

**Wisdom of Crowds, Wisdom of the Few:
Expertise versus Diversity across Epistemic Landscapes**

Patrick Grim, Daniel J. Singer, Aaron Bramson, Bennett Holman, Sean McGeehan & William J.

Berger

a version forthcoming in *Philosophy of Science* as
“Diversity, Ability, and Expertise in Epistemic Communities,”

Abstract

In a series of formal studies and less formal applications, Hong and Page offer a ‘diversity trumps ability’ result on the basis of a computational experiment accompanied by a mathematical theorem as explanatory background (Hong & Page 2004, 2009; Page 2007, 2011). “[W]e find that a random collection of agents drawn from a large set of limited-ability agents typically outperforms a collection of the very best agents from that same set” (2004, p. 16386). The result has been extremely influential as an epistemic justification for diversity policy initiatives. Here we show that the ‘diversity trumps ability’ result is tied to the particular random landscape used in Hong and Page’s simulation. We argue against interpreting results on that random landscape in terms of ‘ability’ or ‘expertise.’ These concepts are better modeled on smoother and more realistic landscapes, but keeping other parameters the same those are landscapes on which it is groups of the best performing that do better. Smoother landscapes seem to vindicate both the concept and the value of expertise.

Change in other parameters, however, also vindicates diversity. With an increase in the pool of available heuristics, diverse groups again do better. Group dynamics makes a difference as well; simultaneous ‘tournament’ deliberation in a group in place of the ‘relay’ deliberation in Hong and Page’s original model further emphasizes an advantage for diversity. ‘Tournament’

dynamics particularly shows the advantage of mixed groups that include both experts and non-experts.

As a whole, our modeling results suggest that relative to problem characteristics and conceptual resources, the wisdom of crowds and the wisdom of the few each have a place. We regard ours as a step toward attempting to calibrate their relative virtues in different modelled contexts of intellectual exploration.

Introduction

Expertise has taken some pretty hard knocks.

Empirical work by James Shanteau and his collaborators shows that there are remarkably low rates of agreement among supposed expert stockbrokers, polygraphers, and livestock judges. Alarmingly, in clinical psychology and medical pathology the work also shows remarkably low rates of agreement between the *same* expert at different times (Shanteau 2000, 2002; Shanteau, Weiss, Thomas & Pounds 2002, 2003; Weiss & Shanteau 2014).

Decades of work by Philip Tetlock finds a strong case for radical skepticism regarding expert political opinion. “When we pit experts against minimalist performance benchmarks—dilettantes, dart-throwing chimps, and assorted extrapolation algorithms—we find few signs that expertise translates into greater ability to make either ‘well-calibrated’ or ‘discriminating’ forecasts” (Tetlock 2005, 20). Within the data, however, Tetlock also finds evidence that one reasoning style does better than another: ‘foxes,’ self-critically following many leads, do measurably better than ‘hedgehogs,’ attempting to expand a single idea to cover new cases (67-120).

This empirical work seems to support the popular idea that there is a ‘wisdom of crowds.’ That work combines anecdotes regarding the surprising accuracy of majority or mean group outcomes (*Who Wants to be a Millionaire?* and Galton’s ox) with a simple statistical explanation. When some members of a population know the answer to a multiple-choice question or can give an accurate estimate on some amount, and when all error or other input is random, the random input will tend to cancel out. In a group verdict, the wisdom of the few knowledgeable members of the crowd can then shine through in the majority vote or the estimate mean (Surowiecki 2004). Interestingly, the wisdom of those few members will shine out without any independent ability on our part to identify who in the crowd those ‘experts’ are.

There is also formal work beyond the simple statistical argument that argues for important epistemic advantages of group, as opposed to individual, decision. The Condorcet Jury theorem shows that a majority decision by individuals with a mean probability of a correct decision $> .5$ has a higher probability of being correct than does the decision of an individual with that mean probability (Condorcet 1785/1995; Anderson 2006; Landemore 2013). Our focus here will be on the more recent Hong-Page model, in which agents use different heuristics in order to explore a hilly epistemic landscape. Hong and Page present that model as demonstrating a ‘diversity trumps ability’ result: under plausible conditions a group of individuals with randomly diverse heuristics can be expected to achieve better epistemic results than a group of the ‘best’ individuals—those with heuristics that score best individually. Diversity trumps ability: “a randomly selected collection of problem solvers outperforms a collection of the best individual problem solvers” (Page 2007, p. 162).

These results have been taken to have profound implications for social policy. Beyond arguments from a social justice perspective, the results suggest that an organization is

epistemically better off by recruiting a diverse set of candidates instead of just selecting the best individual performers. The work has been presented to NASA, cited by the USGS, is one of four works cited in support of positive expected institutional effects of UCLA's (2014) proposed diversity requirement, and has been appealed in support of promoting diversity in the armed forces in a brief submitted to the Supreme Court in the most recent case to adjudicate the issue (Fisher v. University of Texas, Austin, 2016). These results have also been taken to apply to the value of epistemic diversity in scientific communities (Bright, 2016; Martini, 2014; Stegenga, 2016; Thompson, 2013).

Here we show that the reported epistemic superiority of random groups depends crucially on the character of the epistemic landscape being explored. On random landscapes, using heuristic sets like those in their original work, the Hong-Page result is confirmed: groups of random heuristics systematically do better. We argue against interpreting these basic results in terms of 'ability' or 'expertise,' however, since high performance of a heuristic on one random landscape has little correlation with high performance on any other. We also demonstrate, using other parameters as in the original, that the benefits of diversity disappear with only slightly 'smoother' or less random epistemic landscapes.

With smoother landscapes high performance on one landscape is much better correlated with high performance on another, better supporting an interpretation in terms of 'expertise.' But on those smoother landscapes, using Hong and Page's heuristic pool, it is groups of 'experts' that do better than groups of random heuristics. Where something more like expertise becomes recognizable, these results suggest, it is expertise that trumps diversity.

It turns out that the scope of heuristics available for landscape exploration is equally important, however. In the Hong-page model, heuristics for all individuals are composed of

numbers drawn from a set of a given size. For a specific landscape smoothness, groups of best-performing individuals outperform random groups for a heuristic set of a given size. But if the size of that heuristic pool is increased, something like a Hong-Page result is vindicated in that a ‘diversity trumps ability’ result returns. Given a large enough pool of heuristic possibilities, groups of random agents again outperform groups of the individually best-performing.

In a first section we offer a replication of the Hong-Page result. In a second section we argue against interpretation of the original results in terms of ‘ability’ or ‘expertise.’ The third section outlines the central concept of ‘smoothness’ in epistemic landscapes. Keeping other parameters the same, we demonstrate that the Hong-Page ‘diversity trumps ability’ result disappears on smoother and arguably more realistic landscapes: here ‘ability trumps diversity’ instead. The virtues of diversity reemerge, however, when the heuristic pool is widened. Section four offers parameter sweeps across landscape smoothness and the size of the heuristic pool to demonstrate the relative areas of strength for diverse groups and groups of the best performing. In section five we emphasize another parameter: the pattern of group dynamics in navigating a landscape. Here results again favor diversity in a sense, but the real winners appear to be mixed groups, which are composed of both experts and random agents. Consideration of the effect of group size appears in section six.

What the results suggest is a lesson regarding the relative virtues of both expertise and diversity within particular exploratory contexts. The greater value of expertise is vindicated within specific conceptual constraints, using a specific group dynamic, and relative to a specific range of epistemic landscapes. The greater value of diversity is vindicated within a different range of group dynamics, epistemic landscapes, and conceptual resources. Thus the wisdom of crowds and the wisdom of the few each have a place. We regard the current work as an

important next step in attempting to calibrate their relative virtues in different contexts of intellectual exploration.

I. The Hong-Page Result

Hong and Page offer several variations on a formal model of discussion within a group of individuals (Hong & Page 2004, 2007; Page 2007, 2011). In computationally implemented simulations, agents use heuristics to explore an epistemic landscape. The ‘diversity trumps ability’ result is that epistemic outcomes for groups of individuals with a heterogeneous group of heuristics will consistently exceed those for groups composed solely of individuals with heuristics that gain the highest individual scores. Though accompanied by a mathematical theorem intended to offer partial understanding, the main result appears in simulation rather than in formal proof. The same will be true of our work here.

We begin with a version of the model that is close to Hong and Page’s original. Epistemic exploration is along a linear terrain of 2000 points that forms a loop; 10 points to the right of 1995 is point 5, for example. For each of the 2000 points of the terrain random integer between 1 and 100 is assigned; higher points are interpreted as better answers to a question.

Individual agents are assigned a *heuristic*, modeled as an ordered set of k numbers, each of which is a number between 1 and l . We begin, again following the Hong and Page original, with ordered sets of 3 numbers ($k=3$) between 1 and 12 ($l=12$). With those in hand, respecting order but avoiding duplication, we have 1320 agents defined by distinct heuristics.

Individuals use their heuristics as follows. An agent starts at, say, point 112 of the 2000-point terrain, which carries a value of 80. The agent then applies the first digit of its heuristic: Does the point that many steps to the right offer a higher value? If not, it stays put. If so, it moves to that point. In either case it then applies the second digit of its heuristic. Does that offer a point with a higher value? Once the third digit has been tried it returns to the first. An individual stops only when none of its digits can reach a higher point—its local maximum applying the cycled heuristic from the initial point 112. In exploring the terrain in this way, each of our 1320 agents can be scored by the highest value it reaches starting at each of the 2000 points. The average of those is an individual's score. Our 9 'best' individuals will be those with the 9 highest scores.

To model discussion within a group, Hong and Page employ a sequential "relay" among, say, a group of 9 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 she passes the "baton" of that highest point to the next agent as a starting point. He then searches for a higher maximum by employing his heuristics until his search is exhausted, at which point the baton is passed 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. The final decision for the group will be a local maximum from which none of the heuristics can find a higher point. The discussion can be thought of as a conversational relay, proceeding in orderly fashion around a circular table. The average score for the group will be the average over all of our 2000 starting points.

What Hong and Page compare are the results of a modeled discussion of this form for (a) a group composed of the 'best' individuals—those with the highest individual scores—and (b)

random individuals drawn from the heuristics pool at large. The ‘Diversity Trumps Ability’ result is the fact that the random group consistently does better.

In our reproduction of the Hong-Page result we use groups of 9—random and ‘best’—taking the average value across the 2000 points for each group. With random values between 1 and 100 for each of the 2000 points, the average maximum on the 2000-point terrain for the group of the 9 best individuals over 1000 runs was 92.53, with a median of 92.67. The mean average for a group of 9 random individuals was 94.82, with a median of 94.83. Across 1000 runs a higher score was achieved by random agents in 97.6% of all cases.

Thompson (2014) challenges both the intrinsic interest of the mathematical theorem that Hong and Page offer and its relevance to their conclusions. 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 Page and Hong’s simulation and our replication the strict conditions required for the theorem are significantly relaxed. Even here, however, where probability falls below 1, the results above show the central result to be extremely robust. Within the simulation parameters specified, the epistemic success of a collection of random heuristics proves consistently superior to that of a collection of those which individually score the best.

II. Interpreting Hong-Page

The model is extremely suggestive, and has been offered as support for a number of strong conclusions. In both their original work and in later applications Hong and Page allude to diversity as a value in affirmative action (Hong & 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 & Page 2004). It is to their credit, we think, that Page and Hong tend not to use the term ‘experts.’ In reviews and applications of their work, however, it is probably natural that their results are taken as part of the larger case against expertise (Landemore 2013, Gunn 2014, Weymark 2014). The Princeton University Press’s blurb on the back of the book characterizes Page’s *The Difference* as revealing “how groups that display a range of perspectives outperform groups of like-minded experts.” Elizabeth Anderson characterizes Hong-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” (Anderson 2006, 12).

