Historical Social Research
Historische Sozialforschung

Bennett Holman, William J. Berger, Daniel J. Singer, Patrick Grim &
Aaron Bramson:

Diversity and Democracy:
Agent-Based Modeling in Political Philosophy.

Diversity and Democracy: Agent-Based Modeling in Political Philosophy

*Bennett Holman, William J. Berger, Daniel J. Singer,
Patrick Grim & Aaron Bramson* [*]

Abstract: *»Diversität und Demokratie: Agent-Based Modelling in der politischen Philosophie«*. Agent-based models have played a prominent role in recent debates about the merits of democracy. In particular, the formal model of Lu Hong and Scott Page and the associated "diversity trumps ability" result has typically been seen to support the epistemic virtues of democracy over epistocracy (i.e., governance by experts). In this paper we first identify the modeling choices embodied in the original formal model and then critique the application of the Hong-Page results to philosophical debates on the relative merits of democracy. In particular we argue that the "best-performing agents" in the Hong-Page model should not be interpreted as experts. We next explore a closely related model in which best-performing agents are more plausibly seen as experts and show that the diversity trumps ability result fails to hold. However, with changes in other parameters (such as the deliberation dynamic) the diversity trumps ability result is restored. The sensitivity of this result to parameter choices illustrates the complexity of the link between formal modeling and more general philosophical claims; we use this debate as a platform for a more general discussion of when and how agent-based models can contribute to philosophical discussions.

Keywords: Epistemic democracy, epistocracy, expertise, agent-based models, representation, deliberation, cognitive diversity, problem solving, democracy, wise crowds.

[*] Bennett Holman, Underwood International College, 430 Veritas Hall B, 85 Songdo-gwahak-ro, Yeonsu-gu, Incheon, Korea 406-840; bholman@yonsei.ac.kr.
William J. Berger, Department of Philosophy, University of Pennsylvania, 433 Cohen Hall, Philadelphia, PA 19104-6304, USA; zberger@sas.upenn.edu.
Daniel J. Singer, Department of Philosophy, University of Pennsylvania, 433 Cohen Hall, Philadelphia, PA 19104-6304, USA; singerd@phil.upenn.edu.
Patrick Grim, Center for Complex Systems, University of Michigan, Weiser Hall Suite 700, 500 Church Street, Ann Arbor, MI 48109-1042, USA; patrick.grim@stonybrook.edu.
Aaron Bramson, Laboratory for Symbolic Cognitive Development, RIKEN Brain Science Institute, BSI West Building 2F, 2-1 Hirosawa, Wako City, Saitama 351-0198, Japan; bramson@brain.riken.jp.

> The view that it is more proper for the multitude to be sovereign than the few of greatest virtue might be thought to be explicable and to have some justification, and even to be the true view. For it is possible that the many, though not individually good men, yet when they come together may be better, not individually but collectively, than those who are so, just as public dinners to which many contribute are better than those supplied at one man's cost; for where there are many, each individual, it may be argued, has some portion of virtue and wisdom, and when they have come together, just as the multitude becomes a single man with many feet and many hands and many senses, so also it becomes one personality as regards the moral and intellectual faculties. (Aristotle Pol 3.1281a-b)

1.   Introduction

A recognition of the 'wisdom of the crowds' has been a source of optimism for democratic theorists, suggesting democracy has a competitive advantage over alternative regimes. Some political philosophers have made agent-based modeling (ABM) central to this discussion with their claim that Lu Hong and Scott Page's (2004) "diversity trumps ability" result (DTA) formally indicates the epistemic virtues of democracy (Anderson, 2006; Gaus 2016; Landemore 2013a, 2013b; Schwartzberg 2015). But there has also been pushback to this move, questioning whether the Hong-Page results have such an implication (Ahlstrom-Vij 2012; Weymark 2015). Some have argued that the Hong-Page results are compatible with (or positively support) a form of epistocracy, where the knowers or experts dictate the policy of the state (Ancell 2016; Brennan 2016).

Here we attempt to go deeper into the models crucial to the debate, but also attempt to go beyond those specifics in order to reflect more generally on the promise and pitfalls of attempts to draw philosophical lessons from agent-based models. We present our own extensions to the Hong-Page model, illustrating with that example how such an application of ABMs can make philosophical argumentation more rigorous. In Section 2, we discuss the Hong-Page results, describing both the dynamics of the ABM and the modeling decisions therein. In Section 3, we argue that interpreting the Hong-Page results as pertaining to predictive expertise is implausible. We identify test-retest reliability as a minimal criterion of predictive expertise and show that the "best-performing agents" in the Hong-Page model fail to meet such a criterion. In Section 4 we explore a set of extensions to the original model, showing that the Hong-Page 'Diversity Trumps Ability' result fails to hold in parameter subspaces in which claims of predictive expertise are more plausible. In Section 5 we explore the parameter sensitivity of this result and show that the DTA result reemerges with other suites of parameters.

The remainder of the paper pivots from formal models to the wider question of philosophical interpretation and social application. Section 6 provides an

illustration of how such models are used in political philosophy, and how extensions such as ours might reflect on the limited generality of such arguments. We use Helene Landemore's *Democratic Reason* (2013a) as a case study due to its influential application of ABMs to questions in political philosophy. In that regard, we show how our extensions offer further refinements to the conjectures presented in her work. Finally, Section 7 argues for the importance of carefully identifying modeling assumptions and relevant extensions, particularly when leveraging ABMs in practical literatures such as political philosophy. We conclude by suggesting what further steps are needed to bridge the gap between the two.

2. Unpacking the Hong Page Result

The Hong-Page model consists of three important components: (1) Problems and problem characteristics (epistemic landscapes); (2) Agents; (3) Individual problem solving; (4) Group formation; and (5) Group problem solving dynamics. In the model, sets of agents navigate an epistemic landscape in an attempt to identify the landscape's highest point (the global maximum). In this section, we examine the model in two passes. First, we explain these five features of the Hong-Page model dynamics and we explore the modeling choices made in each case. We will return to discuss the importance of these modeling choices later in the paper.

