Diversity, Not Randomness, Trumps Ability*

Daniel J. Singer — singerd@phil.upenn.edu

Penultimate Draft from March 26, 2018
Forthcoming in *Philosophy of Science*. Please cite the published version.

Abstract

A number of formal models, including a highly influential model of group problem solving from Hong and Page (2004), purport to show that diverse groups are often better at solving problems than more homogeneous groups of individually high-performing agents. A natural way to explain the advantage of diverse groups is in terms of their ability to bring more skills, methods of problem solving, background knowledge, perspectives, etc. to the problem than the homogeneous groups. Thompson (2014) argues that in Hong and Page's model, it's actually the fact that the diverse groups are created by a random process that explains their success, not the diversity itself. Here I defend the diversity interpretation of the Hong and Page result, and in doing so, I undercut the concerns created by Thompson's article about institutional, educational, and legal policies that appeal to the epistemic benefits of diverse groups. What the failure of Thompson's argument shows, I argue, is that in order to understand the value of functionally diverse groups, we must be clearer about how we conceive of and measure functional diversity.

Recent research in social epistemology, epistemic democracy, and the social structure of science has lead to a number of formal models of group problem solving that expound the epistemic virtues of diverse groups. The most prominent example of this is a model from Hong and Page (2004), which purports to show that functionally diverse groups of non-experts can be (and often are) more effective than homogeneous groups of experts at solving difficult problems. Hong and Page defend this by showing that groups of randomly-selected individuals often beat groups of the individually best performing agents in finding the peaks of random landscapes. Other notable proponents of the epistemic virtues of diversity include Thoma (2015), who argues that mixed groups of 'explorer' and 'extractor' scientists explore more of a epistemic landscape than either kind of agent alone,[1] and Zollman (2010), who shows that heterogeneous groups can do better by allowing more hypotheses to be considered before a group converges.[2]

Results like these have had significant impacts on discussions about not only functionally diverse groups but also identity diverse groups. The Hong and Page

*I'm very thankful to Scott E. Page, William Z. Berger, Meredith Tamminga, and two anonymous reviewers for comments on an earlier draft of this paper.

[1]Thoma's model is a modification of a model originally presented by Weisberg and Muldoon (2009), who also purport to defend a pro-diversity result. As Thoma argues though, it's not clear that Weisberg and Muldoon successfully defend their claim.

[2]For an important critique of Zollman, see Rosenstock et al. (2017).

result alone has been cited over 3,000 times, appealed to by Landemore (2012) to argue for inclusiveness in deliberative democratic institutions, taken by philosophers of science to demonstrate the epistemic value of diversity in science (e.g. Bright 2017, Martini 2014, Stegenga 2016), cited in support of a diversity requirement by UCLA (2014), and included in a brief to the Supreme Court of the United States supporting promoting diversity in the armed forces (Fisher v. University of Texas, Austin 2016).

If the pro-diversity results are right though, what explains the success of functionally diverse groups? There isn't a universally-accepted answer to this question, but many assume that diverse groups' ability to bring more skills, methods of problem solving, background knowledge, perspectives, etc. to the table must be at least part of the explanation. But, Abigail Thompson (2014) argues that Hong and Page's result cannot be understood that way. What explains the diverse groups' success is that they are created by a random selection process: it's the randomness, not the diversity, that explains the result, she argues. Thompson's arguments have been taken by many to undermine the arguments that appeal to the Hong and Page result, most notably by Brennan (2017, p. 182), who worries that Thompson's result may expose a fatal flaw in Landemore's (2012) defense of democracy.

Here I'll argue that Thompson was mistaken and that the Hong and Page result is explained by diversity, not by randomness. Much of my argument involves a technical exploration of the Hong and Page model and Thompson's critique of it.[3] There's a general lesson in the failure of Thompson's critique though: in order to understand the value of functionally diverse groups, we must be clearer about how we conceive of and measure functional diversity. Moreover, by defending an interpretation of the Hong and Page result in terms of diversity, I undercut the concerns created by Thompson's article about the institutional, educational, and legal policies that were created by appeal to Hong and Page's result. I don't provide a general conception or measure of functional diversity in arguing for my claims here, but my argument will showcase some methods and tools that can be used towards that end in future work.[4]

1 *The "Diversity Trumps Ability" Result and Thompson's Randomness Interpretation*

Hong and Page's (2004) "Diversity Trumps Ability" result shows that on sufficiently hard problems "a randomly selected collection of problem solvers outperforms a collection of the best individual problem solvers" (2007, p. 162). They show this

[3] Kuehn (2017) also canvasses some of the ways proponents of the diversity result can defend themselves while accepting most of Thompson's technical claims (though perhaps saying that they don't apply to the Hong and Page result). Here I argue more directly against Thompson.