Page and Hong’s most careful statement of the central result refers simply to ‘best-performing’ agents. There is good reason to resist interpreting those agents as either ‘experts’ or as individuals with the highest ‘ability.’ In their computational experiment, as outlined above, the landscape on which these are the ‘best-performing individuals’ is a purely random landscape. Because of that, different landscapes produce ‘best-performing individuals’ with very different heuristics. An individual with a set of heuristics that is ‘best-performing’ on one random landscape is very likely to do extremely poorly on another.

Table 1 shows the top 9 heuristic sets in a model runs on different random landscapes. The first thing that leaps out is the redundancy of heuristic numbers across the ‘best-performing’ within each individual landscape. On the first landscape, for example, the numbers 4 and 12 appears in every one of the ‘best-performing’ heuristic sets. The redundancy of the ‘best-performing’ is a major part of Hong and Page’s own analysis of both formal results and social

implications: why hire 5 individuals with the same background if you will just hear the same message five times?

#	Heuristic Sets for the ‘Best-Performing’ Agents								
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)

Table 1. Heuristic sets for the ‘best-performing’ on 10 different random landscapes

On a single random landscape, the ‘best-performing’ individuals will tend to have heavily overlapping heuristic sets: it will be the same numbers, or close to the same numbers, that do well in different orders. But those particularly valuable heuristic numbers will be specific to a given random landscape. Numbers 12 and 4 appear in all of the best-performing heuristics for the first landscape in Table 1, for example, but neither number appears in any of the best-performing heuristics for the second or third landscapes. Table 2 shows the percentage of cases in which each of our 12 numbers appears among the 3 heuristic numbers of the 9 ‘best performing’ agents on 100 random landscapes. Here there appears to be no clear signature of ‘ability’: each heuristic number shows roughly equal representation across the random landscapes as a whole.

Percentage of ‘best-performing’ in which each heuristic value appears											
1	2	3	4	5	6	7	8	9	10	11	12
22.7	21.1	19.2	22.3	22.2	24.7	28.3	23.4	31.6	24.3	31.4	28.3

Table 2. Percentage of cases in which each value appears among the 3 heuristic numbers of the 9 ‘best performers’ on 100 random landscapes.

Because ‘best-performance’ is so specific to a given random landscape, it seems dubious that ‘best-performing’ agents qualify as ‘experts’ in the familiar sense. Their success is the local success of a specific set of heuristics on what is essentially a specific pattern of noise. 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. Genuine expertise and ability, where they exist, should be transferable. An expert on livestock should be able to give us reliable results on various groups of livestock. An individual with an ability to predict the weather should be able to predict the weather in various conditions and on various days. ‘Best-performing’ in the Hong Page model, keyed to a single random landscape, is not transferable in that sense.

Our replication of Hong and Page’s simulations shows that the formal result is secure on random landscapes. Interpretation is another matter. Because of the central role of a random landscape in that result, we suggest that ‘high-performance’ in that original not be interpreted in terms of agent ‘ability’ or ‘expertise,’ and advise that social and policy implications be handled with extreme care. In the next two sections we explore variations on the model using landscapes for which interpretations in terms of transportable ‘expertise’ seems significantly more appropriate. Those are also variations, however, in which the ‘diversity trumps ability’ result disappears.

III. Expertise over Diversity on Smoother Epistemic Landscapes

What we explore here is how robust the Hong-Page results are, concentrating on one variation in the original model. Hong and Page themselves emphasize a number of important conditions on their result. These appear as strict conditions within their formal theorem, but also in a more

relaxed form in the context of simulation and in their discussion of model implications (Hong & Page 2004, Page 2007, 2011). They assume that the pool of agents from which groups are selected must be diverse. The groups themselves must be ‘good-sized.’ And, most importantly for our purposes, they assume that the problem to be solved must be difficult.

Hong and Page’s original specification for a ‘difficult’ problem is that there be no individual problem solver who always finds the global maximum (Page 2007, p. 159). Although our models count as difficult in that sense, we use a more nuanced measure for the character of problems: the ‘smoothness’ of the epistemic landscape at issue.

One elementary way of ‘smoothing’ a random landscape would be the following. Instead of assigning a random value to each of 2000 points, as above, we might assign a random value to a selection of the points—to just the even-numbered points, for example. If we then fill in the landscape by putting mean of the even points’ values at the odd points, we have a landscape intuitively twice as smooth as our random original. For a still smoother landscape, we assign random values to every third point, or every fifth, drawing roughly descending or ascending lines between the assigned points.

Here we construct smoother landscapes using a slightly more sophisticated version of the same idea. We assign a random value to point 1. For a smoothing factor of x , we pick a random integer between 1 and $2x$, assigning a random value at that point. Our assigned points thus average a distance of x without the artificiality of a fixed interval. Points between those assigned are positioned roughly on a line of ascending or descending values between them. Epistemic landscapes with smoothing factors of 0, 5, 10 and 20 are shown in Figure 1.

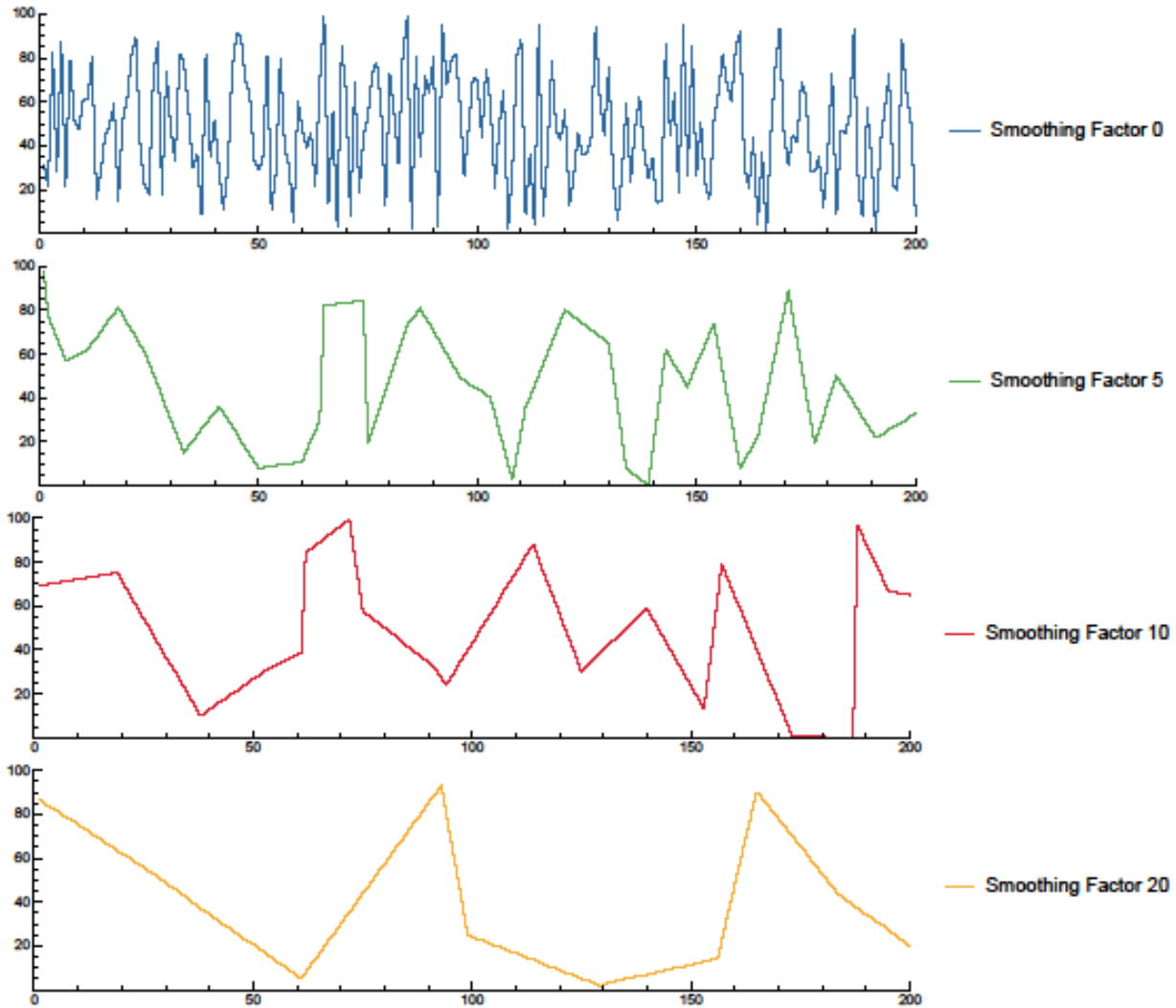


Figure 1 Sample landscapes created with smoothing factors of 0, 5, 10 and 20.

There is a sense in which landscapes become less difficult as they become smoother. Hill-climbing becomes a reliable strategy: unlike a purely random landscape, smoother landscapes offer extended slopes to climb. In another sense, however, increasingly smooth landscapes pose increasingly difficult problems. Peaks become increasingly far between relative to the limited reach of available numbers, and canyons between them become increasingly hard to cross. Problem difficulty is thus a complex notion: we take ‘smoothness’ as a measure of the *character* of a problem space, without trying to translate it directly as ‘difficulty.’

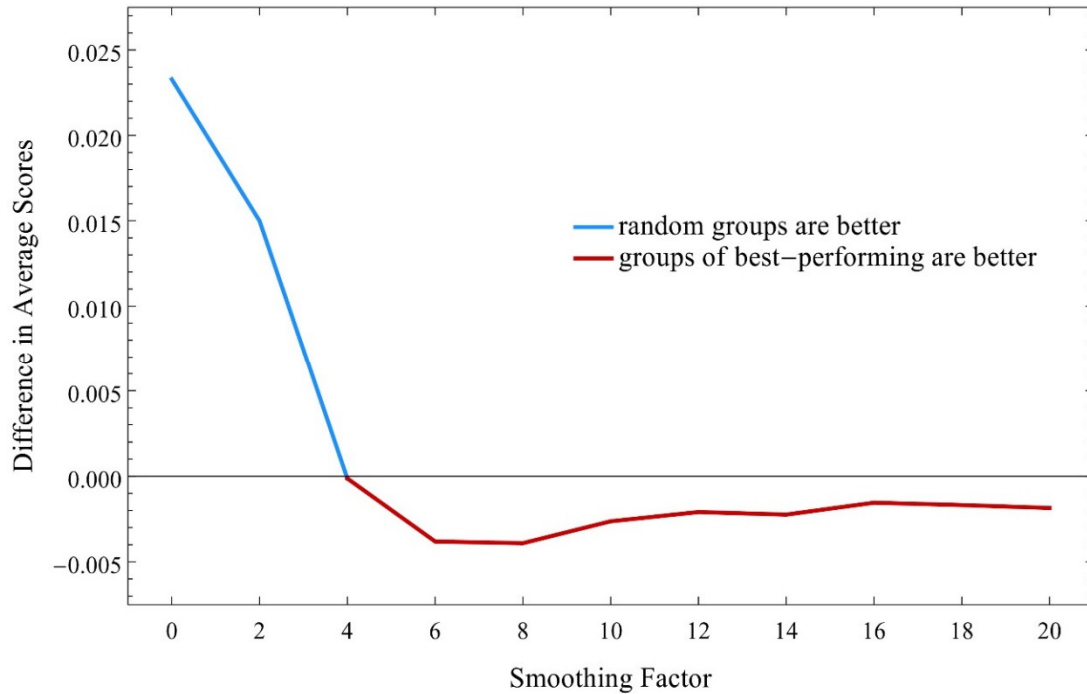


Figure 2. Groups of 9 individuals using 3 Heuristics from a pool of 12. ‘Diversity trumps ability’ only for landscape smoothness less than 4. Differences in averages shown.