2.1 Model Dynamics

Problems and problem characteristics (epistemic landscapes): The 'problem' that the agents are trying to solve is represented through the epistemic landscape. The landscape consists of a set of two thousand points arranged in a ring such that the last point is adjacent to the first – once agents reach the 2,000th spot, they loop back to the first. In the original model, each point is randomly assigned a value between 0 and 100. Points with higher values are understood as better solutions to the problem.
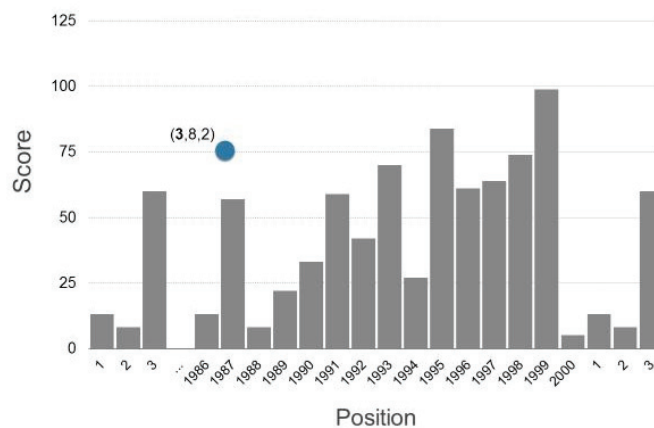
Agents: Each agent is assigned a random 'heuristic,' an ordered set of three natural numbers between one and twelve (e.g. [8,2,9]), which dictates how the agent moves around the landscape. The 'maximum heuristic value' is the greatest number that can appear in a heuristic. There are 1,320 possible heuristics using an ordered set of 3 unique numbers from 1 to a maximum heuristic value of 12.

Individual problem solving: Problem solving is represented as movement across the landscape in search of the best solution (i.e. the point with the highest value). At any given moment an agent occupies a space with some value, but is looking to move to a better solution. To decide whether to move, the

agent uses her first heuristic value, say '5,' to probe the point five spaces to the right. If that point has a higher value than the one she currently occupies, she moves to it and repeats this procedure from her new spot using the next value in her heuristic. If the probed space has a lower value she stays on the original spot and tries her next heuristic value. She stops exploring the epistemic landscape when she cannot reach a higher point with any of the elements in her heuristic.

For example, consider the agent represented in Figure 1. The agent finds herself on point 1987 with the heuristic (3,8,2). The agent first uses the '3' to check point 1990, but stays because the point is lower. She next uses the '8' to check point 1995, which she then moves to because it is higher. From point 1995 she next uses the '2' to check point 1997, then the '3' to check 1998 and '8' to check point 3. Because all of them have lesser values, she halts on point 1995. The final value for the agent starting from point 1987 is thus 82, the height of point 1995.

Figure 1: An Example of One Agent on a Hong-Page Epistemic Landscape

Group formation: Hong and Page create two different groups. The first is created by taking the nine best-performing agents. To determine the best-performing agents, a landscape is generated and a score is determined for the first agent by starting her from each of the 2,000 points on the landscape and averaging the values obtained at every corresponding stopping point. This is repeated for each of the 1,320 possible heuristics. Agents are then ranked according to their average performance, with higher-scoring agents interpreted as having greater ability. The "high ability" group is composed of the nine agents with the highest ability (i.e. the highest average score on the landscape). A second group is composed simply by randomly selecting nine from the pool of all agents, irrespective of individual ability.

Group problem solving dynamics: Both groups of nine agents move around the landscape as in a relay race. When the first group member gets stuck, the second agent tries out their heuristic, continuing until they get stuck, and so on. The group halts at the point at which no agent can identify a better solution. The 'Diversity Trumps Ability' result is the finding that groups of randomly selected agents regularly outperform groups composed of the nine individually best-performing agents (i.e. the group of randomly selected agents finds points with higher values).

2.2   Model Conditions

Hong and Page (2004, 16387) explicitly identify four modeling assumptions: that the agents are minimally intelligent, that the agents are diverse, that the best agent is unique, and importantly for our analysis, that the problem is difficult.[1] But their epistemic landscape itself embeds a number of assumptions as well. The epistemic landscape is a discrete, one-dimensional domain in which each spot has a corresponding value drawn from a uniform distribution. But of course this is not the only way of generating epistemic landscapes. For example, many problem solving situations will be best modeled by multi-dimensional landscapes.[2] Even on a single dimension, the function used to generate the values at each point embodies a modeling choice. One important property of the original model is the independence of adjacent points. The value of an agent's current position on the landscape contains no information about points in the immediate vicinity. The best solution in the landscape is just as likely to be proximate to the second best solution as it is to be proximate to the worst.

The ordering of points is standardly interpreted to capture how well the agent(s) understand the problem they are attempting to solve. A perfect understanding of the problem would create a monotonic ordering of values from lowest to highest. Landscapes that "fail to embed understanding […] result in very rugged landscapes" (Page 2007, 140). The landscape used by Hong and Page is at the latter end of the spectrum as the value of each point is independent of the point before it. As such, the extremely rugged landscape used by Hong and Page represents problem solving in a case where none of the agents possess any understanding of the problem.

The agents, as well as the landscape, embed assumptions. It is worth noting that all agents possess a heuristic vector with the same number of elements.

---

[1]   Hong and Page classify a problem as "difficult" when no individual problem solver regularly finds the global maximum (i.e. where working in groups has the potential to improve on individual performance, cf. Page 2007, 159).