[4] Hong and Page are careful to distinguish functional from identity diversity in their discussion. Identity diversity is differences in demographic characteristics, cultural identities, ethnicity, etc., whereas functional diversity is differences in "how people represent problems and how they go about solving them" (2004, p. 16385). I'll restrict the discussion of diversity here to issues of *functional* diversity. Many have taken the virtues of functionally diverse groups to have implications for the virtues of identity diverse groups, but those arguments aren't the focus of this discussion.

both by proving a general mathematical theorem about agents using different search methods and by computational model. Here I'll focus on the model.

In the model, groups of agents work their way around a ring of 2000 spots, which I'll call the 'landscape'. We can think of the different spots in the landscape as being different candidate answers to the question the group is trying to answer. At any given moment, the entire group inhabits a single spot. They can move forward on the ring, and once they reach the end, they loop back to the beginning. Each of the 2000 spots on the ring has a score associated with it. We can think of the score of a spot as being the epistemic payoff of accepting that answer (i.e. a measure of how good the answer is). In Hong and Page's model, each of the spots is randomly assigned a real-number score between 0 and 100.[5]

Each agent has a *heuristic* that they use to move around the ring. A heuristic consists of an ordered non-repeating list of three integers $\{h_1, h_2, h_3\}$. Agents use their heuristic in the following way: from wherever they are on the ring, they ask themselves if the spot h_1 spots ahead has a higher score than their current spot. If so, they move, and if not, they stay. They then try the next number in their heuristic, and repeat the process (looping back to h_1 after trying h_3) until no number in their heuristic takes them to a spot with a higher score. From every starting spot on the ring, there is a unique ending spot the agent will get to from any starting spot. An agent's *competence* will be the average score the agent receives starting from any spot.

A *group* of agents will be an ordered list of agents $\{a_1, a_2, \ldots a_i\}$. Groups of agents move around a ring similarly to individual agents. From any given starting spot, the first agent uses their heuristic to take the group to the highest spot they can from the starting spot. The baton is then passed to the second agent who leads the group from there using their own heuristic. The baton is passed around the group (looping back to a_1 after a_i) until no agent can take the group any higher. A group's *competence* will be the average score the group receives starting from any spot.

What the "Diversity Trumps Ability" result says is that groups of agents with randomly selected distinct heuristics ("random groups") typically outperform groups of agents with the individually-highest-performing distinct heuristics ("expert groups").[6] In my recreation of the model, with three heuristic numbers that range from 1 to 20, 10 agents in each group, and averaged over 1000 runs, expert groups have an average competence of 93.79 (standard error = 0.041) whereas random groups have an average competence of 96.07 (standard error = 0.016). So the random groups performed significantly better (Welch t-test, $p < 2.2 * 10^{-16}$).[7]

Hong and Page explain this result in terms of the random groups being more functionally diverse than the expert groups. To precisify this, Hong and Page provide

[5]For generalizations of this model and implications for the results, see Grim et al. (2017).

[6]Hong and Page are careful to use the term 'best-performing agents' rather than 'experts.' The difference is irrelevant here, but see Grim et al. (2017) for a careful discussion.

[7]Of course, since these data come from simulations, each of the results reported here can be extended to be as statistically significant as is needed.

a measure of functional diversity: Let's say that two heuristics *overlap* in a spot when they share the same number in that spot in the heuristic. For example, $\{1,2,3\}$ and $\{1,4,5\}$ overlap in only the first spot. $\{1,2,3\}$ and $\{4,5,6\}$ do not overlap in any spots. Let $\delta_{HP}(h_1, h_2)$ be the percent of places that the two heuristics h_1 and h_2 do *not* overlap. Then,

HP-DIVERSITY The HP-diversity of a group of agents is the average of all $\delta_{HP}(h_i, h_j)$ where h_i and h_j are heuristics in the group, and $i \neq j$.

HP-diversity essentially measures the lack of overlap of heuristics of members in the group: the higher the HP-diversity, the less overlap there is among the heuristics.