How robust is the conclusion that ‘diversity trumps ability’ with increasing smoothness of a landscape? Our agent heuristics consist of ordered sets of 3 numbers between 1 and 12, resulting in 1320 possibilities. Over 100 runs, we take average values starting from each of 2000 points for (a) a relay group of 9 random individuals and (b) a relay group of the 9 individuals that perform best individually. For landscape smoothness from 0 to 20, Figure 2 graphs the subtractive difference (best-performing group from random).

What is notable is the cross-over point, indicating a reversal of the ‘diversity trumps ability’ claim. Given the other assumptions carried over from the Hong and Page simulation, groups of 9 random individuals do better than 9 of the best-performing—‘diversity trumps ability’—only when landscape smoothness is less than 4. At landscape smoothness 4 there is a cross-over. Beyond that point ‘ability trumps diversity’: groups of the best-performing outperform the random group. The advantage shown is small: at a smoothness factor of 6, for

example, the average performance over 100 runs is 0.760 and 0.756 for ‘best’ and random groups, respectively. That small advantage of the best over the random is, however, clear and robust beyond the cross-over.¹

It can be argued that smoother landscapes are more realistic in leaving purely random data behind and, at least up to a point, in capturing the difficulties of heuristic ‘reach.’ The fact that groups of the ‘best’ outperform random groups on those smoother landscapes is thus arguably a mark of realism in favor of the value of expertise: ‘ability trumps diversity.’

It is also on those landscapes that the transportability of ‘best-performing’ becomes evident, and thus interpreting ‘best-performing’ in terms of ‘ability’ or ‘expertise’ becomes more plausible than on the original Hong-Page landscape. One measure of whether a task admits of expertise is whether supposed experts consistently do well on repeated trials on different versions of the same task. We therefore compared agents’ performance on pairs of landscapes of equal smoothness, repeating simulations 100 times at smoothness from 0 to 10 for all 1320 different heuristics. We then ran a Pearson correlation for each pair of trials and averaged the 100 repetitions at each smoothness level (Table 3).

From this data, one can further calculate what percent of the variance is explained by the individuals’ heuristics. In essence, this value calculates the proportion of the performance in the second trial that can be predicted from the performance in the first trial (Figure 3).

¹ To this point we are considering the ‘relay’ group dynamic reported in the Hong-Page simulation: a first agent reaches his local maximum, a second agent attempts to go farther from there, and so forth. That dynamic might also be modified to a simultaneous ‘tournament,’ in which all agents report their local maxima and each tries to then go from the highest reported by any. Hong and Page report qualitatively similar results for the simultaneous ‘tournament’ variation. We find a similar cross-over in favor of groups of the best-performing using a simultaneous ‘tournament’ dynamic, though the smoothness factor at which the cross-over occurs is slightly higher. Further and important differences between the two dynamics are outlined in section V.

SF	Correlation	SF	Correlation	SF	Correlation
0	0.1575	7	0.9772	14	0.9674
1	0.8866	8	0.9807	15	0.9694
2	0.9645	9	0.9778	16	0.9631
3	0.9785	10	0.9744	17	0.9615
4	0.9830	11	0.9734	18	0.9600
5	0.9814	12	0.9706	19	0.9591
6	0.9793	13	0.9700	20	0.9549

Table 3. Average over 100 repetitions of Pearson correlations for all heuristics on pairs of landscapes of the same smoothing factor

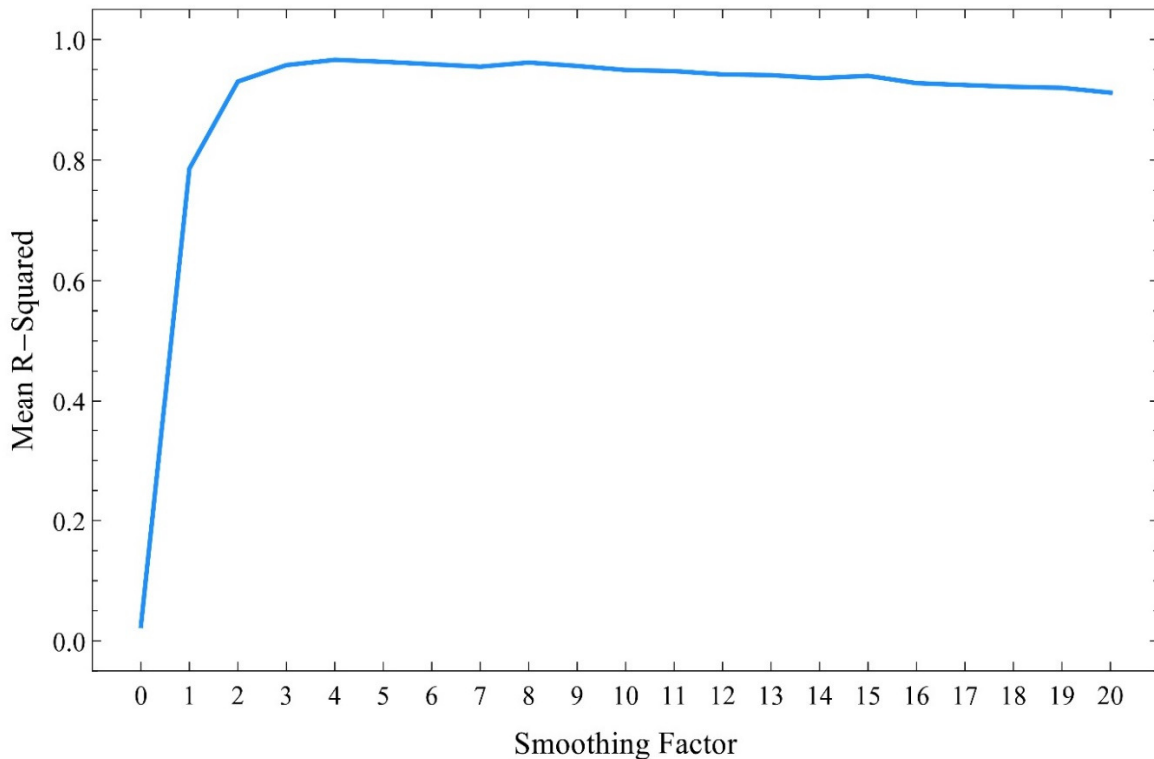


Figure 3. Percentage of variance on two landscapes of the same smoothness explained by individual heuristics.

There is a clear and sudden jump in the transportability of individual performance on one landscape to another of the same smoothness as smoothness increases from 0 to 1 and 2. It is only here, we propose, that one can properly speak of a heuristic as representing ‘ability’ or ‘expertise.’ On that proposal, the ‘ability’ or ‘expertise’ that is trumped by diversity in the original Hong-Page simulation is not true ‘ability’ or ‘expertise.’ At roughly (though not

exactly) that point at which we can more plausibly speak of transportable ability or expertise across landscapes, it is no longer true that diverse groups are better performing. It is roughly (though not exactly) where expertise in this sense emerges that groups of ‘experts’ start to outperform groups of random heuristics.

Table 2 shows the percentage of cases in which each of our 12 numbers appears among the 3 heuristic numbers of the 9 ‘best performing’ agents on 100 landscapes. Here there appears to be no clear signature of ‘ability’: each heuristic number shows roughly equal representation across the random landscapes as a whole.

We can also say something about what that expertise consists in. Table 4 expands Table 2 to show the percentage of cases in which each of our 12 numbers appears among the 3 heuristic numbers of the 9 ‘best performing’ agents on 100 landscapes with different smoothing factors. Where a random landscape of smoothness 0 shows no particular preference for any specific digits among the ‘best-performing,’ a preference immediately emerges at smoothing factor 1: the number 1 appears among all of the best 9 heuristics in all 100 cases. Extreme numbers at the other end, particular 12, approach the importance of 1 as the smoothing factor increases, as do middle numbers in the region of 7. The number 2 disappears entirely at smoothing factor 1, on the other hand, joined by the number 3 at smoothing factor 2. Neither 2, 3, nor 4 appear among the best heuristics at smoothing factor 3. In all these cases, unlike the random landscape, it is clear that there are certain patterns of heuristics—the ‘expert sets’—that tend to do best quite generally across landscapes of a particular smoothness.

		Heuristic Number											
		1	2	3	4	5	6	7	8	9	10	11	12
Smoothing Factor	0	22.7	21.1	19.2	22.3	22.2	24.7	28.3	23.4	31.6	24.3	31.4	28.3
	1	100	0	3.8	23.4	20.4	17.0	21.4	19.9	19.3	22.4	20.6	31.4
	2	100	0	0	0.2	10.3	32.7	35.7	19.6	3.7	11.1	24.4	62.3
	3	100	0	0	0	1.7	21.0	46.4	26.8	4.1	0.4	13.8	85.8
	4	99.4	0.5	0	0.5	7.8	23.4	33.8	27.7	6.6	0.1	3.9	96.1
	5	98.7	1.4	1.7	5.8	14.2	21.6	26.8	21.0	7.7	1.2	0.3	99.7

Table 4. Percentage of cases in which each value appears among the 3 heuristic numbers of the 9 ‘best-performers’ on 100 landscapes for smoothing factors 0 through 5. Column represents the smoothing factor while row represents percentage of cases.

Why exactly these particular heuristic values do well on these smoothing factors is a mystery to us. Possible explanations include that ‘1’ is valuable as a heuristic number 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. With ‘1’ present, the value of 2 tends at best to be 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 0. The same appears to be true for 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 as the highest number available, on the other hand, is that it offers the best promise of leaping over declines to an incline on the other side of a valley—a promise that is of increasing importance as the width of valleys widen as smoothing factor increases.

The details aside, the broader lesson of this first set of results is a warning against accepting the claim that ‘diversity trumps ability’ without a qualification regarding the character of the epistemic landscape at issue. Keeping other values in the Page and Hong simulation constant, it turns out that groups of random agents do better than groups of ‘experts’ or high-

performing individuals only for a very narrow range of fairly random landscapes. For more realistic landscapes—on which successful individuals are more easily interpretable as ‘experts’ in terms of heuristics transportable with comparable success—it is the groups of high-performing individuals that do better. Given the other assumptions in play, it is ability that trumps diversity on smoother landscapes.

IV. Diversity over Expertise with Larger Heuristic Pools

With a heuristic pool limited to numbers between 1 and 12, there is a cross-over in favor of groups of the best-performing once the smoothness factor exceeds 4. At that point, it is no longer true that ‘diversity trumps ability.’ Another cross-over occurs, however, if we increase the size of the heuristic pool. Rather than being confined to numbers between 1 and 12, for example, our heuristics might have any three numbers between 1 and 16 or 1 and 20. With increase in the heuristic pool as a further parameter, diversity again shows its strength.

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 Figure 2). When heuristic numbers are chosen from a set of 24 or more, however, the group of random heuristics again does better. Figure 4 shows subtractive differences in average score for groups of random heuristics and the best-performing as we increase the size of the heuristic pool.

