing landscape smoothness. As before, our agent heuristics consist of ordered sets of three

2. Smooth landscapes are by no means the only variation worth studying. Many problems, for example, including many problems in science, might be better modeled using NK landscapes (Alexander, Himmelreich, and Thompson 2015; Fontanari and Rodrigues 2016).

3. Hong and Page emphasize a number of conditions on their result, one of which is that the problem to be solved is 'difficult'. Their specification for a 'difficult' problem is that there be no individual problem solver who always finds the global maximum (Page 2007, 159). All the landscapes employed in our models count as difficult in this sense. In the conditions outlined for the theorem they offer in explanation of their results, Hong and Page (2004) specify as a third condition that agents are "diverse" in the sense that for any point that is not the global maximum, there exists an agent who can improve the solution. For many landscapes, given the limits of 'reach' defined by a largest heuristic number, the models employed in our simulations will not satisfy that third condition. That is also true of the simulation models in Hong and Page's original, however. Our simulations are patterned directly on theirs, with agents who will be diverse in every sense that theirs are.

4. The use of randomized interval spacing rather than fixed-length intervals avoids potential artifacts in the results, such as those heuristics that evenly divide a fixed-length smoothing factor becoming overrepresented.

https://doi.org/10.1086/701070 Published online by Cambridge University Press

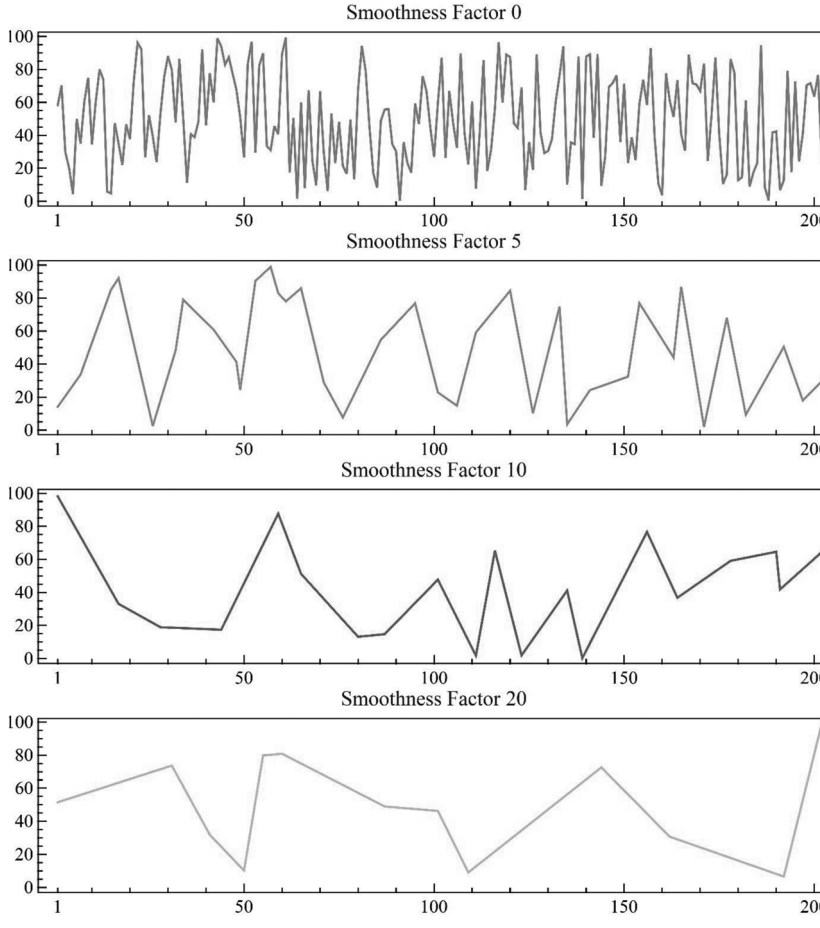


Figure 1. First 200 points of sample landscapes of length 2,000, created with smoothness factors of 0, 5, 10, and 20. Color version available as an online enhancement.

numbers between 1 and 12, resulting in 1,320 possibilities. Over 100 distinct landscapes, we average the values of the final heights reached when starting from each of 2,000 points for (*a*) a relay group of nine random individuals and (*b*) a relay group of the nine individuals who perform best individually. For landscape smoothness factors from 0 to 20, figure 2 plots the difference in performance (random group scores minus best performing group scores). The crossover point at a smoothing factor of 4 indicates the point at which the DTA result no longer holds. Below this value, random groups outperform groups of the highest performing individuals. In Hong and Page's

https://doi.org/10.1086/701070 Published online by Cambridge University Press

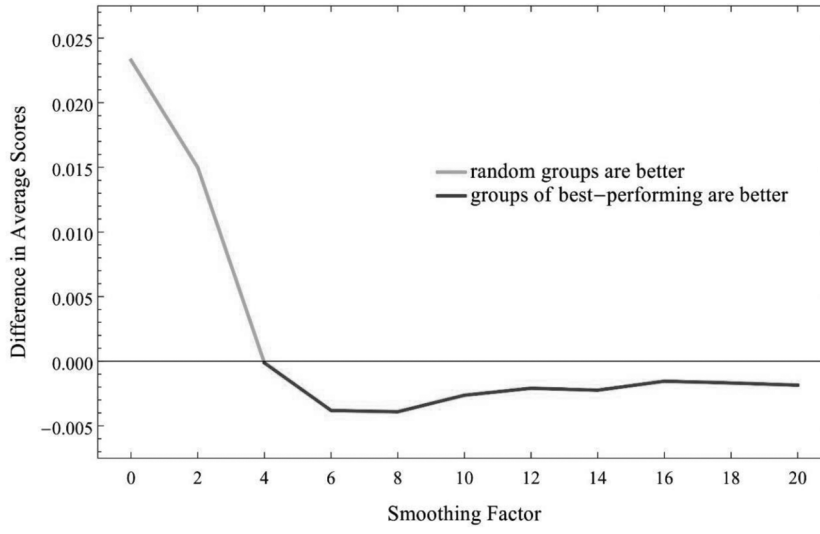


Figure 2. Differences in average performance over 100 different landscapes for groups of nine individuals using three heuristics from a pool of 12. Color version available as an online enhancement.

terms, 'diversity trumps ability'. Above that value, it is groups of the highest performers who do better. Here 'ability' trumps diversity.