As Hong and Page show, on average, the HP-diversity between expert groups and random groups differs in the same way their average competence differs. Expert groups have an average HP-diversity of 87.19% (standard error = 0.0006) whereas random groups have an average HP-diversity of 95.08% (standard error = 0.0014). Hong and Page conclude then that random groups do better because of their higher diversity. More diverse groups bring more skills to the table (here represented as bringing more non-overlapping heuristics), which explains why "diversity trumps ability" (2004, p. 16385).

2 *Thompson's Critique and Randomness Interpretation*

Thompson (2014) presents a number of worries for Hong and Page's original paper. Chief among them is the worry that diversity can't explain why random groups outperform expert groups. Hong and Page draw their conclusion that diversity explains the performance of random groups by showing that the HP-diversity of random groups is higher than that of expert groups. But a mere correlation of these values isn't a strong enough reason to conclude that there is an explanatory connection, Thompson argues (2014, p. 1028). If it's diversity that is doing the work, then maximally diverse groups should be expected to perform well in general. But, in Thompson's reconstruction of the model, she reports that five distinct groups that were each maximally HP-diverse all performed worse than the median performance of 200 random groups. She concludes that it can't be diversity that's doing the work (2014, p. 1028).

What does Thompson think is going on in the model then? Thompson thinks the success of random groups is best seen as an instance of a well-established theme in algorithms research that random algorithms outperform the best known deterministic algorithms in many situations (2014, p. 1028). As she points out, it's widely accepted that "randomization can improve algorithms, and often can improve them dramatically" (2014, p. 1028). Since the diverse groups are created by a random process in Hong and Page's model, the natural explanation, Thompson suggests, is that the model is best understood as showing that "randomness trumps ability," not that "diversity trumps ability."

3 *The Implausibility of Thompson's Critique*

The bulk of this section will be an argument against Thompson's reason for thinking that maximally diverse groups don't perform well compared to random groups. But note that even if she were right about that, we should still be concerned about the plausibility of the randomness interpretation. Here's why:

Randomized algorithms differ from deterministic (and non-deterministic) algorithms in that randomized algorithms operate on a series of random bits in addition to a regular fixed input. This means that the same random algorithm can give different outputs on the same fixed input. As Thompson notes, randomized algorithms have received much praise in complexity research. The purported advantages of these algorithms are their simplicity (understood in terms of humans' ability to grasp and implement them) and their efficiency over *known* deterministic algorithms.[8]

But, since the issue here is about performance rather than simplicity, we should only think the randomness-based explanation of the Hong and Page result is compelling if random algorithms are thought to have performance benefits over deterministic algorithms in general, not just benefits over *known* deterministic algorithms. Otherwise, Hong and Page could just as easily be right that some fixed property of the randomly-generated groups, like their diversity, is what explains their success, not the fact that they were randomly-generated. It is not among the generally accepted benefits of random algorithms that the randomness itself explains their success though.

The question of whether random algorithms are computationally more powerful than deterministic algorithms is often studied in terms of whether the complexity class of bounded-error probabilistic polynomial time algorithms (BPP) is reducible to the class of polynomial time algorithms (P). That question is still an open one (Greathouse 2013). (For a relevant survey, see Vadhan 2012.) But, current research into derandomization shows that many particular random algorithms can be reduced to deterministic algorithms using techniques like the methods of conditional expectations, pessimistic estimators, and bounded independence. Research on the complexity classes involving random algorithms also shows that many subclasses of random algorithms in BPP can be reduced to deterministic algorithms. In light of those results, most complexity researchers believe that the randomness of random algorithms is not essential to their performance (i.e. that P = BPP) (Aaronson 2017, conjucture 31). So, even if Thompson is correct that HP-diversity doesn't explain the success of random groups, we shouldn't think randomness explains it either. We should look for another property of the random groups that can explain their success.

So what is the property that explains why random groups perform better? Thompson argued that diversity can't be that property because maximally HP-diverse don't

[8]In a seminal paper on this topic, Karp (1991, p. 166), for example, says that random algorithms have two kinds of benefits: "First, often the execution time or space requirement of a randomized algorithm is smaller than that of the best deterministic algorithm *that we know of* for the same problem. But even more strikingly, if we look at the various randomized algorithms that have been invented, we find that invariably they are extremely simple to understand and to implement" (emphasis added).

perform well compared to random groups. I'll show that Thompson was mistaken in asserting that though: maximally HP-diverse groups do typically outperform random groups, and there is a natural explanation of why Thompson thought otherwise.