[2]   For one alternative landscape see Alexander, Himmelreich and Thompson (2015) or Fontanari and Rodrigues (2016) for their use of NK-landscapes that allows both higher dimensionality and more "rugged" problem spaces.

Page (2007, 145) notes that, for example, people who score higher on IQ tests likely have more tools (i.e. elements in their heuristic).[3] We could, therefore, consider agents with heuristics of different sizes (i.e. some heuristics with three elements and others with more).

Agents use a "gradient heuristic" (see Page 2007, 62) meaning that agents are always able to accurately determine whether the next step they take is of greater value and never move to a lower value, either accidently (because they misjudge its value) or purposely (to further explore the landscape). Yet, as Page (2007) notes there are other classes of heuristics that agents might use to move across a landscape, including some that allow agents to explore initially less promising paths to see if they ultimately lead to higher ground (e.g. simulated annealing), to assess multiple possible solutions simultaneously, or to restrict movement only to adjacent spaces. Agents are also limited in their ability to anticipate the results of successive application. Unlike a skilled chess player, they can only see "one move ahead."

This short list suffices to give a sense of the way in which the Hong-Page model encodes a particular problem situation. In and of themselves, of course, explicit modeling choices offer no grounds for critique. Any model simplifies reality, excluding a number of features. The best models include what is relevant and leave off what is not; articulating what is not included in the model serves primarily to describe rather than critique. If such factors turn out to be unimportant, their absence in the model is a virtue. We mention details here to give the reader some appreciation for both what is in the model and what has been omitted – a point we will return to in the conclusion in discussing directions for future research. In the next sections we examine the consequences of varying some of these modeling choices and show that the DTA is sensitive to at least some of the original modeling decisions.

3.   Interpreting Hong-Page: Expertise

What can be learned from these modeling results? The headline interpretation is that 'diversity trumps ability,' with the implication that a random collection of skills or approaches will yield better outcomes than a cultivated collection of skills. The core result, more carefully stated, is that for a given type of difficult problem, groups of randomly chosen agents regularly outperform groups of "better performing" agents. But the DTA result has also been taken to imply that a random selection of agents will outperform groups of *experts* (e.g., An-

---

3   Page is quick to point out this does not mean they are "smarter" just that they have more tools applicable to standardize tests. One of their primary points is that ability is context specific and thus high performance on standardized tests need not translate to other settings where people with low test scores might outperform "smarter" individuals.

derson 2006; Gunn 2014). For example, Farrelly argues that Hong and Page show that "the central failure […] of 'epistocracy' stems from the fact that it equates collective ability with the individual ability of experts working in isolation" (2012, 17). We'll argue that this terminology isn't warranted, however, at least in terms of the original model. In explaining why, we will motivate a model extension that is more plausibly interpreted in terms of expertise.

"Best performing" is an uncomplicated measure of success in the Hong-Page model: it refers to those agents that score the highest, on average, on a given landscape. In at least one intuitive sense, true 'expertise' requires something more. As Goldman and Blanchard (2016) put it, "[a]n expert in any domain will know more truths and have more evidence […] , and these things can be used to form true beliefs about new questions concerning the domain." In this sense of expertise, which we'll label 'predictive expertise,' we should expect the success of an expert on one problem (landscape) to be correlated with their success on a new problem (landscape). When there is no such correlation, the better performance on a single landscape fails to meet this minimal condition. It is still the case that some agents do better than others, but 'best performance' in this case might better be viewed as a result of randomness or luck than any general ability or transportable expertise. Focusing on predictive expertise, we'll investigate how and whether the Hong-Page Diversity Trumps Ability ('DTA') result can be generalized to a Diversity Trumps Expertise ('DTE') result.

Consider Robert Hodgson's (2008) assault on expert wine tasters. To determine if expert wine ratings were truly tracking the quality of the wine, he inserted multiple pourings of the same wine into the tastings at the California State Fair Wine Competition (the oldest commercial competition in the US). On a twenty-point scale, the median range for test-retest reliability was about 4 points. Only 10% of the judges in a given year were "consistently consistent" in that they routinely rated wines in the same medal category (a four-point spread) each time. These results might be taken to indicate that only a small percentage of even the *crème de la crème* are true predictive experts. However, by tracking judges across years, he showed that a judge's consistency one year does not predict the judge's consistency in the following year. In other words, the judges that appeared to be consistently consistent were just lucky, no more consistent than other judges. In subsequent work Hodgson (2009) analyzed wines that had been entered in multiple competitions and found that the probability of winning gold in one event was probabilistically independent of winning gold at another competition.

Examples like this could be multiplied, but the upshot is the same: predictive expertise is demonstrated by repeated success.[4] Applying these lessons to the Hong-Page model, we can ask how we might distinguish luck from skill. One thing that will not work is rerunning the same agents on the same landscape (i.e. one with the same values for each spot). Since the process is determinate, the outcomes of a literal test-retest will always be the same. Alternatively, we might create an artificial test-retest measure by dividing the landscape into two halves and asking to what extent performance on the front half correlates with performance on the back half. As it turns out, the answer to this is: not very much. A Pearson correlation of such a test-retest found that performance on the first half of the landscape only predicted 1% of the variance of the agent's performance in the second half – leaving the remaining 99% unexplained. If expertise requires the ability to repeat successes on the same kind of problem, these best-performing agents are not 'experts'– at least not in the predictive sense.

On these grounds, we claim that the "best-performing agents" in the original Hong Page model should not be interpreted as modeling predictive expertise or the performance of experts. A better interpretation of the result might not be 'diversity trumps ability,' with its tempting generalization to expertise. The better interpretation might be '*In problems where there are no real experts, it is a diverse group that typically does best.*'

4. Hong–Page Model Extension: Correlated Landscapes

In an attempt to better model predictive expertise, we need to construct a problem space in which individual performance is correlated across different runs of the model. Above we noted that in the original model, the value at each point in the landscape is uncorrelated with the next, thus modeling a situation in which no agent has any understanding of the problem. It is perhaps unsurprising that such problems don't allow for the emergence of expertise. One way to induce correlation in the landscape would be to assign a new value to every other point, or every third point, or every 10th point, and then fill in the intermediate points with intermediate values.