When groups of the highest performing individuals do better, the advantage is small: at a smoothness factor of 6, for example, the average performance over 100 landscapes is 0.756 and 0.760 for random and 'best' groups, respectively.[5] That small advantage of the best over the random is, however, clear and robust beyond the crossover value of 4.

With these tools in hand we return to the issue of ability and expertise. If we think of a landscape as representing a particular question within a particular discipline or subject matter, for the best performing agents to be considered to be experts, we would expect to see their skill as transportable: we would expect them to perform roughly as well on other related questions.[6]

5. Performance of each is significantly lower than that reported for a purely random landscape in sec. 1, despite the fact that a smoother landscape might be thought to represent a less 'difficult' problem. Increasing 'smoothness' encourages hill climbing, but increasing smoothness with a fixed maximum of heuristic numbers can also create valleys more 'difficult' to cross. For that reason we have resisted interpreting 'smoother' landscapes as modeling 'less difficult' problems in any simple sense

6. Thinking of alternative landscapes as different specific questions within the same general problem space follows Page (2007, 2011). An interpretation that demands a content distinction between sets of questions in different fields would require significantly more complex modeling assumptions.

https://doi.org/10.1086/701070 Published online by Cambridge University Press

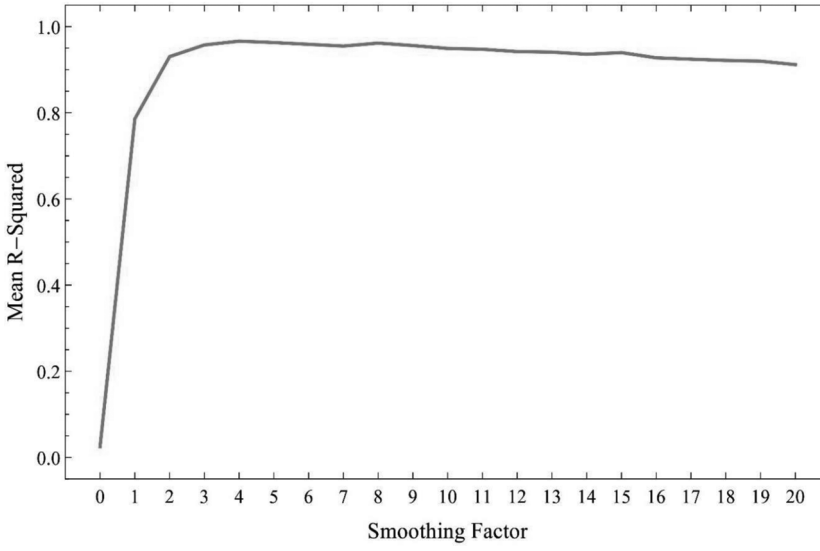


Figure 3. Average percentage of variance of performance of an agent on paired landscapes of the same smoothness that is explained by individual heuristics. Color version available as an online enhancement.

Smoother landscapes do exhibit this form of transportability of best performing heuristics. Interpreting 'best performing' as reflecting expertise thereby becomes more plausible on smoother landscapes than on the random landscapes of Hong and Page's original model.

Here is how we can tell: we generated pairs of landscapes of equal smoothness and found the Pearson correlation of the performance of each of the 1,320 agents on those landscapes. This process was repeated 100 times to obtain the average value for a given smoothness factor, and the entire process was repeated for each smoothness factor from 0 to 20. The square of the Pearson correlation (R^2) tells us what percentage of the variance in performance is explained by an individual's heuristic and thus to what extent performance on one landscape predicts performance on the next (fig. 3).

Figure 3 shows a clear and sudden initial jump in the correlation of individual performance on one landscape to another of the same smoothness. It is only at and after the jump, we propose, that interpreting results in terms of 'expertise' becomes plausible, since it is only then that one's performance on one landscape is correlated with one's performance on another.[7] In this

7. One might reject our understanding of expertise in this model and instead think that ability needs only be transportable to different places in the same landscape to count as expertise. The same story plays out if we compare intralandscape performance, however. We

https://doi.org/10.1086/701070 Published online by Cambridge University Press

case, we see the highest level of transferability occurs roughly when it is no longer true that diverse groups are best performing. Moreover, as the figures above suggest, in this case, viewing the highest performing agents as experts becomes more plausible as the relative strength of the random groups decreases.[8]

By looking at who the experts are in these smoothed landscapes, we can also say something more about what is involved in the expertise modeled. Recall table 2 showing the percentage of cases in which each of our 12 numbers appears among the three heuristics of the nine 'best performing' agents on 100 landscapes. In that case there appears to be no clear signature of the transportability required for general 'ability' to exist: each heuristic number occurs roughly equally across the random landscapes. Table 3 expands table 2 to show the percentage of cases in which each of our 12 numbers appears among the three heuristic numbers of the nine 'best performing' agents on 100 landscapes with increasing smoothing factors.

While a purely random landscape (smoothness 0) shows no consistent bias toward any specific heuristic numbers, a pattern immediately emerges at smoothing factor 1; specifically, the heuristic 1 appears among all of the best nine heuristics in nearly all cases. The largest heuristic number, 12, becomes increasingly prevalent as the smoothing factor increases, as do middle numbers initially around 7 but moving progressively toward shorter steps. For example, the number 2 disappears entirely at smoothing factor 2, joined by the number 3 and then 4 by smoothness 3, as the heuristic pool focuses on 6, 7, and 8. The upper numbers remain rare with further increases in the smoothing factor, but the lower numbers recover their prevalence as the set of weakly selected middle numbers shifts to the left. We also see a gradual decrease in the prevalence of the heuristic 1 because these other low numbers also work as sufficient hill climbers on smoother landscapes. Unlike the random landscape, it is clear that there are certain patterns of heuristics in all these cases—the 'expert sets'—that tend to do the best quite generally across most landscapes of a particular smoothness.

There are a number of at least partial explanations for why we see this correlation between heuristic value and performance on these smoothing factors. The heuristic 1 is valuable because it is the ultimate hill climber: should other numbers in rotation not interfere, repeated access to '1' alone would allow a heuristic to climb to the highest point on any incline to reach a local

tested the average percentage of variance of performance of an agent on the first half of a landscape and the second half of the same landscape that can explained by individual heuristics, and the graph was virtually identical to the one described above. Just as with cross-landscape comparisons, it showed a clear and sudden initial jump in the correlation of individual performance on both halves of the landscape as smoothness increases.