In arguing that maximally HP-diverse groups don't perform well compared to random groups, Thompson reported testing 5 maximally HP-diverse groups against the median performance of 200 random groups.[9] She found that all of them performed worse than the median of the random groups. My reconstruction of the result further confirms that the groups Thompson tested were in fact very weak compared to random groups. Testing Thompson's groups against 100 random groups on each of 100 landscapes shows that Thompson's groups lose to random groups 93.62% of the time.

So if Thompson's selected groups were representative of maximally HP-diverse groups, it would be reasonable to conclude that maximally HP-diverse groups don't perform better than random groups. But, that's not the result we see. In 1 million runs of randomly generated groups on random landscapes, the average competence of all groups was 96.04 (standard error = 0.0005, n = 1,000,000). The average competence of groups with maximal HP-diversity was 96.33 (standard error = 0.0278, n = 261). So maximally HP-diverse groups, on average, do perform better than random groups (Welch Two Sample t-test, $p < 2.2 * 10^{-16}$).[10] Moreover, when we look at all of the runs, we can see a clear correlation between HP-diversity and competence. Figure 1 shows a scatterplot of HP-diversity and group competence for 50,000 runs that are representative of a 1 million run sample.

We can see from figure 1 that HP-diversity is correlated with competence (Pearson correlation = 0.224, 99.99% CI [0.220, 0.228]), pace Thompson (2014). Moreover, though it's not clear in the scatterplot, there are maximally HP-diverse groups that perform markedly better than random groups, like this one:

```
Super-MaxHP Group = [1,2,3],[4,5,6],[7,8,9],
                    [10,11,12],[13,14,15],[16,17,18],
                    [19,20,1],[2,4,8],[6,10,14],[12,16,20]
```

The Super-MaxHP Group almost never loses to random groups. On 100,000 random landscapes, facing a new random group on each, the Super-MaxHP Group lost only 42 times. Moreover, when it won, it won by 0.897 on average (standard deviation = 0.364), but when it lost, it only lost by 0.041 (standard deviation = 0.038). So the Super-MaxHP is the polar opposite of what Thompson's maximally HP-diverse groups were like.

What happened with Thompson's groups then? Inspecting an example of one of Thompson's groups will be helpful in answering that question. Consider this one (Thompson 2017):

[9] Thompson reports that the groups she tested were *maximally* HP-diverse, though inspection of her code shows that only 3 of the 5 were, in fact, maximially HP-diverse. Groups 1 and 5 had HP-diversity of .993 and .970, respectively (Thompson 2017). This was presumably due to typos in the code, but those don't affect her overall argument.

[10] Unless stated otherwise, the results given are for 1 million runs of groups of size 10 with 3 heuristic numbers, a maximum heuristic of 20, and a ring of length 2000 with real numbered scores in [0,100).