Below we adopt a similar, though slightly more sophisticated version of the same idea by introducing a "smoothing factor" (SF). As in the original model, we assign a random value to the first point in the landscape. Instead of repeating this for every point, we next randomly choose an integer between 1 and 2n+1, where n is the smoothness parameter. We then move forward that many

---

[4] This is the same intuition embedded in hypothesis testing. We know that two sample means differ, what we want to know is how confident we can be that the difference was not due to mere chance.

spaces on the epistemic landscape, assign *that* point a new random value and assign intermediate values to points in between. As n increases, the landscape becomes less random. More crucial for our purposes, as seen in Table 1, when the landscape becomes less random, the test-retest reliability of agents emerges (and does so quite quickly). The test-retest correlations indicate that at a smoothness of 1, the performance in the first half of the landscape explains roughly 60% of the performance in the second half and between 85% and 92% of the variance at smoothness values between 2 and 20.[5]

Keeping other parameters the same, we reran the simulation on 100 landscapes at each smoothness to compare the average performance of groups of best-performing agents and randomly selected groups. Table 1 shows that the DTA result fails to hold for smoothness values greater than 3.[6] For the majority of the parameter settings considered here, where the result supports an interpretation of ability or expertise, groups of the best-performing agents tend to outperform randomly selected groups.

Table 1: Average Pearson Correlations Comparing Agent Performance between the Front and Back Halves of the Model over 100 Repetitions for All Heuristics at Landscapes of a Given Smoothing Factor

SF	Correlation	DTA	SF	Correlation	DTA	SF	Correlation	DTA
0	0.09	Y	7	0.96	N	14	0.94	N
1	0.78	Y	8	0.95	N	15	0.93	N
2	0.93	Y	9	0.96	N	16	0.93	N
3	0.95	Y	10	0.94	N	17	0.92	N
4	0.96	N	11	0.95	N	18	0.92	N
5	0.96	N	12	0.93	N	19	0.92	N
6	0.96	N	13	0.94	N	20	0.93	N

The Column label "DTA" records whether the result obtains. The original landscape is SF = 0.

5.   Assessing Robustness

Given other parameter settings in the original Hong Page model, Section 4 shows that the DTA result holds only for a limited range of landscape smoothness. There are other parameters that might also be expected to impact DTA results and in their original paper Hong and Page investigate a number of parameters to demonstrate that their result is robust. Here we adopt the same strategy to assess the robustness of our result and determine whether the DTA

---

[5] Percent of variance explained is calculated by squaring the correlation coefficient.

[6] At SF = 4 the groups are roughly equal and from 4-20 groups of experts are superior. For finer detail on these simulations, see the parameter sweeps in Figures 3 and 4 (group size = 9, max-heuristic = 12).

result only obtains in portions of the parameter space that fail to support an interpretation of predictive expertise. We conducted parameter sweeps across smoothing factors 0-30 for each of the following parameters: (a) for maximum heuristic values 4-20, (b) for groups of 3, 6, and 9, and (c) using two kinds of group dynamics: relay and tournament.

First, a word about the last of these. In the original model, as in our extension above, groups of agents used a relay dynamic. In the relay dynamic each agent moves until reaching her stopping condition, at which point she "passes the baton" over to the next agent who takes over, and so on until no agent in the group can find a better solution. In their original paper Hong and Page (2004) also consider a tournament dynamic in which every agent announces the point they would reach applying their heuristic from the group's current starting point. The group then moves to the highest point reachable by any agent and all of the agents again announce the point they would reach starting from this point. The group halts when no agent can move from the current starting point.

The results across parameter sweeps are shown in Figures 2 through 7. Figures 2 (relay dynamic) and 5 (tournament dynamic) display the information in its most basic form by recording whether randomly selected groups or groups of best-performing agents did better on average. Figures 3 (relay dynamic) and 6 (tournament dynamic) display the average difference between the groups and is thus sensitive to the degree to which one group is outperformed by the other. Figures 4 (relay dynamic) and 7 (tournament dynamic) ignore the difference between the paired samples and simply record what percentage of cases one group outperforms the other in a sample of 100 runs.

With respect to the interpretation of these results, the maximum heuristic, can be thought of as the total number of unique conceptual resources or skills that are available to agents. Each heuristic is a different potential move forward, but there is nothing intrinsically better about having heuristic 4 or heuristic 12. In the original model, Hong and Page (2004) considered both cases where the maximum heuristic was 12 and 20 and found qualitatively similar results. In contrast, as can be seen in each figure, as the maximum heuristic increases, so does the portion of the parameter space where DTA obtains.

A second parameter is group size. In stating their original result, Hong and Page require that groups be 'good sized'; the groups of 9 we use 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 some point– increase the portion of the parameter space that favors diversity. In groups of three the vast majority of the parameter space favors experts.
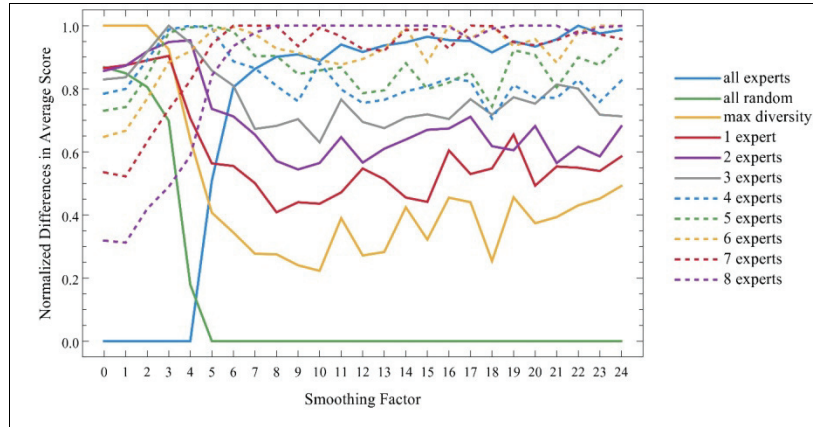
Our third parameter is group dynamics. While Hong and Page report no difference between the two dynamics on the original landscapes, we find significant divergence between the results, particularly as they interact with the other variables. In general, the tournament dynamics increase the portion of the parameter space in which the DTA result obtains. This is especially true for large

groups (n = 9) with at least a moderate maximum heuristic (roughly 10 or greater).