8. As a generalization, this is, of course, subject to other parameters considered in the following sections, including, for example, tournament dynamics.

https://doi.org/10.1086/701070 Published online by Cambridge University Press

TABLE 3. Percentage of Cases in Which Each Number Appears among the Three Heuristic Numbers of the Nine 'Best Performers' on 100 Landscapes for Smoothing Factors 0–20

Smoothing Factor	Heuristic Number											
	1	2	3	4	5	6	7	8	9	10	11	12
0	22.0	20.0	24.5	20.4	25.5	21.3	29.7	25.6	29.1	25.8	29.2	26.9
1	99.9	.1	3.4	22.7	19.8	19.2	21.6	19.0	18.8	19.7	23.2	32.6
2	100.0	.0	.0	.1	9.6	30.5	39.3	18.8	3.2	8.9	27.8	61.9
3	100.0	.0	.0	.0	1.8	20.7	45.3	29.0	3.1	.8	14.0	85.2
4	99.6	.4	.0	.6	6.4	22.9	33.6	27.8	8.3	.3	3.6	96.4
5	99.5	.6	1.5	6.2	15.0	22.7	26.9	20.2	6.9	.4	.5	99.5
6	99.3	2.4	6.5	13.4	18.8	20.6	19.5	13.2	5.5	.7	.2	99.9
7	99.2	5.5	11.2	17.2	20.2	18.7	14.6	8.9	3.8	.8	.1	99.9
8	98.8	9.5	16.6	21.4	20.2	14.8	9.8	5.8	2.3	.7	.1	99.9
9	97.9	13.0	19.6	22.9	19.4	12.4	8.2	4.3	1.9	.5	.1	100.0
10	97.3	15.8	21.4	23.0	18.8	12.0	6.9	2.9	1.4	.5	.1	100.0
11	96.4	18.3	22.4	23.8	18.3	10.0	6.0	2.9	1.3	.5	.2	99.9
12	95.8	20.1	24.1	24.2	17.6	9.5	4.6	2.3	1.3	.4	.2	99.9
13	95.5	22.2	25.8	24.8	16.0	8.6	4.0	1.8	.9	.4	.1	100.0
14	95.0	22.7	25.7	25.1	17.3	7.6	3.9	1.8	.6	.2	.1	99.9
15	93.9	23.0	26.3	25.4	17.0	8.1	3.7	1.7	.7	.2	.1	99.9
16	92.2	25.6	26.8	25.3	16.2	7.6	3.4	1.8	.6	.3	.1	99.9
17	91.0	27.8	28.2	24.6	15.4	6.9	3.8	1.5	.6	.2	.2	99.9
18	90.1	27.9	28.2	25.2	15.5	7.0	3.2	1.5	.7	.5	.3	99.8
19	89.4	30.1	28.9	26.0	14.8	6.4	2.6	1.0	.6	.2	.3	99.8
20	88.4	32.0	28.0	24.9	14.8	6.9	2.9	1.2	.4	.2	.3	99.8

maximum. With 1 present, 2 is at best redundant on landscapes with smoothing factor 1, and potentially disruptive—pushing one over the top of a local maximum to a decline on the other side—hence, its total disappearance at smoothing factor 1. This phenomenon also accounts for the disappearance of 2 and 3 given a smoothing factor of 2, and of 2, 3, and 4 given a smoothing factor of 3. The value of 12—the highest number available—is that it offers the best hope of leaping over declines to an incline on the other side of a valley. Because the widths of valleys widen as the smoothing factor increases, this capability becomes increasingly important. Values in the middle are plausibly useful for jumping over more narrow valleys. These may be weakly selected for as a third number that will put an agent in the top nine once 1 and 12 anchor high performance at the ends. We do not have a full explanation for the specific pattern of middle value shift that we see in the data, however.

Those details aside, the broader lesson of these results is a warning against accepting the DTA result without a qualification regarding the character of the epistemic landscape at issue. Keeping other values in the Hong and Page simulation constant, groups of random agents do better than groups of high-performing individuals only for a narrow range of highly random landscapes.

https://doi.org/10.1086/701070 Published online by Cambridge University Press

For smoother landscapes, those on which successful individuals are more plausibly viewed as experts, it is the groups of high-performing individuals that do better. Given the other assumptions in play, it is ability that trumps diversity when that ability is portable.

4. Diversity over Ability with Larger Heuristic Pools. In the previous section, we showed that with a heuristic pool limited to numbers between 1 and 12, there is a crossover in favor of experts past a smoothness factor of 4. Beyond that point, DTA no longer holds. What happens when the heuristics can be any triplet of numbers from 1 to 16 or from 1 to 20, instead of 1–12 though? What we see is that diversity again shows its strength. We explore that variation on the model here.

Recall that at a smoothness of 8, for example, the best performing do better than a random group when the three numbers of heuristics are chosen from a set of 12 numbers (see fig. 2). When heuristic numbers are chosen from the set 1–24 (or more), however, the group of random heuristics again does better at this smoothness. Figure 4 shows the difference in average score for groups of random heuristics minus the best performing as we increase the size of the heuristic pool for a smoothing factor of 8.

A similar crossover occurs for other smoothing factors, and as we vary the smoothness we have found a very rough 'rule of three' for the value where this

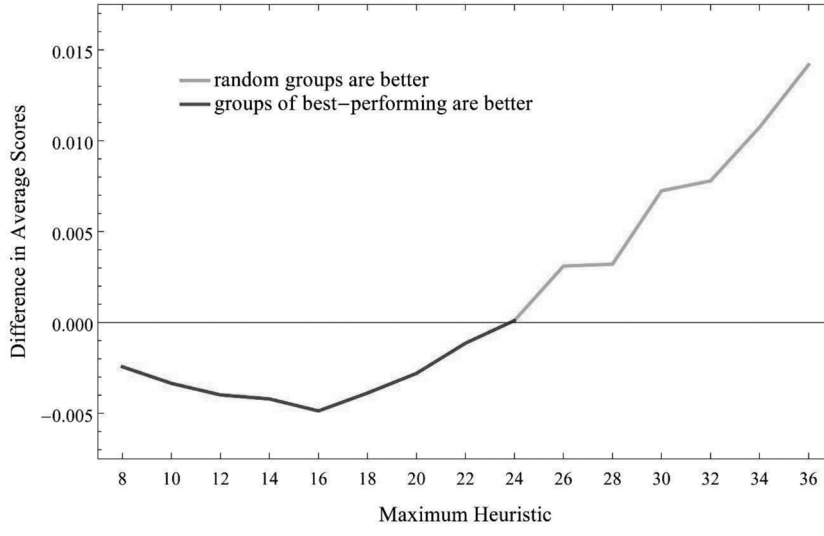


Figure 4. Using a smoothness of 8 as an example, a crossover in favor of random groups occurs when the maximum heuristic number is 24. Differences in averages over 100 landscapes shown. Color version available as an online enhancement.