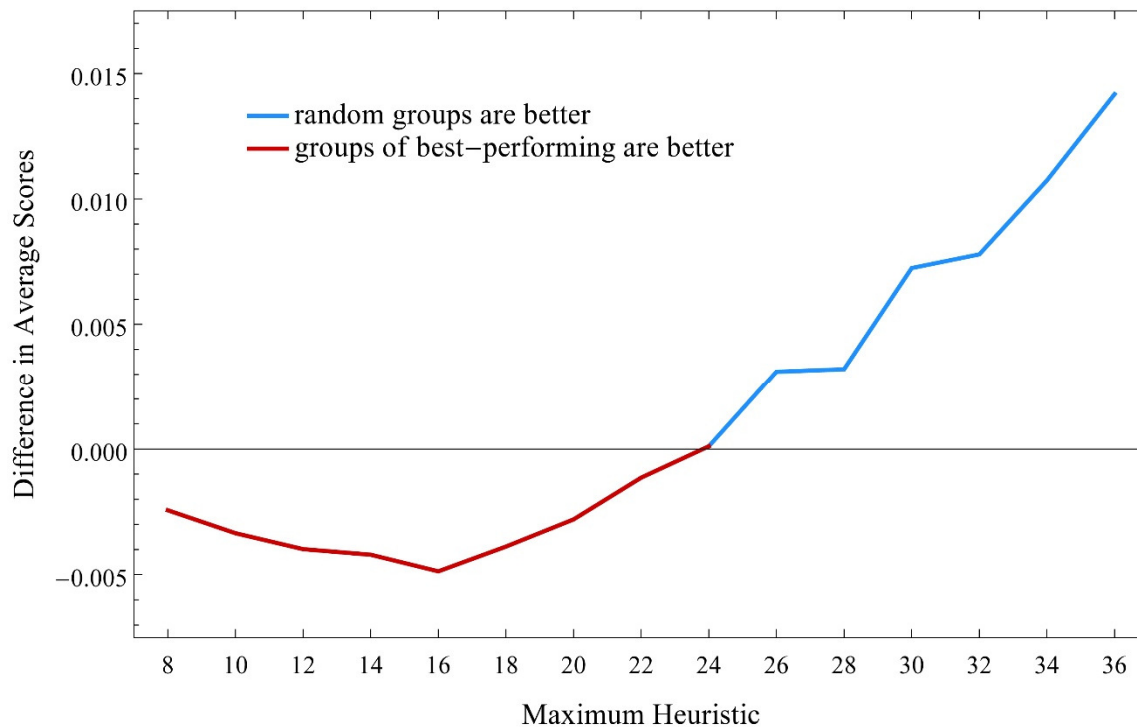


Figure 4. Cross-over in favor of random groups at heuristics pool of 24 with a landscape smoothness of 8. Differences in averages over 100 runs shown.

Across various points of smoothness we have found a very rough ‘rule of three.’ For heuristic pools that are less than three times the smoothing factor of the landscape, groups of 9 composed of 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 ‘diversity trumps ability’ effect. Although increases in landscape smoothness favor expertise, such an advantage is always relative to the maximum size of the heuristic pool from which strategies are drawn.

The virtues of diversity and ‘expertise’ are therefore relative to the interaction of at least *two* important factors: landscape smoothness and heuristic pool. Figures 5 through 7 show a parameter sweep across both variables, indicating distinct areas of relative strength for diversity and expertise.

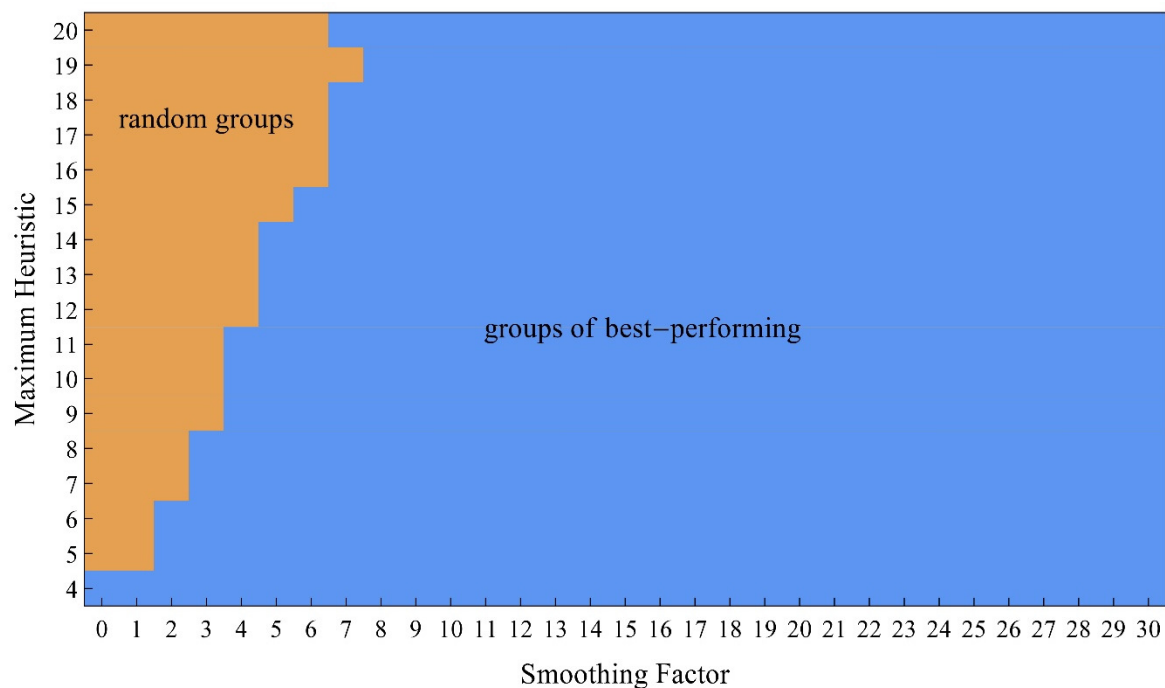


Figure 5 Areas in which groups of random heuristics do best (brown) and areas in which groups of the best-performing do best (blue) across a parameter sweep of landscape smoothness and the size of the heuristic pool.

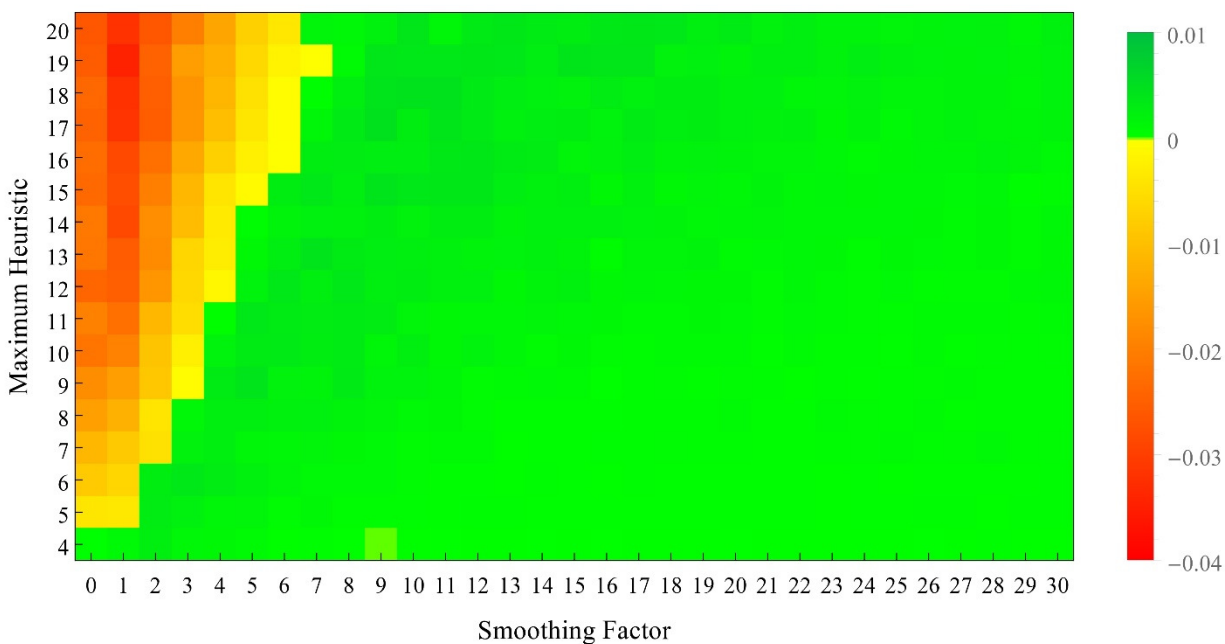


Figure 6 Differences in averages for groups of random heuristics and groups of the best-performing over 100 runs at different parameters for smoothing factor and heuristic pool. Positive values (in green) show higher averages for groups of the best-performing. Negative values (in yellow and red) show higher averages for groups of random heuristics.

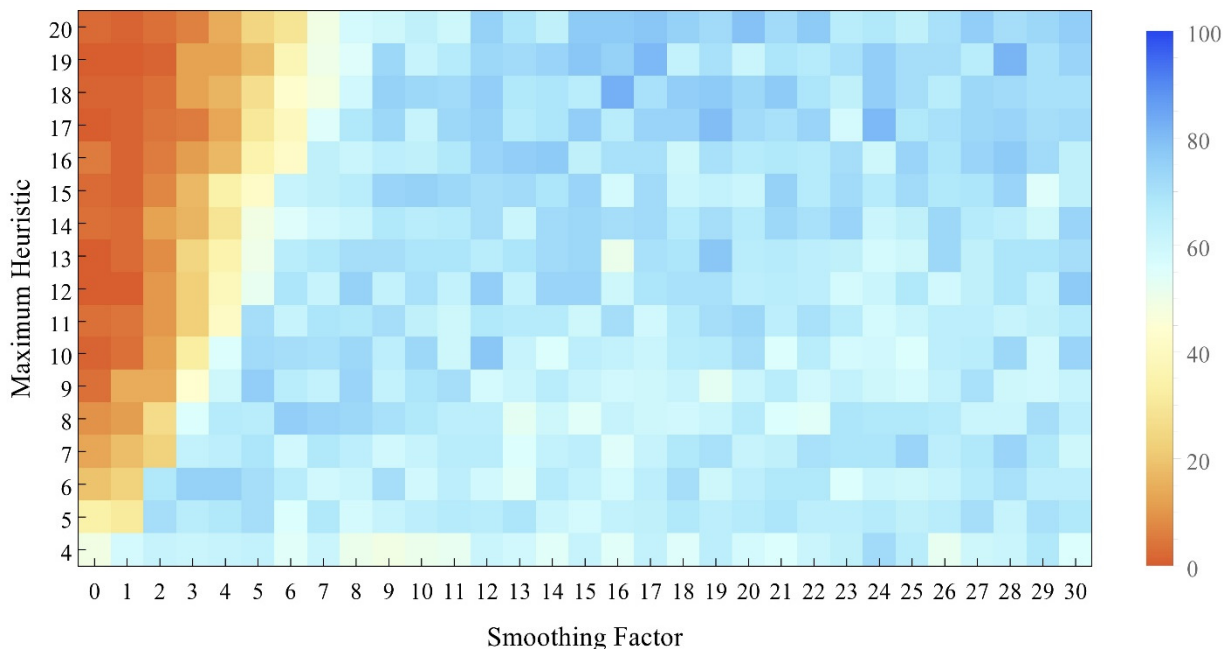


Figure 7 Percentages of runs in which groups of the best-performing do better than groups of random heuristics. Percentages greater than 50% colored in blue; those less than 50% colored in yellow and orange.