In addition to these parameter tests, we test group composition by altering the ratio of best-performing agents and randomly assigned agents in groups for both relay (Figure 8) and tournament (Figure 9) dynamics. The groups can contain all randomly assigned agents, all 'experts,' or some mixture in between. We also construct a group with maximum diversity by choosing a random digit for the first number in the heuristic of the first member, constraining successive choices for later members of the group so as to duplicate heuristic numbers as little as possible. For example, if the number 11 is already assigned to the heuristic of a previous group member, it is removed as a candidate for further assignments. Once every number has been assigned once, they are all made freshly available for the next assignment.
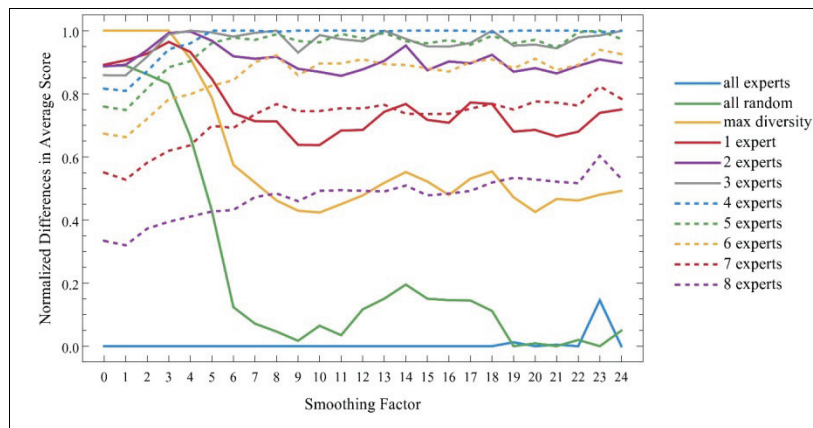
Here we find a number of interesting results. First, both max-diversity and all-random groups do best initially and DTA cleanly obtains in the relay dynamic for a smoothing factor of 0. Max-diversity performs best overall for smoothing factors 0-2 in relay and 0-3 in tournament play. This indicates as above, that in some parameter settings with low levels of within-landscape correlation, heuristic diversity holds an advantage over expertise. That being said, adding one expert to the group allows it to perform at least as well (relay) if not better (tournament) than all-random at every smoothing factor. Indeed, at more highly correlated landscapes all-random and max-diversity do quite poorly. By contrast, a group of all experts does even worse in tournament mode, but does quite well for higher smoothing factors in relay play. Within the standard interpretation of the model, our results show that diversity is important where there is a poor understanding of the problem (i.e. where landscape is highly random), but as the problem becomes better understood, the best groups are increasingly made up of experts. Whether, and under what circumstances we should take these results to apply to actual groups of problem solvers will be considered in the next section.

Figure 8: Normalized Averages for Pure and Mixed Groups of Size 9 Using a Relay Group Dynamics

Groups with maximum diversity (yellow line) perform well in areas of the parameter space with a high degree of randomness. In areas where the problems are better understood, the best groups are made mostly of experts (dotted purple line).

Figure 9: Normalized Averages for Pure and Mixed Groups of Size 9 Using a Tournament Group Dynamics

In sum, these results qualify and add nuance to our conclusions in Section 4. There we found that diversity trumped ability only for lower smoothing factors. As the landscape became less random, groups of experts regularly outperformed groups of randomly selected agents. Here, we find larger portions of the parameter space where the DTA result holds. Moreover, we find that not only

are there conditions under which diversity trumps ability, but that it makes sense to go further and claim that it is possible for diversity to trump expertise, since groups of randomly selected agents come to outperform groups of the best-performing agents in some portions of the parameter space which support an interpretation in terms of predictive expertise. However, the parameter sensitivity of the results should inspire caution in those whose interest lies in moving from these models to real world policy prescriptions. DTA's dependence on a number of modeling assumption emphasizes the need to gather further evidence before relying on such results in application, a point we expand upon in the conclusion.

6. Using Model Extensions to Probe Philosophical Arguments

The original Hong-Page model has received significant attention among practical philosophers. David Estlund (1997, 2009) and Jason Brennan (2011, 2016), among others, have asked why epistocracy is not preferable to democracy, given that experts are more apt to arrive at the truth than the *hoi polloi*. The DTA result has been cited as evidence for the epistemic superiority of democracy, providing formal evidence against these worries. While there are others who make use of DTA, we focus here on its deployment by Helene Landemore (2013a) in light of her close and careful consideration of the application of the original model. Doing so provides an excellent illustration of how ABMs and model extensions can shape and inform philosophical arguments.

The Hong-Page result gives us reason to believe that a diverse set of agents selected from the masses might well have the capacity to identify better solutions than a collection of rarified technocrats. Landemore makes use of the Hong-Page model to argue for the superiority of democracy in deliberative contexts by noting that its "results show that in groups of problem solvers it is often more important to maximize cognitive diversity […] than individual competence" (2013a, 121). Unlike voting, which is most often conceived of as a majoritarian process, deliberation is a project of consensus building (Landemore 2013a, 92). As arguments are made and alternatives discussed, some will be revealed to be better than others.