https://doi.org/10.1086/701070 Published online by Cambridge University Press

crossover occurs: for heuristic pools that are less than three times the smoothing factor of the landscape, the best performing outperform random groups (as outlined in the previous section). For heuristic pools roughly three times the smoothing factor or greater, we once again see a DTA effect. Although increases in landscape smoothness favor groups of the best performing, such an advantage is always relative to the maximum value in the heuristic pool from which strategies are drawn.

The virtues of diversity and 'ability' are therefore relative to the interaction of at least two important factors: landscape smoothness and the heuristic pool. Figures 5–7 show a parameter sweep across both variables, indicating distinct areas of relative strength for diverse groups as opposed to groups of the individually best performing.

Figure 5 presents the data in the roughest form, showing those areas in which the average score for each is greater over 100 landscapes. Figures 6 and 7 show the more nuanced reality behind this result. Even when an average over 100 runs is higher for diversity as opposed to 'ability', the difference may be very slight. Figure 6 shows the same data mapped in terms of the difference in the average scores. Figure 7 shows the percentage of 100 runs in which a random group or group of the best performing does better at each setting of maximum heuristic and landscape smoothness. This demonstrates that even in combinations for which best performer or random groups generally win out, there are still some landscapes for which the other set does better.

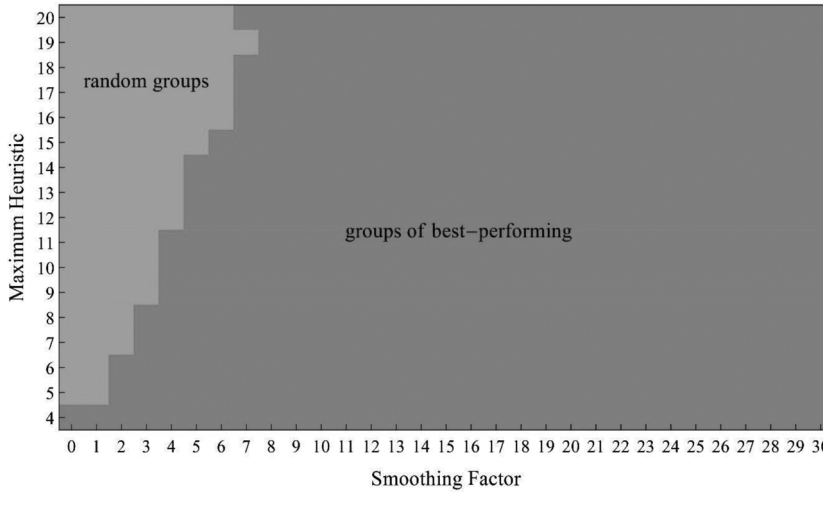


Figure 5. Parameter combinations for which groups of random heuristics do better (*light gray*) and areas in which groups of the best performing perform better (*dark gray*). Color version available as an online enhancement.

https://doi.org/10.1086/701070 Published online by Cambridge University Press

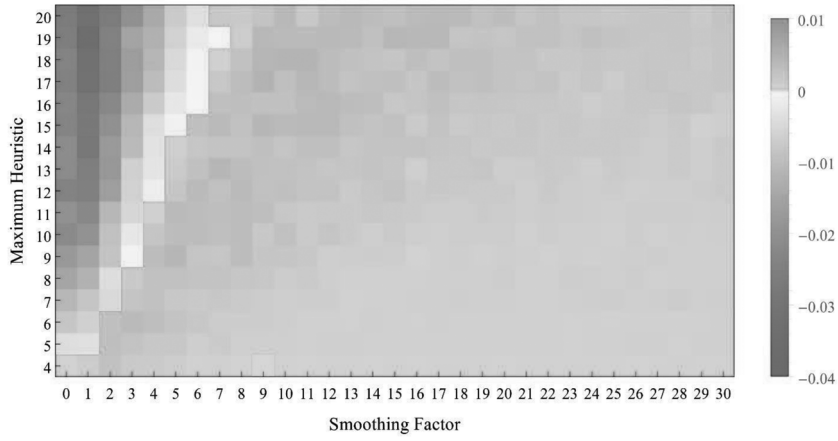


Figure 6. Differences in average scores for groups of random heuristics and groups of the best performing over 100 landscapes at different combinations of smoothing factors and heuristic pools. Positive values (*bottom right*) show higher averages for groups of the best performing. Negative values (*upper left*) show higher averages for groups of random heuristics. Color version available as an online enhancement.

If we think of the heuristic pool as representing the conceptual or problem-solving resources available to tackle the problem, what these results suggest is that there are different niches in which groups of the best performing and diverse groups are each of particular value. Groups of the best performing are better for a wide range of smoothness but only where the available conceptual

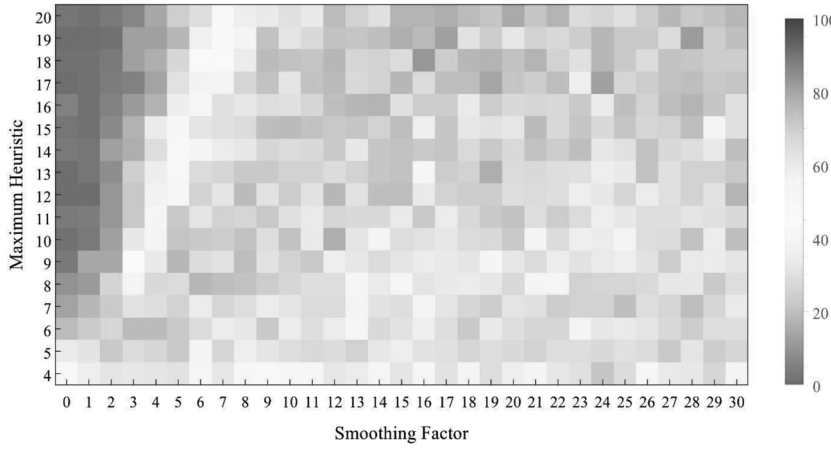


Figure 7. Percentages of runs in which groups of the best performing do better than groups of random heuristics. Color version available as an online enhancement.