Figure 5 presents the data in the roughest form, showing those areas in which the average score for each is greater over 100 runs. Figures 6 and 7 show the more nuanced reality behind this result. Even where an average over 100 runs is higher for diversity as opposed to expertise, the difference may be very slight. Figure 6 shows the same data mapped in terms of difference in average. Figure 7 shows the percentage of 100 runs in which a random group or expert group does better at each setting of maximum heuristic and landscape smoothness.

Landscape smoothness can be taken as a measure of the character of a problem, with one aspect of difficulty increasing as the landscape is ‘stretched.’ We can think of the size of the heuristic pool as a measure of the conceptual resources available to tackle the problem represented by a particular landscape. What our results suggest is therefore particular niches in which expertise and diversity are each of particular value. Expertise is favored for a wide range of ‘smoothness,’ but only where the available conceptual resources are importantly limited.

With a wider pool of conceptual resources, a diverse group will do better even on problems of that same character.

A key to understanding many of these results, we think, is the extent of heuristic coverage represented in a group of individuals. As landscapes increase in smoothness from random, best-performing individuals tend to be very much alike, as indicated in table 4. A small number of the available heuristic numbers will be best on landscape of that smoothness, and all experts will share that small set of numbers. A *group* of experts will therefore show high redundancy: their collective numbers will not be dense on the space of heuristic numbers. We suggest that what gives random groups of heuristics an advantage is not that they are random (pace Thompson 2014) but that they will offer a more complete coverage of the space of heuristic numbers. Between them a group of random strategies will have more numbers to try, and so have a prospect of reaching higher peaks and avoiding more local maxima.

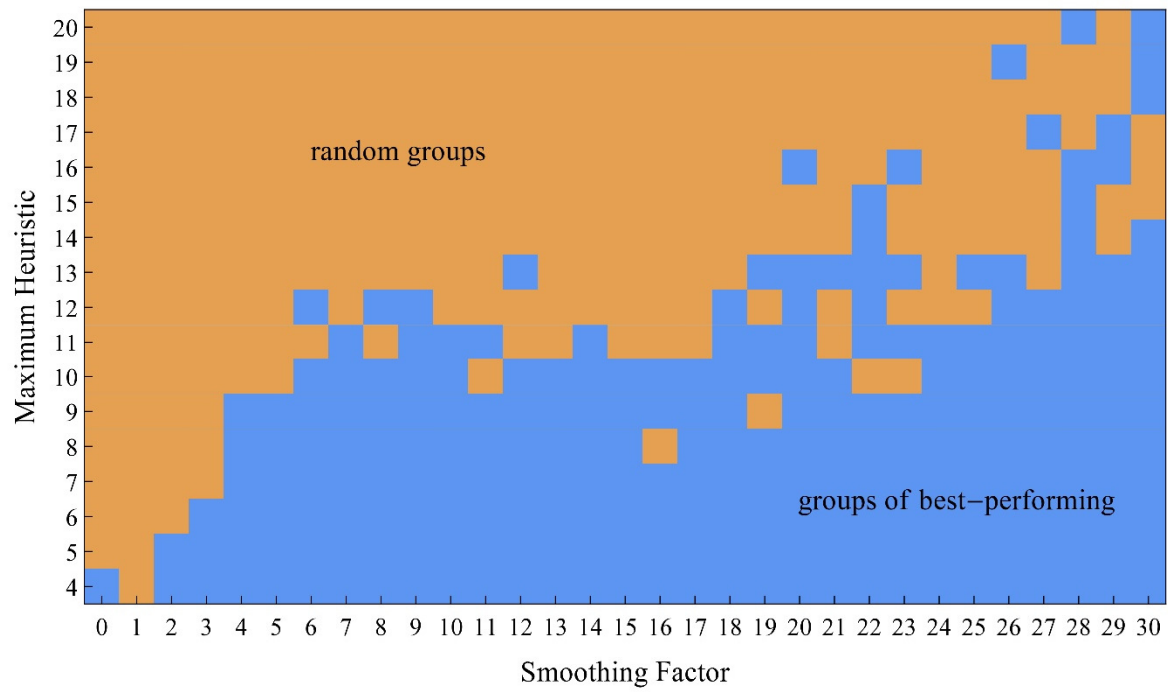
As Page and Hong hint in their original work, greater coverage explains the success of groups of random heuristics on random landscapes: the random groups have more heuristics to work with in their union, and so have a greater number of options to pursue in finding highest peaks. Coverage also helps to explain the fact that for a landscape smoothness at which experts do better, random heuristics can pull ahead with an expanded heuristic pool. In a larger heuristic pool the percentage of ‘expert’ numbers is smaller; the relative redundancy, we might say, is larger. 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 than experts with increased heuristic pools seems to be because their coverage of available heuristic numbers increases with a larger pool.

V. Group Dynamics and Composition: Diversity and Expertise

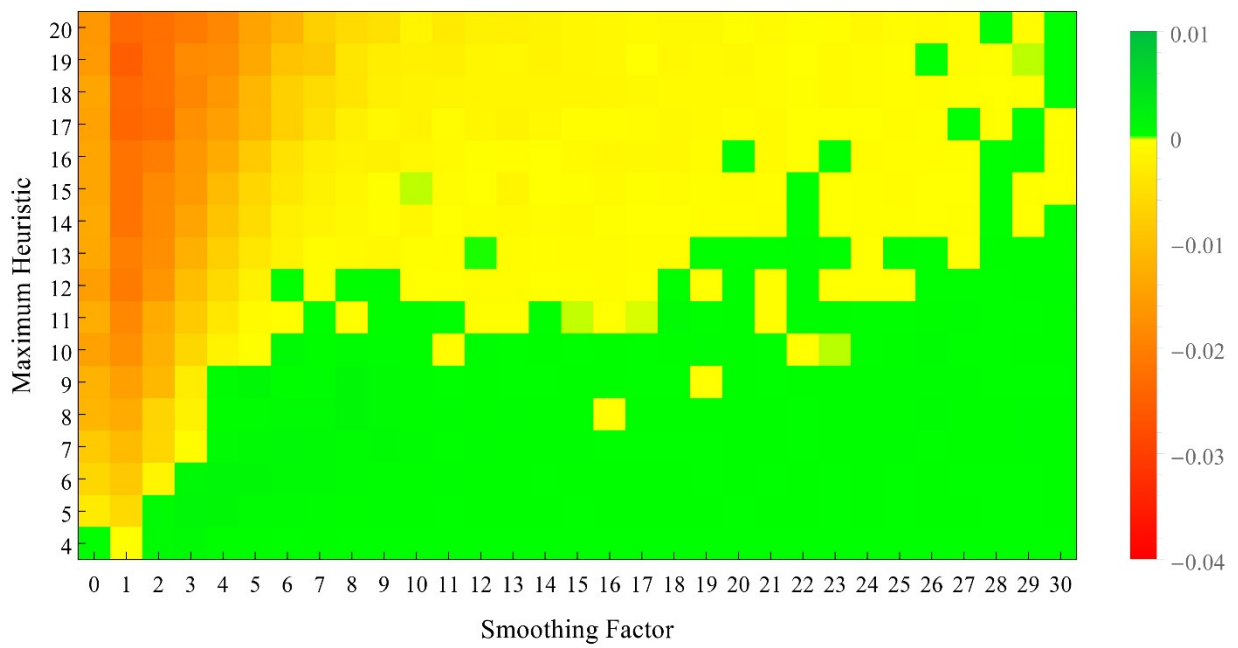
There is a further factor that surprisingly and dramatically favors random heuristics, largely ignored in the original Hong-Page results. In all the results above we use the ‘relay’ dynamics employed in Hong and Page’s original simulation. Starting from a given point, an assigned first agent in the group finds the highest point her heuristic will reach. The second agent then starts from that point in search of a higher and so forth. Once all members of the group have sequentially sought for the highest point from that 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 for their highest point. From the highest of any of those, all again try for a higher, and so on. What is eliminated in tournament dynamics is the artificial around-the-table sequential updating of a relay. Hong and Page consider both dynamics, saying that “our results do not seem to depend on which structure was assumed” (2004, 16386). For many group dynamics, we would argue, it is ‘tournament’ rather than ‘relay’ that is a more realistic model of epistemic exploration. And here results do depend on which structure is assumed. Just as group size and the conceptual headroom of a larger heuristic pool favor diverse groups, so does the use of tournament over relay dynamics. In comparison with Figures 5 through 7, Figure 8 shows results for tournament dynamics in place of relay.

a



b



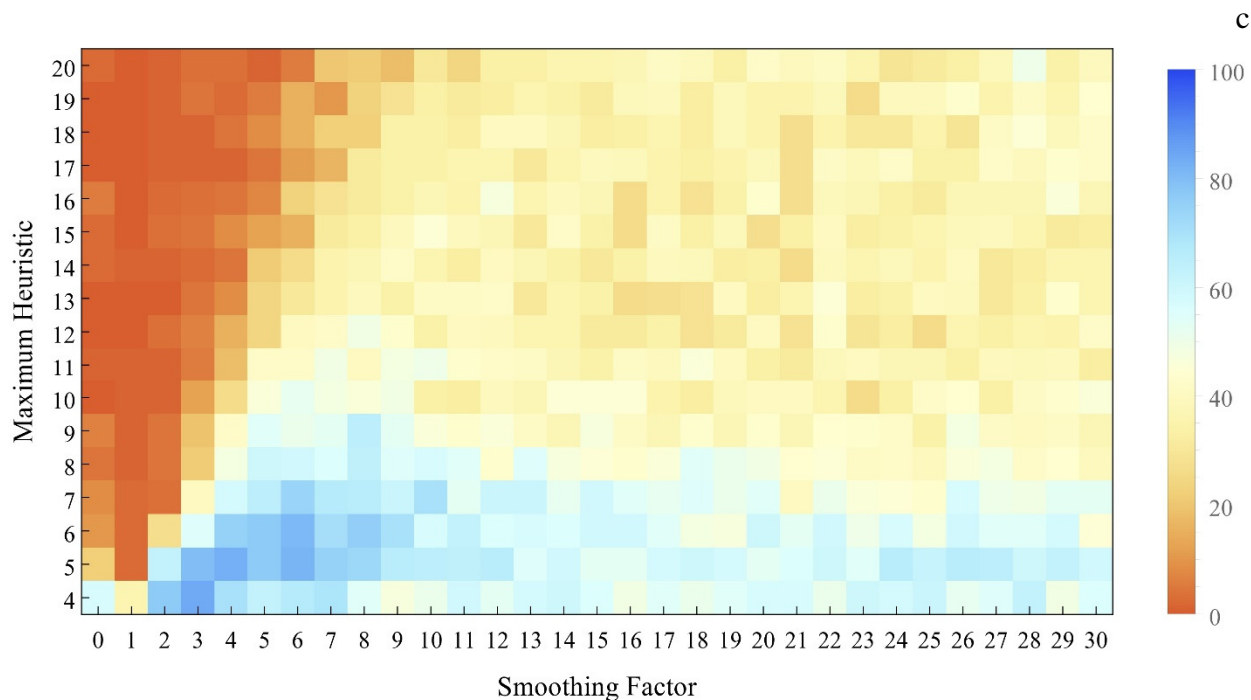


Figure 8. Corresponding to Figures 7 through 9, parameter sweeps across smoothing factor and heuristic pool size with tournament rather than relay group dynamics. (a) Areas in which groups of random (brown) and groups of best-performing (blue) do best. (b) Differences in averages over 100 runs, with positive values (green) showing advantage to the best-performing, negative values (yellow and red) for random groups. (c) Percentages of runs in which each group does better, with blue values for best-performing, red and yellow for random heuristics.