Consider the example of a jury. As jurors deliberate about the evidence and the merits of a case, they change their beliefs and are ideally either all persuaded of guilt or they come to a verdict of not guilty (or not proven). They all agree on the nature of the problem, and they seek to uncover the relevant evidence in order to arrive at the right conclusion – what we can think of in the context of the Hong-Page model as the global maximum.

Deliberation is one process that comes to mark democratic procedures by positing "an ideal speech situation where there are no time and information

constraints, to reach an uncoerced agreement on the 'better argument'" (Landemore 2013a, 92). The Hong-Page model fits well as a model of such an idealization, as it puts no information constraints between agents. Each agent is capable of locating, identifying, and communicating the value of their identified solution, free from time constraints.

While mindful of the details of the model – including the difficulty of the problem, basic agent competency, existence of local optima, and large population size – Landemore (2013a) believes that DTA gives us reason to think that democratic deliberation proffers a particular efficacy. We might ask why *democratic* deliberation is special, however, as opposed to say internal deliberation in which one sits alone thinking very hard about a particular problem.[7] Landemore's answer here – one motivated by the Hong-Page model – is that "larger deliberating groups are simply more likely to be cognitively diverse […] the more inclusive the deliberation process, the smarter the solutions resulting from it should be, overall" (2013a, 104). Diverse groups need not be better, but in the face of uncertainty (a characteristic of many of the problems we face) it is better to expand the sphere of politics to include more voices, rather than discriminate in favor of nominal experts.

She goes on, however, to consider representative domains, in which a sample must be drawn from among the population in order to deliberate on some matter – be it a jury or the Senate Foreign Relations Committee. And while she admits that it is appealing to consider selecting those with a particular set of cognitive resources ("oversampling cognitive minorities" as she calls it), such an impulse would be ultimately unwise. "This is due, essentially, to the unpredictable nature of political questions" (Landemore 2013a, 110).

Given that Landemore's (2013a) argument relies on her interpretation of the Hong-Page model, the model's properties bear on her argument for the superiority of diversity. Our modeling extensions can shed further light on her central conjectures, or at least articulate the appropriate scope of her claims. Consideration of the range of modeling results provides an example of how modeling can be used both to support and constrain philosophical argumentation.

Landemore claims that the Hong-Page "results show that in groups of problem solvers it is more important to maximize cognitive diversity […] than individual competence" (2013b, 1210). Our creation of the max-diversity group and our extension of the Hong-Page model allows us to explore this conjecture. If her analysis is right, a 'maximum diversity' group should do better than a random group (and *a fortiori* better than a group of the best-performing agents). In support of Landemore's claim that it is diversity *per se* that drives

---

[7]   In the original results, the best individual was regularly outperformed by both groups of experts and groups of randomly chosen agents (Hong and Page 2004).

the DTA, we do indeed find that the max-diversity group outperformed all other groups under the original Hong-Page parameter settings.[8]

Similarly, Landemore relies on the Hong-Page model to support the claim that because "cognitive diversity can trump individual ability, more inclusive groups are likely to be smarter" (2013b, 1217) and that this will hold even if it means "dumbing down" the average performance of the group. As before our extensions allow us to probe Landemore's claim by creating modeling analogues of her assertion. Restricting our attention to the original parameter settings (relay dynamic, max-heuristic = 12, group size = 9, smoothness = 0), we find support for Landemore's claim. The max-diversity group outperformed all other group including the group of randomly selected agents. Groups that had all or nearly all of its nine members selected randomly were roughly equivalent in terms of their performance. Moreover, Landemore's claim that group performance can go up even as the individual average goes down finds support. By construction, replacing a best-performing agent with a random agent will nearly always reduce the average ability of the individuals in the group. We find that beyond a ratio of 6 randomly selected agents to 3 best-performing agents, there was a steady decrease as each randomly selected agent was replaced by one of the best-performers. The worst group of all was composed entirely of best-performing agents.

Yet given that Landemore (2013b) is interested in applying these claims to areas where there are genuine subject experts, the original Hong-Page model offers no support. While the results on the original landscapes are consistent with Landemore's (2013b) claim, as we argued in Section 4, the original model does not support an interpretation in terms of ability, in the sense of predictive expertise. As illustrated in Figures 8 and 9, while "max-diversity" outperforms a randomly chosen group in all cases, it is outperformed by any group with even a single expert for the majority of the parameters settings considered here. Moreover, in the relay dynamics, groups tend to be better off the more experts they have, although it is worth noting that the best groups have one or two randomly chosen members amongst the experts.

In summary, our results show that some of the lessons philosophers have drawn from the work of Hong and Page can be better supported by creating variations on the original model to test the conjecture – whether maximizing cognitive diversity lead to better outcomes, for example. Our work also underscores the importance of assessing the robustness of a result by varying parameters in the model. While the original model lends support for Landemore's (2013b) claim that we would be better off trusting our decisions to randomly selected groups, our results suggest a more nuanced view: Random and diverse

---

8   To be clear, the construction of the max-diversity group confirms Landemore's understanding of what is driving the DTA modeling result, not whether this result applies to any real world situation. We discuss using models to inform policy discussions in the next section.

groups can perform quite well, and even optimally (e.g. when problem solving performance is a matter of luck), but they are not the best at identifying optimal solutions under all circumstances. When problems admit of true ability, for instance those tackled by bureaucratic agencies, expertise is plausibly important. Thus the Hong-Page model is not an unqualified argument in favor of democracy. Under some salient conditions a mixture of experts and non-experts proves to be epistemically optimal.