https://doi.org/10.1086/701070 Published online by Cambridge University Press

resources are relatively limited. With a wider pool of conceptual resources, a diverse group will do better even on problems of that same character.[9]

One key factor to understanding many of these results, we think, is the extent of heuristic coverage represented in groups. On landscapes of any smoothness, the best performing individuals are very much alike, as indicated in tables 1 and 3. So, the group of the best performing will show high redundancy: their collective numbers will cover little of the space of heuristic numbers. Thompson (2014) suggests that it is the random group's randomness that explains their epistemic virtue. As she sees it, the random group's success is an instance of a greater ability of random algorithms to outperform deterministic ones in many applications (1028). We do not think this explanation works, however.

In our tests, deterministically generated groups that maximize their coverage of the heuristic space significantly outperform the random groups. We suggest that it is not the randomness that is an epistemic virtue of random groups but the extent to which their heuristics collectively cover the available space. Among all the members, a group of random strategies will have more numbers to try on average, and so have a prospect of reaching higher peaks by avoiding getting stuck at local maxima.

This supports the claim of Hong and Page in their original work that the greater coverage of diverse groups helps explain the success of groups of random heuristics on random landscapes. Because their measure of diversity (Hong and Page 2004, 16386) does not closely track the coverage of a group's collective heuristic numbers in the space of all heuristic numbers, Thompson is able to show that maximally diverse groups in that sense do not generally beat out random groups (Thompson 2014; Abigail Thompson, pers. comm., 2017). That is not true when maximally diverse groups are thought of as groups that maximally cover the space of heuristic numbers.[10]

Employing coverage, rather than randomness, also helps explain why groups of the best performing do better with smaller heuristic pools but random groups pull ahead with an expanded heuristic pool: in the larger heuristic pool, the percentage of 'expert' numbers is smaller. Even groups of random heuristics have some redundancy, but in a larger heuristic pool that expected redundancy will be smaller. At least one reason why random groups do better with increased heuristic pools than do groups of the best performing seems to be because their coverage of available heuristic numbers increases with a larger pool.

9. As documented in the appendix, we have tested the robustness of these results with groups of size 3, 6, and 9. The smaller the group, the greater the advantage for groups of the best performing. The larger the group, all else being equal, the greater the advantage for groups of random heuristics.

10. For further details, see our discussion of mixed groups in the appendix. Additional exploration of this result as it pertains to debates about the epistemic merits of democracy between Landemore (2013) and Brennan (2017) can be found in Holman et al. (2018). For a discussion of how the result undermines Thompson, see Singer (2019).

https://doi.org/10.1086/701070 Published online by Cambridge University Press

5. Discussion Dynamics: Diversity and Expertise. There is a further factor that surprisingly and dramatically favors diverse groups and is largely ignored in other discussions of the Hong-Page result. The original Hong-Page model uses a 'relay' dynamics, as we have done above. Starting from a given point, the first agent in the group finds the highest point her heuristic will reach. The second agent then starts from that point in search of a higher one and so forth. Once all members of the group have sequentially sought for the highest point from the last point of their predecessor, the baton is passed again to the first agent of the group.

A clear alternative to 'relay' dynamics is a 'tournament' in which all agents of a group simultaneously strive to identify the point on the landscape that earns them the highest value. In tournament dynamics, the point with the highest value that any agent could identify in the first round then becomes the starting point for everyone in the next go-round. What is eliminated in tournament dynamics is the around-the-table sequencing of a relay. Hong and Page consider both dynamics, saying that their results "do not seem to depend on which structure was assumed" (2004, 16386). Yet our results across varied smoothness factors and numbers of available heuristics do depend on which dynamic is used. Specifically, the use of tournament dynamics generally favors diverse groups over relay dynamics. In comparison with relay results in figures 5–7, figures 8–10 show the corresponding results for a tournament. For a maximum heuristic over 10, 'tournament' rather than 'relay' updating gives a strong advantage to random groups. Indeed, the tournament group dynamic dramat-

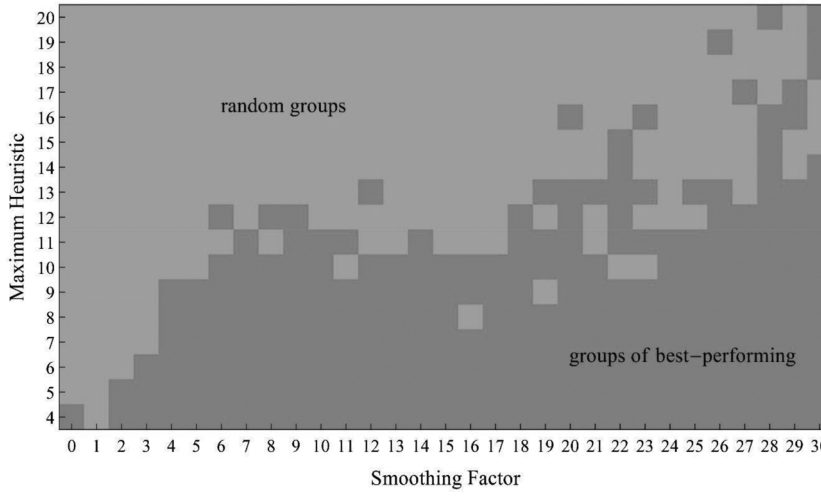


Figure 8. Tournament results corresponding to figure 5's relay results showing parameter combinations in which groups of random (*light gray*) and groups of best performing (*dark gray*) do best. Color version available as an online enhancement.