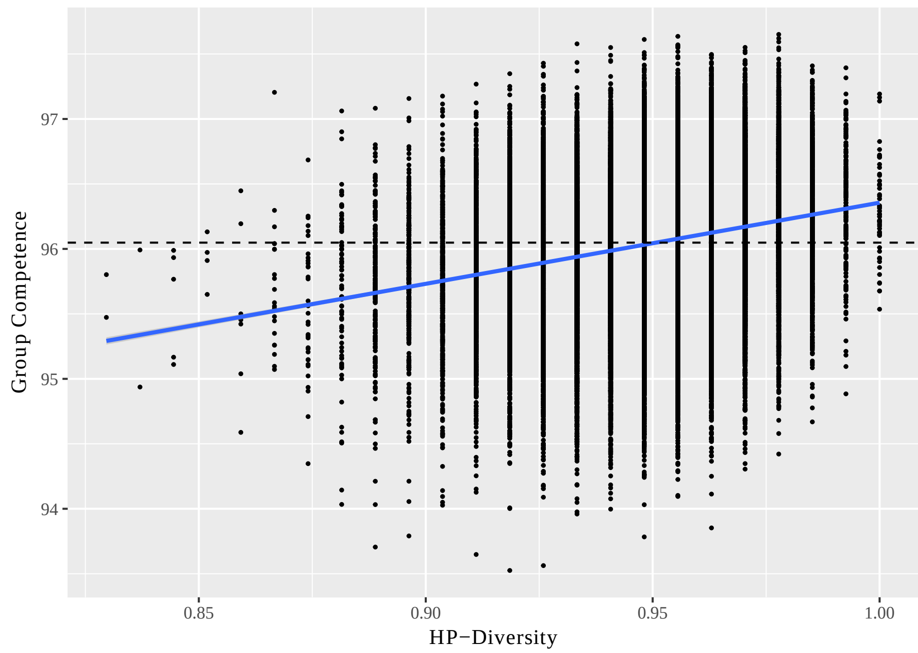

Figure 1: Comparison of HP-Diversity to Competence for 50,000 runs. Includes linear regression (error bars too small to be visible). Average score of all random groups marked by dashed line.

```
Thompson's Group = [1,2,3],[2,3,1],[3,1,2],[4,5,6],[5,6,4],
                   [6,4,5],[7,8,9],[8,9,7],[9,7,8],[10,11,12]
```

It's easy to check that this group is maximally HP-diverse, since no two heuristics have the same heuristic number in the same location. Is this group really diverse (rather than just HP-diverse) though? The maximum heuristic number used in these runs was 20, even though the maximum heuristic in Thompson's group is 12. So forty percent of the possible heuristic numbers don't even show up in this group, even though 9 heuristic numbers are each repeated 3 times. Given that, it seems like there's a sense in which this group isn't very diverse at all. HP-diversity doesn't track this other sense of diversity, the sense which can be measured by the proportion of the heuristic numbers are present in the group. I'll call this sense of diversity 'coverage diversity,' since it measures how much of the heuristic space is covered by the group's combined heuristics.

C-DIVERSITY The C-diversity of a group of agents is the percentage of the heuristic numbers that are represented in any spot in any heuristic in the group.

The example group has a C-diversity of 60%, the minimum possible diversity of a maximally HP-diverse group.

In fact, all of the groups Thompson test have a C-diversity of 60%. C-diversities that low are *very* uncommon among maximally HP-diverse groups at these parameters though. In a test of 1 million random-generated maximally HP-diverse groups, only 0.0131% had a C-diversity less than 0.7, and none had a C-diversity the same as Thompson's (0.6). So it would be *incredibly* unlikely for Thompson to have selected 5 groups with 0.6 C-diversity at random from the set of maximally HP-diverse groups. If one were to construct maximally HP-diverse group by hand though, a natural way to do it would be to avoid overlaps by intentionally repeating the same number in different places in different heuristics in the group. Doing this will result in a maximally HP-diverse group that only uses a small percentage of the available heuristic numbers (since it guarantees repeats of the used numbers). Presumably this is why Thompson's maximally HP-diverse groups were so unrepresentative of the whole class.

So, Thompson's reason for rejecting the diversity interpretation of Hong and Page's result is flawed. Maximally HP-diverse group do generally perform better than random groups. Thompson missed this, I hypothesize, because the groups she presumably made by hand are unrepresentative of the whole class of maximally HP-diverse groups.

4 *It* is *Diversity, Not Randomness*

Once we reject Thompson's argument for thinking that HP-diversity can't explain the random groups' successes and notice that randomness isn't a good option either, we're back in the dialectical position created by Hong and Page in 2004: HP-diversity is the presumptive best candidate to explain the success of random groups. But, why think it's high HP-diversity rather than some other property of random groups? One mark in favor of HP-diversity is its correlation with group competence (detailed above), but might there be another property that can explain the success of random groups?

In fact, C-diversity can explain the success of random groups much better than HP-diversity. Compare the scatterplot of C-diversity and group competence in figure 2 to the previous plot of HP-diversity and group competence (figure 1). Here again, the 50,000 runs is representative of a 1 million run sample. In the whole sample, the Pearson correlation between C-diversity and group competence is 0.553 (99.99% CI [0.551, 0.554]). Compared to the correlation between HP-diversity and competence ($r = 0.224$), C-diversity is significantly more correlated with competence (Williams's test of dependent correlations, $t = 356$; Steiger's $z = 341$). So knowing the C-diversity of a group gives us significantly more information about the group's ability than knowing the HP-diversity does.

This point is strengthened when we look at the informativeness of HP-diversity in sets of cases where C-diversity is held constant. Figure 3 shows a collection of scatterplots. In each plot, the C-diversity of each group is the same and the HP-diversity of the group is plotted against its competence.

From the definition of HP-diversity, it follows that the HP-diversity of a group is higher when pairs of heuristics in the group either do not share the same heuristic numbers or share the same heuristic numbers but in a different order. Since C-diversity tracks when the heuristics contain more heuristic numbers, changes in HP-diversity when C-diversity is held fixed must be explained by changes in the ordering either of heuristics in the group or heuristic numbers in particular heuristics. As such, we can see the scatterplots in figure 3 as showing the impact on group competence of a third kind of diversity, namely diversity of order, or *O-diversity*.