On the broader point, however, these extensions illustrate how ABMs can be used to craft, hone, and revise philosophical conjectures. Landemore does an outstanding job of deploying DTA with attention to the model's constraints. However, the original model is silent about the tradeoff between a mixed-composition group of experts and randomly selected agents. Beyond clarifying the effect our extensions have on the political philosophical debate here, the current work illustrates the broader point that by recruiting ABMs we can offer constructive amendments to central claims merely by tweaking the underlying model. Relying on ABMs allows for easier more lucid philosophical argumentation, in which the grounds of disagreement can be made more transparent.

6.   Conclusion

Our results demonstrate that a modeling perspective alone is insufficient to determine whether a group of experts or randomly selected agents should be expected to be epistemically superior. We first showed that the original diversity trumps ability result obtained on random landscapes in which "best-performing agents" were best thought of as lucky rather than good. When we modified the landscapes so that they might support an interpretation in terms of predictive expertise or ability we found that the best groups were composed of experts.

Changing still other parameters, such as group dynamic, dramatically increased the portion of the parameter space in which diversity trumps ability. 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. In short, the variety and sensitivity of these results argues against the straightforward application of the Page-Hong result in favor of either democracy or epistocracy. That being said, we do not wish to undersell the importance of the Hong-Page result – which we still take to be an important insight into the nature of group problem solving – or its relevance for political theorists.

Some clarity may be brought to the matter by considering recent classification of the many of uses to which models can be put. Grim et al. (2013) distinguish between four distinct intellectual contributions models can make, though

for our purposes we need just focus on two: emergent explanation and prediction.

When a model is used to generate an emergent explanation of some phenomenon, the modeler attempts to show that complex outcomes may be the result of simpler underlying rules. The output is taken as given and the modeler looks for input and mechanisms that could possibly produce such results. Echoing Page (2007, esp. chapter 13), we think there is evidence that suggests that diversity has epistemic benefits under some circumstances, though as Page (2007) himself cautions readers when he begins to discuss the empirical effects of diversity:

> we are not testing the logical truth of the models […] we're interested in whether the hypotheses and insights developed in these stark models shed any light on the complex multifaceted real world. Before we begin, two caveats are in order. First, the models that we covered do not provide the only candidate explanations for the empirical phenomena we discuss […]. Second, even if the evidence does align with the models, it need not always do so. (2007, 300)

Hong and Page have presented one possible way that such advantages might come about. Our results indicate similar advantages within a range beyond their original findings.

Political philosophers often have different purposes. In general, they do not tend to focus on how some particular event occurred, but wish to know what justifies a particular organization of political society or its institutions. One way to understand how political philosophers are using models is related to what Grim et al. (2013) classify as the 'predictive' capacity of models. While they are not interested in predicting some particular outcome, political philosophers do make arguments that compare the expected results of different ways of organizing political life (i.e. "In general, will we be better off leaving decisions to experts or groups that are more representative of the population?"). In such cases the input and the mechanism are taken as given and the outcome of the model is new information. Grim et al. suggest some further direction for those wishing to use a model in this way.

Specifically, one must ensure that both the input and the mechanism match the situation in which the model is being put to use. Moreover, because arguments pertain to general outcomes, either such work must be done across the range of possible situations or it must be shown that the set of relevant models robustly yields the same result across the parameter space. Additional empirical work is therefore required to show that both the inputs to the simulation and the mechanisms in the simulation sufficiently capture the real-world situation being modeled. A complete run down of the Hong-Page model is beyond the scope of the paper, but a couple of examples of correspondence should suffice to provide a feel for the additional warrant needed if we are to rely on models in this capacity.

As discussed above, the landscape of the problem turns out to have important implications for whether diverse groups prove superior. As mentioned in Section 3, our choice of 'smooth' landscapes represents one of many ways to construct landscapes. We chose smoothed landscapes to stay as close as possible to the original model while illustrating ways in which agents in the original landscape failed to qualify as predictive experts. It may well be that some problems are like this – smooth gradients separated by large troughs, where a solution can be easily found once a person moves across the conceptual trough. Yet, as others have suggested, many problems might be better modeled using alternative landscape topologies (Alexander, Himmelreich and Thompson 2015; Fontanari and Rodrigues 2016). Again, whether a landscape is sufficient as a justificatory model is a matter of empirically determining that the relevant aspects of the real-world target are indeed captured.

That said, empirical correspondence is not an intrinsic good. Models need not capture every aspect of the target phenomena, only the essential aspects. The non-correspondence of irrelevant aspects is a virtue of a good model. Which aspects need to be captured is determined by what users want to know and the true nature of the target system. As mentioned in Section 2.2, there are several potential factors which could be added to the model, but which are left out. This abstention does not necessarily undercut the utility of the model. The absence of such factors can also be seen as articulating the problem-solving dynamics in areas where such elaborations do not make an empirical difference. Again we trust that one example will suffice.

One aspect that we noted was that problem solving agents could not anticipate the result of multiple steps. One situation where this might plausibly apply is insight-learning.[9] In such cases, possible solutions come in roughly distinct wholes. If a person gets close enough to the solution, the whole solution quickly follows and appears all at once. Nevertheless, there are surely many situations relevant to political decision-making where solutions are iterative and agents consider multiple steps and reach multiple dead-ends before arriving at a solution. Here future modeling work might show that the results discussed above are robust to such elaborations and thus their non-inclusion in the model does no harm. Alternatively, it might show yet another factor which needs to be investigated empirically to assess correspondence.

---

[9] The classic example of insight learning is provided in the work of Wolfgang Köhler (1925). In his study of problem solving in chimpanzees, he placed bananas out of reach and left objects in the room, some that could be used for retrieval. For example, in one trial the bananas were placed too high to be reached, but the room contained crates that could be moved and stacked to make the food accessible. Chimpanzees did not blindly try random solutions until they achieved success (trial-and-error learning) but displayed a "eureka moment" and then engaged in goal-directed actions.