https://doi.org/10.1086/701070 Published online by Cambridge University Press

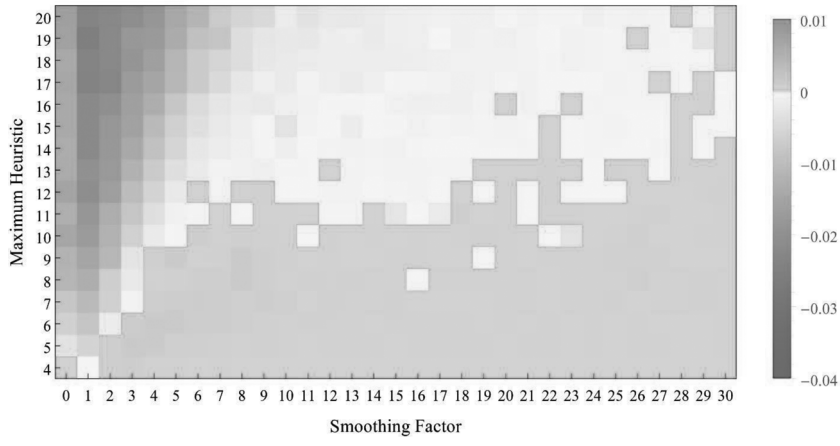


Figure 9. Tournament results corresponding to figure 6's relay results showing differences in averages over 100 landscapes, with positive values (*bottom right*) showing advantage to the best performing and negative values (*upper left*) for random groups. Color version available as an online enhancement.

ically reduces the advantage possessed by groups of the best performing on smoother landscapes with the relay dynamic. Again here, these are areas in which one can more reasonably interpret best performance as expertise.[11] So group dynamics in the form of a tournament make an important difference in the relative value of diversity and expertise; specifically, when there are more available heuristics, the tournament style makes greater coverage of the heuristic pool more valuable than redundant collections of the most generally useful heuristics.

In additional work (available online in an appendix), we explore other variations of the Hong-Page model.[12] As opposed to groups composed exclusively of the best performing agents or random agents, we look at the performance of mixed groups, consisting of both kinds of agent. What we see there is that, in many respects, mixed groups do better than either kind of pure group, but their performance is importantly affected by the group dynamics. We also explore the performance of groups of different sizes. We find that smaller groups are more advantaged by having the best performers, but random agents

11. As documented in the appendix, we have tested the robustness of these results with groups of size 3, 6, and 9. The smaller the group, for tournament as well as relay dynamics, the greater the advantage for groups of the best performing. The larger the group, all else being equal, the greater the advantage for groups of random heuristics.

12. Some elements of our results here are used as a platform for a more general discussion of philosophical interpretation of agent-based models in Holman et al. (2018).

https://doi.org/10.1086/701070 Published online by Cambridge University Press

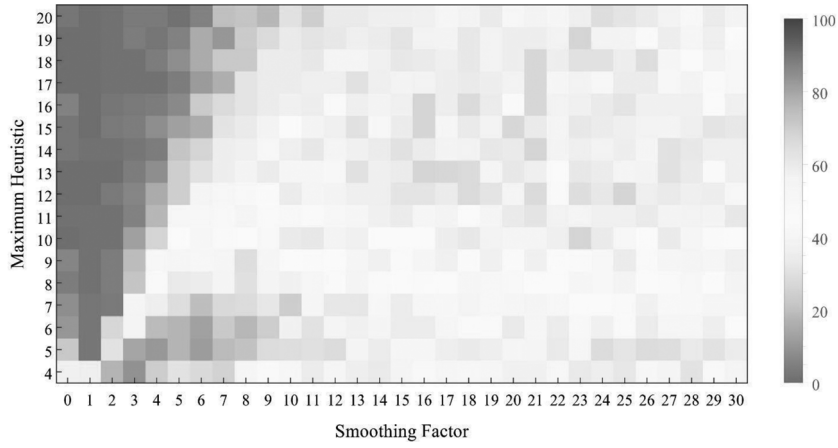


Figure 10. Tournament results corresponding to figure 7's relay results showing percentages of runs in which each group does better, with bottom most values reflecting more wins by best performing and upper values reflecting groups using random heuristics. Color version available as an online enhancement.

do best for larger groups. As above, this is naturally explained by the amount of heuristic coverage offered by the agents in the group.

6. On the Broader Implications of These Results. Despite the accolades and applications of Hong and Page's result, both its validity and soundness have been critiqued, most prominently by Thompson (2014), echoed in Brennan (2017). Responses appear in Megginson (2015), Page (2015), and Kuehn (2017), with a deeper critique ongoing.[13] We think our results start to address that continuing controversy, but they also take us further by emphasizing questions of diversity and expertise that demand a richer model. In this section we offer some reflections on Thompson and review the ways in which our results allow us to distinguish the types of claims that the Hong-Page result genuinely supports from some of the overstated conclusions that have sometimes been drawn from it. In so doing, we hope to refocus the debate onto what additional research is needed for a better understanding of the epistemic roles of diversity and expertise.

In evaluating the controversy fueled by Thompson, it is important to distinguish the 'computational experiment' of Hong and Page (2004) from the

13. Singer (2019) shows that Thompson's claim that randomness rather than diversity is at issue in the Hong and Page results is unwarranted in light of recent developments in complexity research. Singer goes on to show that on some ways of measuring diversity, diverse groups are indisputably stronger than random groups, which undermines Thompson's argument.

https://doi.org/10.1086/701070 Published online by Cambridge University Press

mathematical theorem they offer as explaining "the logic behind [the] result." We take the core of Hong and Page to be the 'computational experiment', replicated in section 1. Thompson focuses on the theorem, noting the lack of a required tie-breaking condition. She offers a minor correction that leaves the main result standing, however, although Page (2007) had already added an assumption sufficient to deal with the problem.

On the interpretational level, Thompson charges the model with artificiality—identifying the number sets with human abilities, for example. Such a critique could be applied with equal justification against almost any computational model in the social sciences. Rejecting this blanket condemnation, we believe a more nuanced critique saves some of the implications of the Hong-Page result and points the way forward to adjudicate the difficult, yet socially important, empirical questions.