What Figure 3 shows us is that C-diversity (almost entirely) screens off HP-diversity, i.e. that when C-diversity is held fixed, HP-diversity gives us very little (if any) new information about group competence (partial Pearson correlation = 0.0019, $p < .05$). From that, we can infer that O-diversity doesn't track competence — it's not diversity in terms of the order of agents or how they use their heuristic numbers that explains the success of groups.

We can also see the power of C-diversity as an indicator of group competence by flipping the plots and looking at the correlation between C-diversity and competence when HP-diversity is held fixed. In figure 4, we see that holding constant HP-diversity, C-diversity still correlates with group competence almost as much as it did without holding HP-diversity fixed (partial Pearson correlation = 0.5188, $p < 2.2 * 10^{-16}$).

Altogether, these data show that C-diversity is a much better indicator of group competence than HP-diversity. But why? One possibility is that, because the landscapes are completely random, groups that have the collective ability to see more of the spots ahead are less likely to miss a peak. Or perhaps it's that highly C-diverse

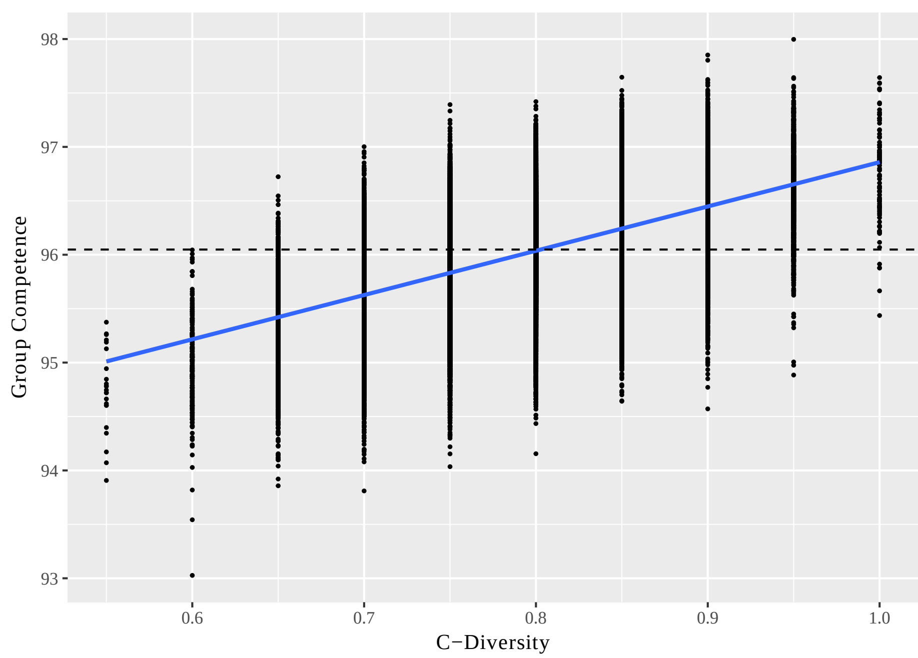

Figure 2: Comparison of C-Diversity to Competence for 50,000 runs. Includes linear regression (error bars too small to be visible). Average score of all random groups marked by dashed line.