For a maximum heuristic over 10, ‘tournament’ rather than ‘relay’ updating gives a strong advantage to random heuristics. Indeed one might almost say that a group dynamics characterized by simultaneous tournament rather than sequential relay reverses the advantage to experts offered by increasing landscape smoothness.

Interesting results occur when we compare not merely groups (a) entirely of experts or (b) entirely of random heuristics, but also groups with some proportion of each. The difference between tournament and relay group dynamics also plays out on these mixed groups. Returning to our initial pool of 12 heuristics, we show results for relay dynamics in Figure 9.

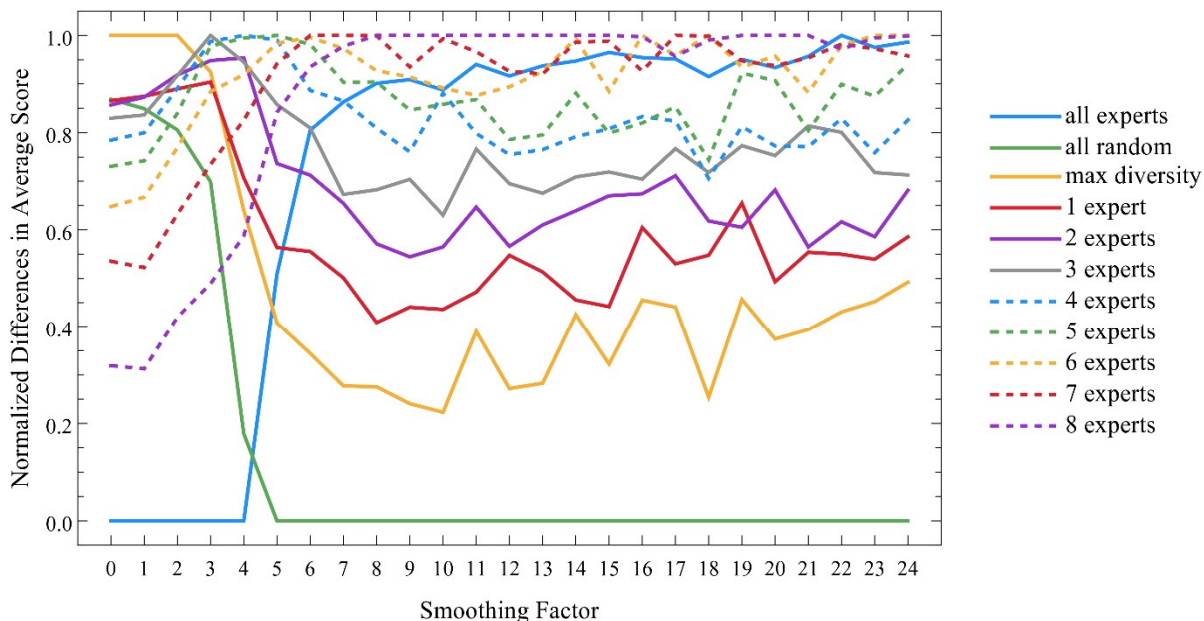


Figure 9. Normalized averages for pure and mixed groups of size 9 using a relay group dynamics

In Figure 9 that group that scores the highest average at any smoothness factor is normalized to 1, that which scores the lowest is normalized to 0. The dramatic cross-over between experts and randoms thus plays out in the cross-over between smoothness 5 and 6. For 5 and below it is experts that do the worst; for 6 and above it is the random group that performs worst.

We track not merely groups of experts and random heuristics but others as well. Our analysis of the Hong-Page result, consistent with indications in their work, is that what groups of the ‘best-performing’ have going against them is redundancy. They are too much alike, thereby losing the exploratory spread of individuals selected randomly. If that analysis is right, a ‘group’ consisting of a single best agent should have even higher redundancy and so should do worse. That is indeed the result for the range shown: a single best agent does so far worse than any of those groups shown in Figure 9 that we left it off rather than distort the readability of the chart.²

² Even this can be complicated by additional factors. In relay dynamics (though not simultaneous), with a heuristic pool of 24 and smoothness factor between 20 and 30 the single best expert outperforms the random group.

Contrary to Thompson 2014, we have suggested that it is not the randomness of random groups that is an epistemic virtue, but the extent to which their heuristics jointly cover the available space. If that analysis is right, a ‘maximum diversity’ group should do better than a random group. A group with maximum diversity might be constructed by choosing a random heuristic for the first member but then constraining successive choices for later members of the group so as to duplicate heuristic numbers as little as possible. If the number 11 is already ‘taken’ in the assignment of heuristics to three previous group members, for example, it is removed as a candidate for further assignments. Figure 9 shows the performance of such a group in gray, significantly outperforming a group of random agents at every point.

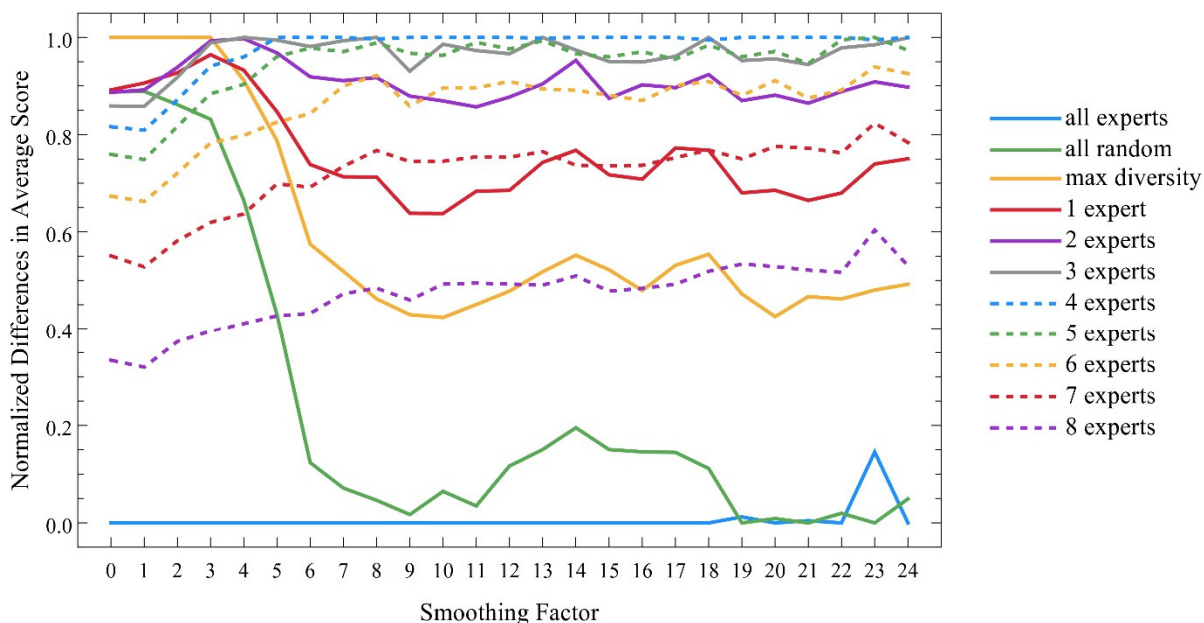


Figure 10 Normalized averages for pure and mixed groups of size 9 using a tournament group dynamics

The other groups shown are composed of 1 expert with 8 random, 2 experts with 7 random, and so forth. For smoothness factors above 8 the pattern is clear. Even 1 expert among randoms does better than our max diversity group. Groups with 2 experts do better still, and the

‘added value’ of experts in place of randoms increases to roughly 7 or 8, though groups with 1 or 2 random members still do better than groups composed of experts alone. The difference between tournament and relay dynamics is clear when we construct a similar group for tournaments results, shown in Figure 10.

It remains true in tournament play that a single expert does so far worse than others that he is left out in order to avoid distorting the chart. It also remains true that a ‘maximum diversity’ group outperforms a random group at all points.

In tournament dynamics, however, groups of 9 experts do far worse than in relay dynamics, occupying the normalized bottom of the chart at almost all points. Random groups do slightly better than experts at most points, though they tend to join them at the bottom with higher smoothing factors.

The performance of mixed groups is particularly interesting and importantly different than in the case of relay dynamics. In tournament dynamics, mixed groups with at least one expert and at least one random agent do better than either pure experts or pure randoms at all points. If one traces groups with 1 expert, 2 experts, 3 experts, or 4, these score progressively higher values in the graph. From that point, however, increasing the percentages of experts proves a disadvantage: groups with 5, 6, 7, or 8 experts do worse than those with 3 or 4. We have found very similar results showing advantage to roughly half-mixed groups with heuristic pools of 24 in place of 12.

Group dynamics makes an important difference in the relative value of diversity and expertise. In a nutshell, tournament dynamics favors random groups over experts but emphasizes the value of mixed composition groups over either.

VI Group Size

Are there other parameters that advantage either diversity or expertise? In stating their original result, Hong and Page require that groups be ‘good sized’; the groups of 9 we have used throughout are very close to the groups of 10 they use in simulation. In limited confirmation of that requirement, it appears that larger groups—at least up to a limit—favor groups of random heuristics over groups of experts. Results hold for both relay and tournament dynamics.

We replicated the smoothing factor and max heuristic graphs of the previous sections for groups of 3 and 6 rather than 9. Previous results for 9 are included for comparison. Up to groups of 9, it is clear that increased group size in relay dynamics advantages the relative performance of groups of random heuristics over groups of experts. Figures 14 through 16 show similar comparisons for simultaneous tournament dynamics.

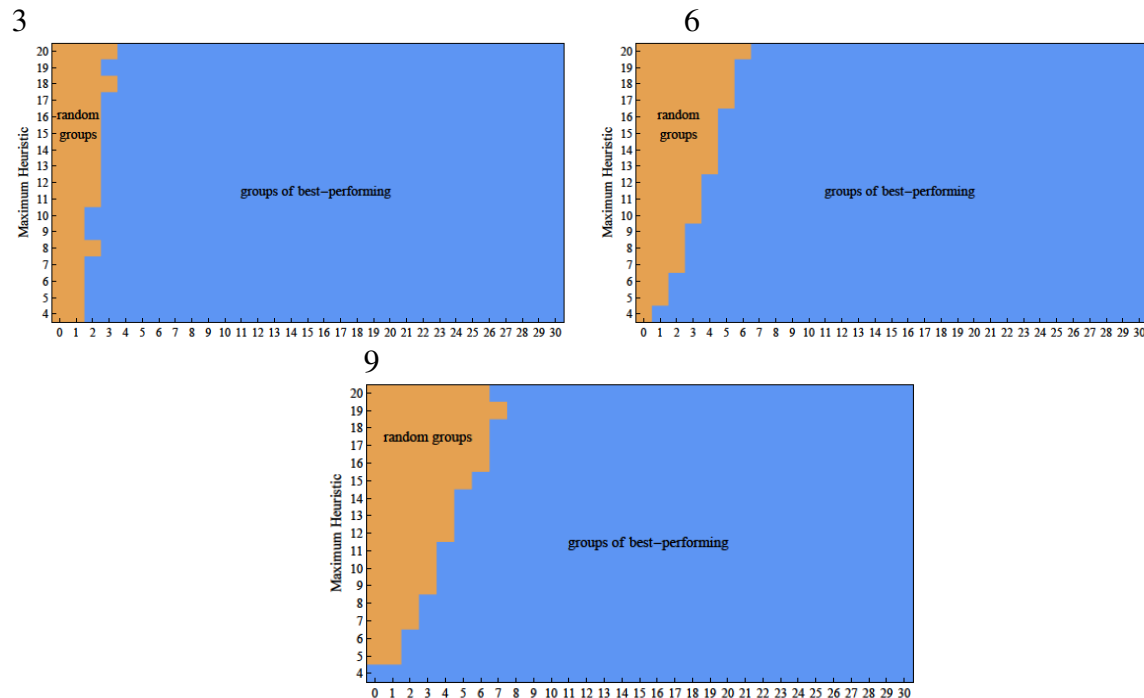


Figure 11 Relay: areas in which groups of random heuristics do best (brown) and areas in which groups of the best-performing do best (blue) across a parameter sweep of landscape smoothness and max heuristic for groups of 3, 6, and 9