First and foremost, our results show that even at the abstract level of agent-based modeling, whether diverse groups outperform groups of the best performing agents depends on a number of parameters in the model—smoothness, deliberation dynamic, and heuristic pool among them. But despite the parameter sensitivity of the model, nothing we have shown here threatens the more guarded claim for which the Hong-Page result has been cited: diversity *can* trump ability. What this means is that, in some cases, diverse groups do perform better than groups of best performers. This more guarded claim is typical of the use of a number of other modelers (Mayo-Wilson et al. 2011; Zollman 2011, 2013; Bright 2017) and is of significant importance. The Hong-Page model is an early example of what Mayo-Wilson et al. (2011) call the independence thesis, for example, the claim that the properties of an epistemic community can differ from the properties of its agents. It thus aligns with other formal models supporting the independence thesis, including Kitcher's (1993) classic argument that scientists selfishly trying to promote their own careers in a properly structured community may nevertheless maximize the community's chances of discovering the truth. Although we have argued against interpreting the original Hong-Page model in terms of expertise, our results indicate that there are closely related models that more plausibly sustain an interpretation in terms of expertise and in which the DTA result remains.

But, our results indicate that the slogan 'diversity trumps ability' is often overstated. For simple expansions of the original Hong-Page model to a wider variety of landscapes, the DTA result of random groups outperforming groups of the best performing holds only within a small window of low landscape smoothness. Within roughly that same window, moreover, the success of the best performing heuristics on a specific landscape is limited to that specific landscape: success on one random landscape cannot be expected to yield success on another. The suggestion for such models is that 'diversity trumps ability' only in those cases in which it is unclear that best performance should really be considered expertise.

https://doi.org/10.1086/701070 Published online by Cambridge University Press

On smoother landscapes, there is a connection between performance on one landscape and another. Interpreting a set of landscapes as specific questions within a given domain, successful heuristics have a strong claim to be capturing a natural notion of expertise, and it is here that expertise shows its value. For landscape smoothness above 4, using the Hong-Page relay dynamics, groups of the individually best performing agents outperform groups of random agents. With an increase in landscape smoothness, leaving other parameters in place, it is ability that trumps diversity.

Diversity again shows its strength, however, when other parameters are changed. Widening the pool from which heuristic numbers are drawn increases the advantage for random groups. All else being equal, groups of random agents perform better (relative to groups of the best performing) as we increase the size of the heuristic pool.

Contrary to Hong and Page's indication of little difference between the relay dynamics used in their simulation and an alternative 'tournament' dynamics, we find a major difference between the two. In 'tournament' dynamics, agents deliberate and navigate a problem landscape with simultaneous suggestions from the floor rather than in a round-the-table 'relay'. It turns out that a tournament dynamic further favors the value of diversity. Many of the points at which groups of the best performing show an advantage within a relay dynamic disappear in favor of groups of random heuristics once the dynamic is changed to a simultaneous tournament.

Parameter sensitivities across the range of models point to a far more subtle conclusion regarding epistemic phenomena of ability and diversity than has often been drawn from the Hong-Page model. The fact that diversity can trump ability within particular parameters of epistemic landscape, heuristic pool, group size and composition, and deliberation dynamics by no means indicates that diversity will trump ability within other parameters or in general. That diversity trumps ability in some situations clearly does not entail that it will do so in any particular situation, yet this is precisely what the Hong-Page result has been taken to imply by those who cite it in support of a particular policy. For example, whether increased gender diversity will "translate into greater STEM discovery, education, and achievement" (Handley et al. 2015, 13205) is no more supported by Hong and Page (2004) than do the results in section 3 provide evidence that increased diversity would hinder STEM discovery. What is lacking between the formal model and the policy implications—in this as in other cases—is a sustained research program studying the characteristics of the target system and assessing the degree to which important aspects of it are adequately captured by specific models in question (Alexandrova and Northcott 2009).

A research program of such a type, if successful, linking modeling to experiment, to empirical data and policy specifics, would quiet concerns that the model at issue fails to capture the target phenomena (like those brought

https://doi.org/10.1086/701070 Published online by Cambridge University Press

up by Thompson). Whether the Hong-Page model, or extensions explored here, succeeds in that task remains a challenging and open question, but it is one that might reasonably be pursued. Our modeling results indicate that diversity is most beneficial in larger groups, for example, but this is especially true for tournament dynamics. Our results similarly indicate that it is mixed groups of experts and nonexperts, rather than homogeneous groups of either, that will be optimal for specific kinds of problems.[14] Both of these modeling results immediately suggest hypotheses suitable for empirical investigation. That further investigation would be required, however, in order to responsibly support specific policy implementations.

7. Conclusion. The variety and sensitivity of our results show that the pervasive uncritical applications of the Hong-Page result are risky. What our results emphasize is that diversity does not always trump ability. Our attempt has been to take some first steps toward understanding a range of parameters relevant to the interplay between diversity and expertise.

Policy makers across the board must consider the specific character of the problem sets at issue and the decision procedures to be employed. Here, as elsewhere, moving from formal results to real world applications is a long, laborious, and, most importantly, empirical process (Alexandrova and Northcott 2009). Policy makers wishing to assess whether a particular issue might benefit from a more diverse community have a significant amount of additional bridgework to engage in before they can derive support for policies from these modeling results (for details, see Grim et al. 2013). Our results show that diversity is not always epistemically beneficial, even at the model level. Diverse groups and groups of the individually best performing both have a place. Our expanded model of group inquiry reveals a nuanced interplay between them and points toward a greater understanding of the strengths of each.

REFERENCES

Alexander, Jason M. 2013. "Preferential Attachment and the Search for Successful Theories." *Philosophy of Science* 80 (5): 769–82.

Alexander, Jason M., Johannes Himmelreich, and Christopher Thompson. 2015. "Epistemic Landscapes, Optimal Search, and the Division of Cognitive Labor." *Philosophy of Science* 82 (3): 424–53.

Alexandrova, Anna, and Robert Northcott. 2009. "Progress in Economics: Lessons from the Spectrum Auctions." In *Oxford Handbook of Philosophy of Economics*, ed. H. Kincaid and D. Ross, 306–36. Oxford: Oxford University Press.

Anderson, Elizabeth. 2006. "The Epistemology of Democracy." *Episteme* 3 (1): 8–22.

Brennan, Jason. 2017. *Against Democracy*. Princeton, NJ: Princeton University Press.