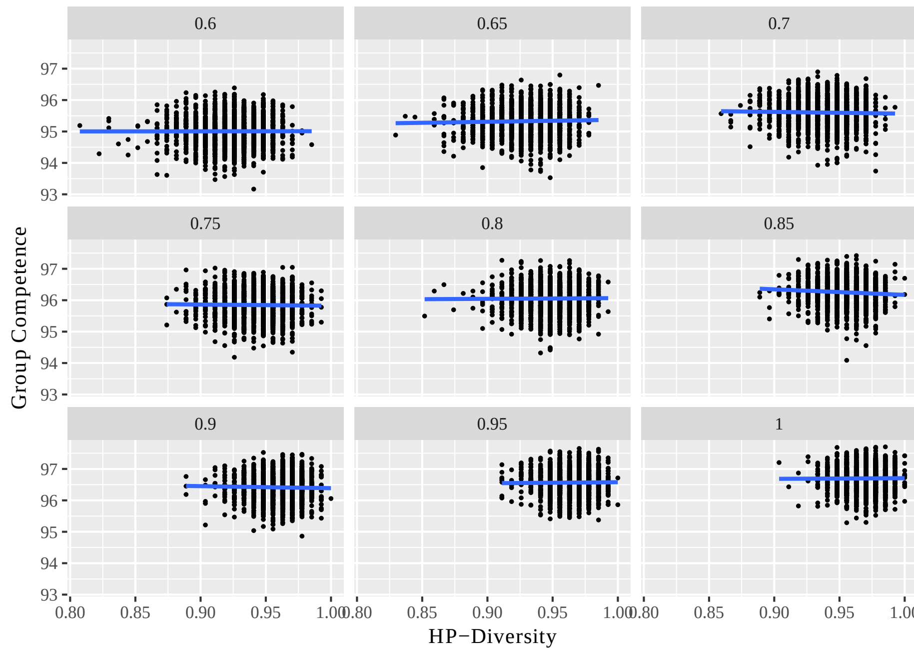

Figure 3: Comparison of HP-Diversity to Competence for 2,000 representative runs of each C-diversity in a 1 million run sample when $n \geq 2000$. Includes linear regression (error bars too small to be visible).

groups won't get stuck in some places that less C-diverse groups would.[11] If either of these were the right description of the mechanism, then it really would be that random groups bring more heuristics to the problem is what explains their success. Unlike what Hong and Page thought, order would be completely irrelevant. These stories are likely missing some details though. If it were *only* about groups not missing any peaks or not getting stuck, it would seem that the order of the heuristics in the group should matter, since an agent without any low-valued heuristics may cause the group to jump over a nearby peak. So further research into the complex dynamics of the model are necessary to fully answer this question.

Might there be even better measures of diversity (or other properties) that beat C-diversity in predicting success in this model? Perhaps a measure that tracks coverage as well as whether high- and low-valued heuristics are evenly spread-out in the group? I imagine measures like this would do better, but I won't attempt to discover those here. Here the goal was only to show that diversity of some kind is a better explanation of the success of random groups than randomness. Being more C-diverse means bringing more skills, methods of problem solving, background knowledge, perspectives, etc. to the table, and the advantage of random groups can be given a (partial) explanation in those terms, contra Thompson.

[11]Since groups only stop in the model once everyone has tried every one of their heuristics, groups that check more of the points ahead are more likely to find a higher point than groups who check fewer points.

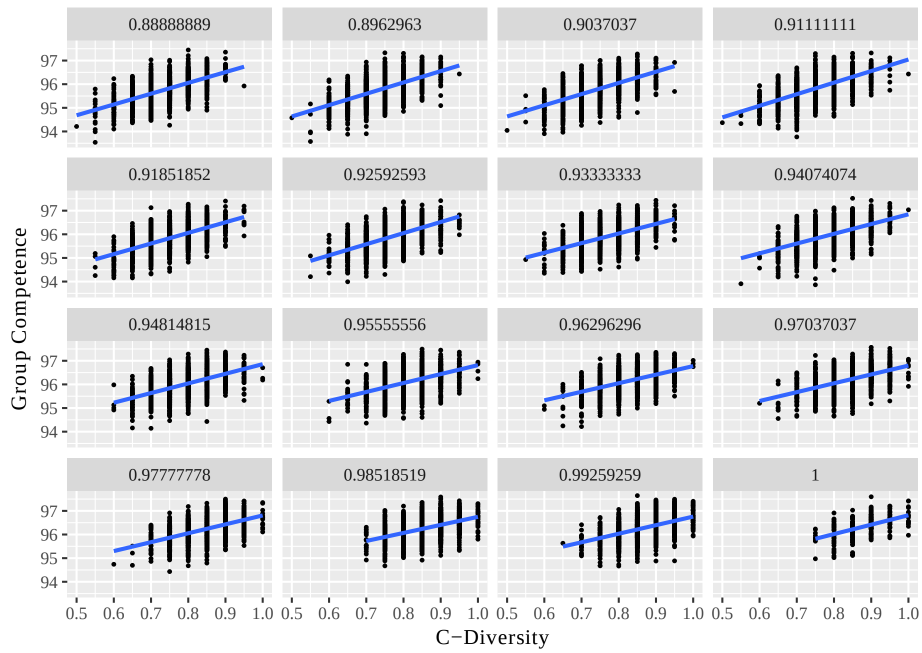

Figure 4: Comparison of C-Diversity to Competence for 2,000 representative runs for each HP-diversity in a 1 million run sample when $n \geq 2000$. Includes linear regression (error bars too small to be visible).