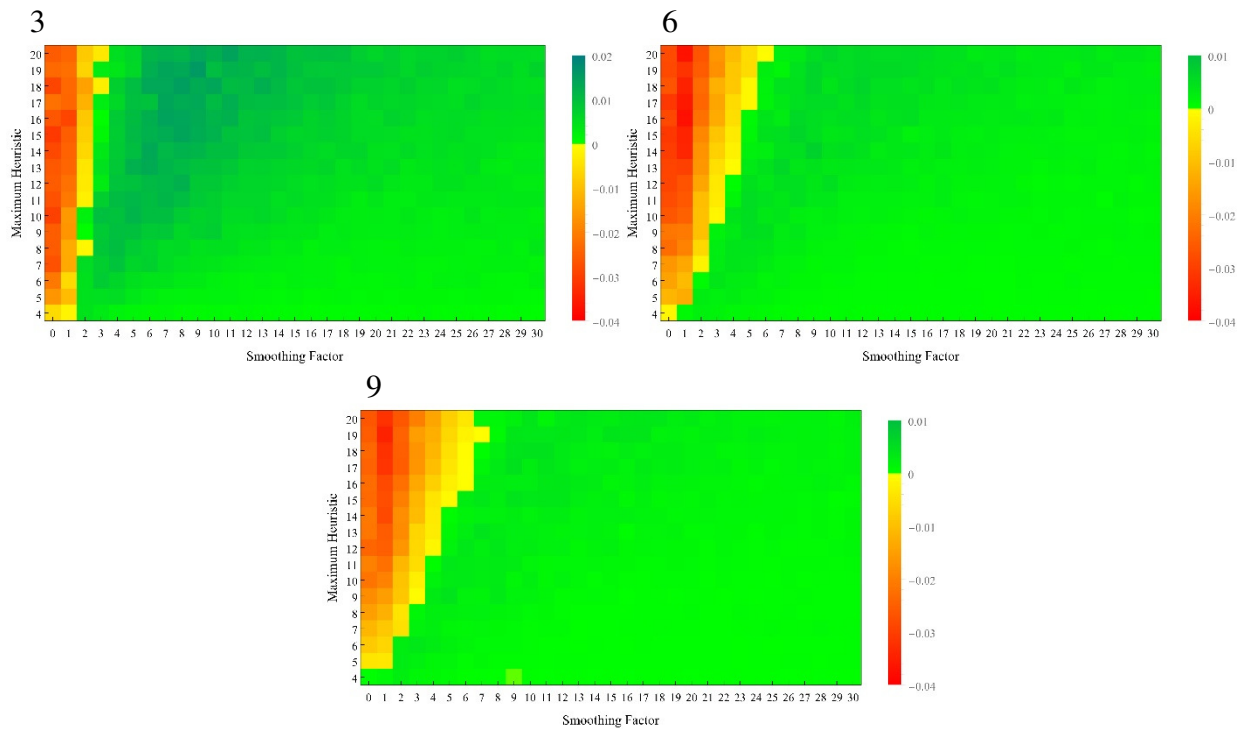


Figure 12 Relay: Differences in averages for groups of random heuristics and groups of the best-performing for groups of 3, 6, and 9.

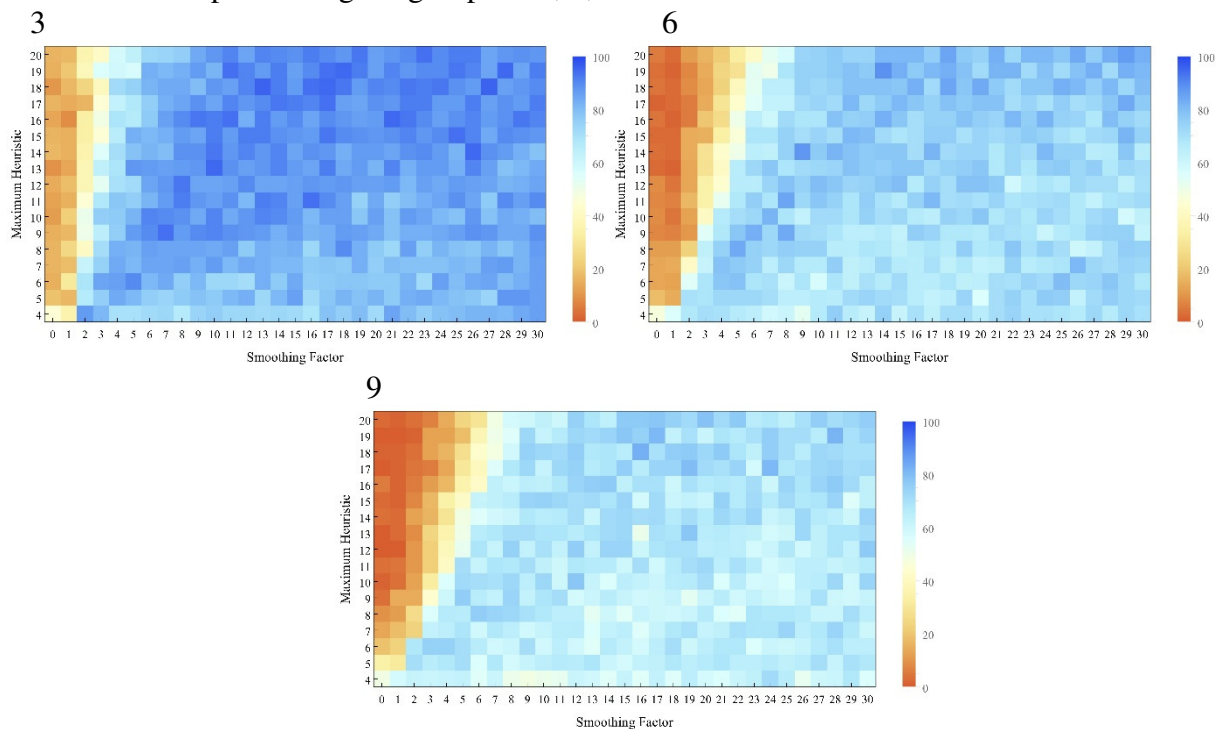


Figure 13 Relay: Percentages of runs in which groups of the best-performing do better than groups of random heuristics.

3

6

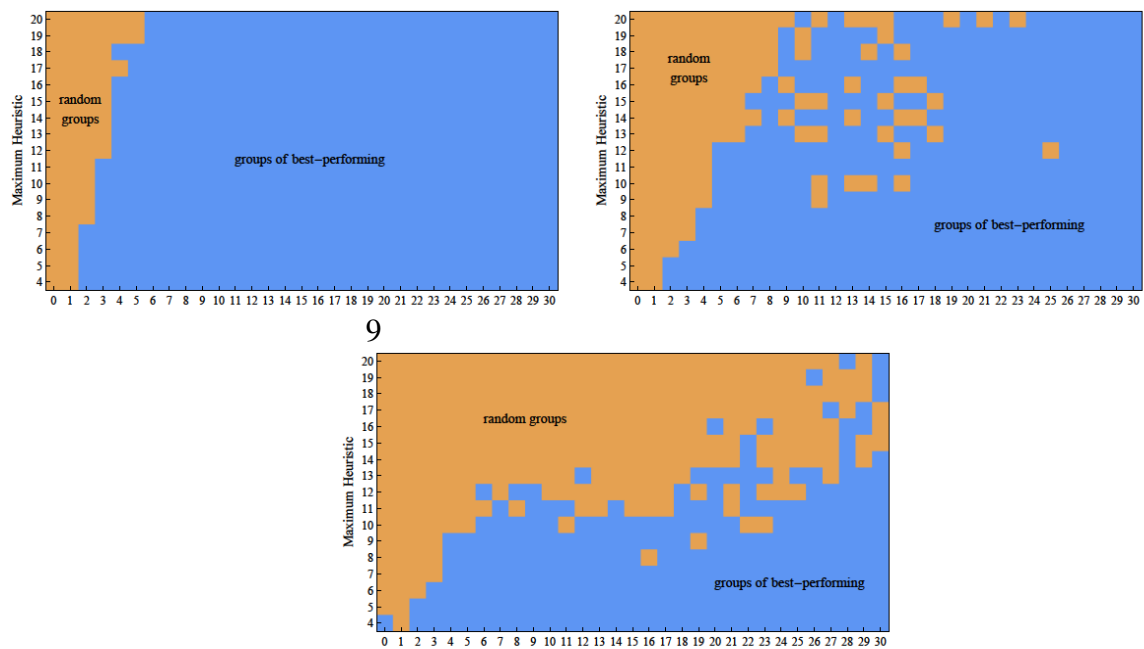


Figure 14 Tournament: areas in which groups of random heuristics do best (brown) and areas in which groups of the best-performing do best (blue) across a parameter sweep of landscape smoothness and max heuristic for groups of 3, 6, and 9

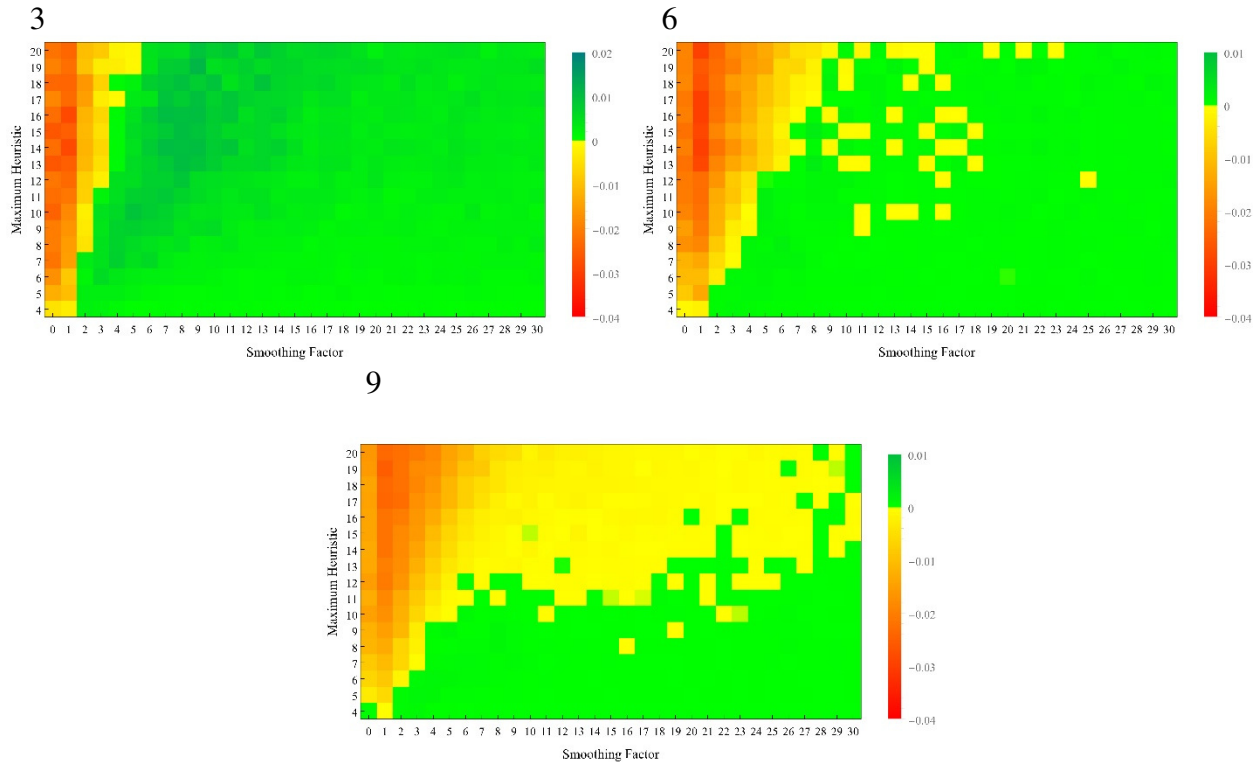


Figure 15 Tournament: Differences in averages for groups of random heuristics and groups of the best-performing for groups of 3, 6, and 9.