14. See the appendix for further details, and for the clearest visual representation of this interaction, compare fig. 12 with fig. 15.

https://doi.org/10.1086/701070 Published online by Cambridge University Press

Bright, Liam K. 2017. "Decision Theoretic Model of the Productivity Gap." *Erkenntnis* 82 (2): 421–42.
Cheryan, Sapna, Victoria C. Plaut, Caitlin Handron, and Lauren Hudson. 2013. "The Stereotypical Computer Scientist: Gendered Media Representations as a Barrier to Inclusion for Women." *Sex Roles* 69 (1–2): 58–71.
Chesler, Naomi C., Gilda Barabino, Sangeeta N. Bhatia, and Rebecca Richards-Kortum. 2010. "The Pipeline Still Leaks and More than You Think: A Status Report on Gender Diversity in Biomedical Engineering." *Annals of Biomedical Engineering* 38 (5): 1928–35.
Elga, Adam. 2007. "Reflection and Disagreement." *Noûs* 41 (3): 478–502.
Ferrini-Mundy, Joan. 2013. "Driven by Diversity." *Science* 340 (6130): 278.
Fontanari, José F., and Francisco A. Rodrigues. 2016. "Influence of Network Topology on Cooperative Problem-Solving Systems." *Theory in Biosciences* 135 (3): 101–10.
Gibbs, Kenneth D., Jr., John McGready, Jessica C. Bennett, and Kimberly Griffin. 2014. "Biomedical Science Ph.D. Career Interest Patterns by Race/Ethnicity and Gender." *PloS One* 9 (12): e114736.
Goldman, Alvin I. 2001. "Experts: Which Ones Should You Trust?" *Philosophy and Phenomenological Research* 63 (1): 85–110.
Grim, Patrick. 2009. "Threshold Phenomena in Epistemic Networks." In *Complex Adaptive Systems and the Threshold Effect: Views from the Natural and Social Sciences; Papers from the AAAI Fall Symposium*, ed. Mirsad Hadzikadic and Ted Carmichael, 53–60. Menlo Park, CA: AAAI.
Grim, Patrick, Robert Rosenberger, Adam Rosenfeld, Brian Anderson, and Robert E. Eason. 2013. "How Simulations Fail." *Synthese* 190 (12): 2367–90.
Gunn, Paul. 2014. "Democracy and Epistocracy." *Critical Review* 26:1–2, 59–79.
Handley, Ian M., Elizabeth R. Brown, Corinne A. Moss-Racusin, and Jessi L. Smith. 2015. "Quality of Evidence Revealing Subtle Gender Biases in Science Is in the Eye of the Beholder." *Proceedings of the National Academy of Sciences* 112 (43): 13201–6.
Hardwig, John. 1985. "Epistemic Dependence." *Journal of Philosophy* 82 (7): 335–49.
Holman, Bennett, William Berger, Patrick Grim, and Aaron Bramson. 2018. "Diversity and Democracy: Agent-Based Modeling in Political Philosophy." *Historical Social Research* 43 (1): 259–84.
Hong, Lu, and Scott E. Page. 2004. "Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers." *Proceedings of the National Academy of Sciences* 101:16385–89.
Kitcher, Philip. 1993. *The Advancement of Science: Science without Legend, Objectivity without Illusions*. Oxford: Oxford University Press.
Kuehn, Daniel. 2017. "Diversity, Ability, and Democracy: A Note on Thompson's Challenge to Hong and Page." *Critical Review* 29 (1): 72–87.
Landemore, Helen. 2013. *Democratic Reason: Politics, Collective Intelligence, and the Rules of the Many*. Princeton, NJ: Princeton University Press.
Martini, Carlo. 2014. "Experts in Science: A View from the Trenches." *Synthese* 191 (1): 3–15.
Mayo-Wilson, Connor, Kevin J. Zollman, and David Danks. 2011. "The Independence Thesis: When Individual and Social Epistemology Diverge." *Philosophy of Science* 78 (4): 653–77.
Megginson, Robert. 2015. "A Plea for Civility and Collegiality." *Notices of the American Mathematical Society* 62 (1): 9.
Nunn, Robin. 2012. "Many-Models Medicine: Diversity as the Best Medicine." *Journal of Evaluation in Clinical Practice* 18 (5): 974–78.
Page, Scott E. 2007. *The Difference*. Princeton, NJ: Princeton University Press.
———. 2011. *Diversity and Complexity*. Princeton, NJ: Princeton University Press.
———. 2015. "Diversity Trumps Ability and the Proper Use of Mathematics." *Notices of the AMS* 62 (1): 9–10.
Pickett, Christopher L., Benjamin W. Corb, C. Robert Matthews, Wesley I. Sundquist, and Jeremy M. Berg. 2015. "Toward a Sustainable Biomedical Research Enterprise: Finding Consensus and Implementing Recommendations." *Proceedings of the National Academy of Sciences* 112 (35): 10832–36.
Quissell, Kathryn, and Gill Walt. 2015. "The Challenge of Sustaining Effectiveness over Time: The Case of the Global Network to Stop Tuberculosis." *Health Policy and Planning*. doi:10.1093/heapol/czv035.

https://doi.org/10.1086/701070 Published online by Cambridge University Press

Singer, Daniel J. 2019. "Diversity, not Randomness, Trumps Ability." *Philosophy of Science*, in this issue.
Stegenga, Jacob. 2016. "Three Criteria for Consensus Conferences." *Foundations of Science* 21 (1): 35–49.
Thompson, Abigail. 2014. "Does Diversity Trump Ability? An Example of the Misuse of Mathematics in the Social Sciences." *Notices of the American Mathematical Society* 61:1024–30.
UCLA. 2014. "Proposed Diversity Requirement." College Diversity Initiative Committee. https://ccle.ucla.edu/pluginfile.php/743624/mod_resource/content/6/082014%20REVISED%20DiversityReqProposal.pdf.
Weymark, John A. 2015. "Cognitive Diversity, Binary Decisions, and Epistemic Democracy." *Episteme* 12 (4): 497–511.
Zollman, Kevin. J. 2007. "The Communication Structure of Epistemic Communities." *Philosophy of Science* 74 (5): 574–87.
———. 2011. "Computer Simulation and Emergent Reliability in Science." *Journal of Artificial Societies and Social Simulation* 14 (4): 15.
———. 2013. "Network Epistemology: Communication in Epistemic Communities." *Philosophy Compass* 8 (1): 15–27.

https://doi.org/10.1086/701070 Published online by Cambridge University Press