There's a broader point to be drawn from this discussion though. The formal models mentioned above are part of a larger discussion about the epistemic and practical import of functionally diverse groups and, less directly, identity diverse groups.[12] In this discussion, we often assume that measuring diversity is obvious and straight-forward. It's often measured as simply the lack of uniformity. The lesson is that we undermine our understanding of the issues in employing such anemic conceptions of functional diversity.[13] What we should learn from Thompson's critique is that we shouldn't be so casual about measuring this central notion.

Neither Thoma (2015) nor Zollman (2010) give a measure of diversity in their discussions, but there are natural candidates that the authors gesture at in both cases.[14] When we measure diversity in models like these, we should ask why we think that we're measuring an interesting notion of diversity rather than mere difference or non-uniformity. As I've shown in the case of Hong and Page's model, we have the tools to understand what counts as diversity that fosters the desired values. So, insomuch as we aim to justify theory or policy with results from these models, getting clear about what counts as diversity should be an essential part of the project.

[12] Exemplars of the broader class include Rubin and O'Connor (2017) and O'Connor and Bruner (forthcoming).

[13] I doubt we're blameworthy here, since we're at the early stages of using these models.

[14] For Thoma (2015) (and Weisberg and Muldoon (2009)), it's the ratio of 'explorers' ('mavericks') to 'extractors' ('followers'). For Zollman (2010), it's his maximum permissible α and β parameters.

References

Aaronson, S. (2017). P=?NP. Unpublished Manuscript. Available online: https://eccc.weizmann.ac.il/report/2017/004/.

Brennan, J. (2017). *Against Democracy*. Princeton University Press.

Bright, L. K. (2017). Decision Theoretic Model of the Productivity Gap. *Erkenntnis*, 82(2):421–442.

Fisher v. University of Texas, Austin (2016). Brief for Lt. Gen. Julius W. Becton, Jr., et al., as Amici Curiae in Support of respondents. (no. 14-981).

Greathouse, C. R. (2013). "P v. BPP" on Open Problem Garden.

Grim, P., Singer, D. J., Bramson, A., Holman, B., McGeehan, S., and Berger, W. J. (2017). Diversity, Ability, and Expertise in Epistemic Communities. Unpublished Manuscript.

Hong, L. and Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences of the United States of America*, 101(46):16385–16389.

Karp, R. M. (1991). An introduction to randomized algorithms. *Discrete Applied Mathematics*, 34(1-3):165–201.

Kuehn, D. (2017). Diversity, Ability, and Democracy: A Note on Thompson's Challenge to Hong and Page. *Critical Review*, 29(1):72–87.

Landemore, H. (2012). *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton University Press.

Martini, C. (2014). Experts in Science: A View From the Trenches. *Synthese*, 191(1):3–15.

O'Connor, C. and Bruner, J. (forthcoming). Dynamics and Diversity in Epistemic Communities. *Erkenntnis*, pages 1–19.

Page, S. E. (2007). *The Difference*. Princeton University Press.

Rosenstock, S., O'Connor, C., and Bruner, J. (2017). In Epistemic Networks, is Less Really More? *Philosophy of Science*, 84(2):234–252.

Rubin, H. and O'Connor, C. (2017). Discrimination and collaboration in science. *Philosophy of Science*.

Stegenga, J. (2016). Three Criteria for Consensus Conferences. *Foundations of Science*, 21(1):35–49.

Thoma, J. (2015). The Epistemic Division of Labor Revisited. *Philosophy of Science*, 82(3):454–472.

Thompson, A. (2014). Does Diversity Trump Ability? *Notices of the AMS*, 61(9):1–024.

Thompson, A. (2017). Personal Communication. Unpublished Code Referred to in Thompson (2014).

UCLA (2014). Proposed Diversity Requirement. https://ccle.ucla.edu/pluginfile.php/743624/mod_resource/content/6/082014%20REVISED%20DiversityReqProposal.pdf.

Vadhan, S. P. (2012). Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7(1–3):1–336.

Weisberg, M. and Muldoon, R. (2009). Epistemic Landscapes and the Division of Cognitive Labor. *Philosophy of Science*, 76(2).

Zollman, K. J. S. (2010). The Epistemic Benefit of Transient Diversity. *Erkenntnis*, 72(1):17–35.