3

6

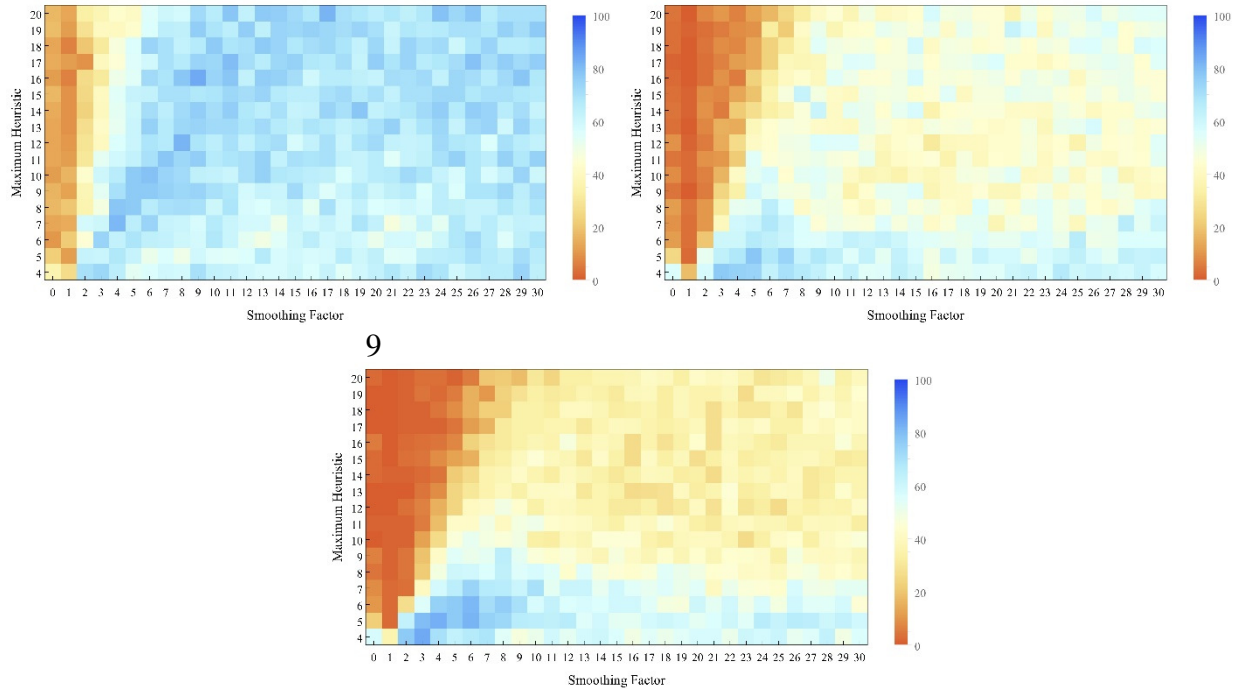


Figure 16 Tournament: Percentages of runs in which groups of the best-performing do better than groups of random heuristics.

The smaller the group, these figures indicate, the greater the advantage of expertise. The larger the group, all things considered, the greater the advantage for random heuristics. Here again heuristic coverage offers an explanation. Larger groups of experts will have a small reduction in redundancy, since our experts are genuinely distinct, but larger groups of random heuristics can still be expected to outstrip the best-performing in terms of coverage on the available heuristic set. At least one reason why larger groups favor random heuristics over expert heuristics is because the ratio of coverage for randoms over experts increases with group size.

VI. Conclusion

Our results indicate that the slogan ‘diversity trumps ability’ can easily be overstated, and that some of the knocks taken by expertise are undeserved. With other parameters patterned on the original Hong-Page simulation, we’ve shown that ‘diversity trumps ability’—that random groups

outperform groups of the best-performing—only within a small window of low landscape smoothness toward the random end. Within roughly that same that window, moreover, the success of the best-performing heuristics on a specific landscape is limited to that specific landscape: success on one random landscape tends not to correlate with success on another. Genuine ability or expertise, on the other hand, would seem to demand transportability from one landscape to another. We therefore warn against interpreting a heuristic’s success on these relatively random landscapes as ‘ability’ or ‘expertise.’ Within the other original parameters, ‘diversity trumps ability’ or ‘diversity trumps expertise’ only where it is unclear that what it trumps should really be considered ability or expertise.

For ‘smoother’ and arguably more realistic problem landscapes, again with other parameters the same, correlation from one landscape to another jumps for the ‘best-performing.’ Here successful heuristics have a better claim to be modeling ability or expertise, and here that expertise shows its value. For landscape smoothness above 4, using the Hong-Page relay dynamics, groups of experts outperform groups of random heuristics. 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. Increase in the pool from which heuristic numbers are drawn increases the advantage for random groups. Given a landscape smoothness factor at which groups of experts do better with a given set of available heuristics, groups of random heuristics perform better once we increase the conceptual space to a larger 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 have found a major difference between the two. In an arguably more realistic ‘tournament’ dynamics, agents

deliberate and navigate a problem landscape with simultaneous suggestions from the floor rather than in a round-the-table ‘relay.’ A tournament group-dynamics, it turns out, further favors the value of diversity. Many of the points at which groups of experts show an advantage within a relay dynamics disappear in favor of groups of random heuristics once the dynamic is changed to a simultaneous tournament. We find it particularly intriguing that within an arguably more realistic tournament dynamics, across arguably more realistic problem landscapes, it is neither pure groups of experts or pure groups of random heuristics that do best but mixed groups including both.

The variety and sensitivity of these results undercut any uncritical application of the Page-Hong result in favor of diversity initiatives in all contexts. While such policies may be supported by social justice considerations, it is the positive impact on educational quality that is offered as a primary consideration in the Supreme Court’s ruling on the matter (*Grutter v. Bollinger*, 2003), and the Hong-Page result is often cited as part of the small body of evidence in support of diversity (e.g. UCLA, 2014; Kreiter, 2013). What our results indicate is that diversity does not always trump ability. Policy makers must also consider aspects of the problem set at issue and the decision procedures to be employed.

Relative to problem characteristics, conceptual resources, and group dynamics, the wisdom of crowds and the wisdom of the few each have a place. We’ve attempted to offer some first modeling steps in understanding the role each has to play.

References

- Anderson, Elizabeth (2006). The Epistemology of Democracy. *Episteme* 3.1, 8-22.
- Bright, L. K. (2016). Decision Theoretic Model of the Productivity Gap. *Erkenntnis*, 1-22.
- Brief for Lt. Gen. Julius W. Becton, Jr., Gen. John P. Abiziad, Adm. Dennis C. Blair, Gen. Bryan Doug Brown, Gen. George W. Casey, Lt. Gen Daniel W. Christman, Gen. Wesley K. Clark, Adm. Archie Clemins, Gen. Ann E. Dunwoody, Gen. Ronald R. Fogleman, Adm. Edmund P. Giambastiani, Jr., et al., as Amici Curiae in Support of respondents. *Fisher v. University of Texas*, Austin, 2016 (no. 14-981)
- Condorcet, M. J. A. N. d. C., Marquis de. (1785/1995). An Essay on the Application of Analysis to the Probability of Decisions Rendered by a Plurality of Votes (Fifth Part). In I. McLean & A. B. Urken (Trans. & Ed.) *Classics of Social Choice*, pp. 91-112. Ann Arbor: University of Michigan Press.
- Gunn, Paul (2014) "Democracy and Epistocracy." *Critical Review*26:1-2, 59-79.
- Grutter v. Bollinger*. 2003. 539 U.S. 306.
- Hong, Lu & Scott E. Page (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*. 101, 16385-16389.
- Hong, Lu & Scott E. Page (2009). Interpreted and Generated Signals. *Journal of Economic Theory* 144, 2174-2196.
- Kreiter, Clarence D. "A Measurement Perspective on Affirmative Action in U.S. Medical Education." *Medical Education Online* 18 (2013): 10.3402/meo.v18i0.20531. *PMC*. Web. 17 Dec. 2015.
- Landemore, Hélène (2013). *Democratic Reason: Politics, Collective Intelligence, and the Rules of the Many*. Princeton, NJ: Princeton University Press
- Martini, C. (2014). Experts in science: A view from the trenches. *Synthese*, 191(1), 3-15.
- Page, Scott E. (2007) *The Difference*. Princeton University Press.
- Page, Scott E. (2011) *Diversity and Complexity*. Princeton University Press.
- Shanteau, James (2000), "Why Do Experts Disagree?" In B. Green, R. Cressy, F. Delmar, T. Eisenberg, B. Howcroft, M. Lewis, Do Schoenmaker, J. Shanteau & R. Vivian, eds., *Risk Behavior and Risk Management in Business Life*, Dordrecht: Kluwer.

Shanteau, James (2002) "Domain Differences in Expertise,"
http://www.researchgate.net/profile/James_Shanteau/publication/228710111_Domain_differences_in_expertise/links/0c9605181938feeb93000000.pdf

Shanteau, James, David J. Weiss, Rickey P. Thomas, Julia C. Pounds (2002). "Performance-based assessment of expertise: How to decide whether someone is an expert or not." *European Journal of Operational Research* 136, 253-263.

Shanteau, James, David J. Weiss, Rickey P. Thomas, Julia C. Pounds (2003) "How Can You Tell if Someone Is an Expert? Performance-Based Assessment of Expertise." In Sandra L. Schneider & James Shanteau, eds., *Emerging Perspectives on Judgment and Decision Research*. Cambridge: Cambridge University Press, pp. 620-640.

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

Surowiecki, James (2004) *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. New York: Doubleday Anchor.

Tetlock, Philip E. (2005). *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton, NJ: Princeton Univ. Press.

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

Thompson, C. (2013). A general model of a group search procedure, applied to epistemic democracy. *Synthese*, 190(7), 1233-1252
 UCLA (2014). Proposed Diversity Requirement. Retrieved from
https://ccle.ucla.edu/pluginfile.php/743624/mod_resource/content/6/082014%20REVISED%20DiversityReqProposal.pdf

Weiss, David J. & James Shanteau (2014) "Who's the Best? A Relativistic View of Expertise." *Applied Cognitive Psychology* 28 (4) 447-457.

Weymark, John A. (2014) "Cognitive Diversity, Binary Decisions, and Epistemic Democracy," Vanderbilt University Center for the Study of Democratic Institutions, Working paper 5-2014.