Until such work is done, modeling only shows increased diversity in our political representatives *can* yield more epistemically successful groups. This is certainly something for political philosophers to bear in mind, but as we show here, closely related models also show the possibility of the opposite effect. For just these reasons supporters of epistocracy cannot derive support from our results without further work to warrant such conclusions. That is, if the purpose of the model is to compare the expected results of different ways of organizing political life.

Prediction is not the only function of modeling. Models, even obviously false ones, can serve numerous productive functions (Wimsatt 1987). One great virtue of formal models is that assumptions are made explicit. Because of that virtue, models make clean targets both for empirical work and for building increasingly sophisticated and realistic models. We've attempted to offer some first steps in developing these models for group problem solving and the epistemology of democracy. Many more steps are surely needed.

References

Ahlstrom-Vij, Kristoffer. 2012. Why deliberative democracy is (still) untenable. *Public Affairs Quarterly* 26 (3): 199-220.

Alexander, Jason M., Johannes Himmelreich, and Christopher Thompson. 2015. Epistemic landscapes, optimal search, and the division of cognitive labor. *Philosophy of Science* 82 (3): 424-53.

Anderson, Elizabeth. 2006. The Epistemology of Democracy. *Episteme* 3 (1-2): 8-22.

Brennan, Jason. 2011. The Right to a Competent Electorate. *The Philosophical Quarterly* 61 (245): 700-24.

Brennan, Jason. 2016. *Against Democracy*. Princeton: Princeton University Press.

Estlund, David. 1997. Beyond fairness and deliberation: The epistemic dimension of democratic authority, vol. 32. Cambridge, MA: MIT Press.

Estlund, David. 2009. *Democratic authority: A philosophical framework.* Princeton: Princeton University Press.

Fontanari, José F., and Francisco A. Rodrigues. 2016. Influence of network topology on cooperative problem-solving systems. *Theory in Biosciences* 135 (3): 101-10.

Gaus, Gerald. 2016. *The Tyranny of the Ideal: Justice in a Diverse Society*. Princeton: Princeton University Press.

Goldman, Alvin, and Thomas Blanchard. 2016. Social Epistemology. In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. <https://plato.stanford.edu/entries/epistemology-social/> (Accessed January 27, 2018).

Grim, Patrick, Robert Rosenberger, Adam Rosenfeld, Brian Anderson, and Robb E. Eason. 2013. How simulations fail. *Synthese* 190 (12): 2367-90.

Gunn, Paul. 2014 Democracy and Epistocracy. *Critical Review* 26 (1-2): 59-79.

Hodgson, Robert T. 2008. An examination of judge reliability at a major US wine competition. *Journal of Wine Economics* 3 (02): 105-13.
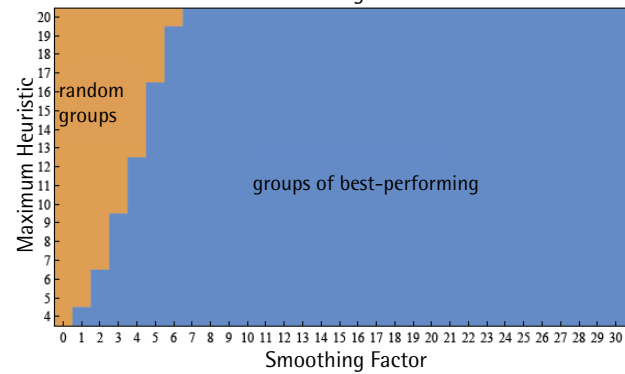
Hodgson, Robert T. 2009. An analysis of the concordance among 13 US wine competitions. *Journal of Wine Economics* 4 (01): 1-9.

Hong, Lu, and Scott E. Page. 2004. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences* 101: 16385-9.

Kohler, Wolfgang. 1925. *The mentality of apes.* London: Kegan Paul, Trench, Trubner and Co., Ltd.

Landemore, Helen. 2013a. *Democratic Reason: Politics, Collective Intelligence, and the Rules of the Many.* Princeton: Princeton University Press

Landemore, Helen. 2013b. Deliberation, cognitive diversity, and democratic inclusiveness: an epistemic argument for the random selection of representatives. *Synthese* 190: 1209-31.

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

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

Weymark, John A. 2015. Cognitive Diversity, Binary Decisions, and Epistemic Democracy. *Epistem*, 12 (04): 497-511.

Wimsatt William C. 1987. False Models as Means to Truer Theories. In *Neutral models in biology*, ed. Matthew H. Nitecki and Antoni Hoffman, 23-55. New York: Oxford University Press.
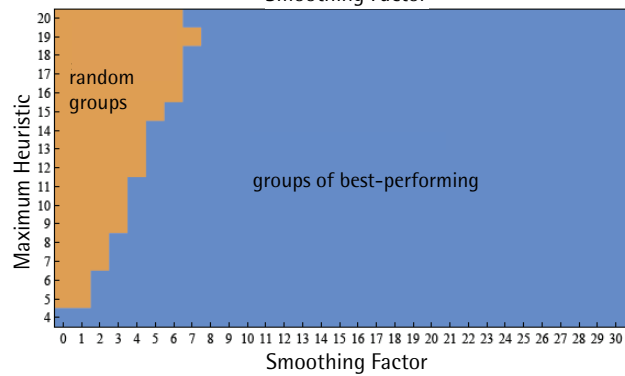
Appendix

Figure 2: Relay: Areas in Which Groups of Random Heuristics Do Best (�) and Areas in Which Groups of the Best-Performing Do Best (▇) across a Parameter Sweep of Landscape Smoothness and Max Heuristic for Groups of 3, 6, and 9.

Groups of 3.

Groups of 6.

Groups of 9.

Figure 3: Relay: Differences in Averages for Groups of Random Heuristics and Groups of the Best-Performing for Groups of 3, 6, and 9
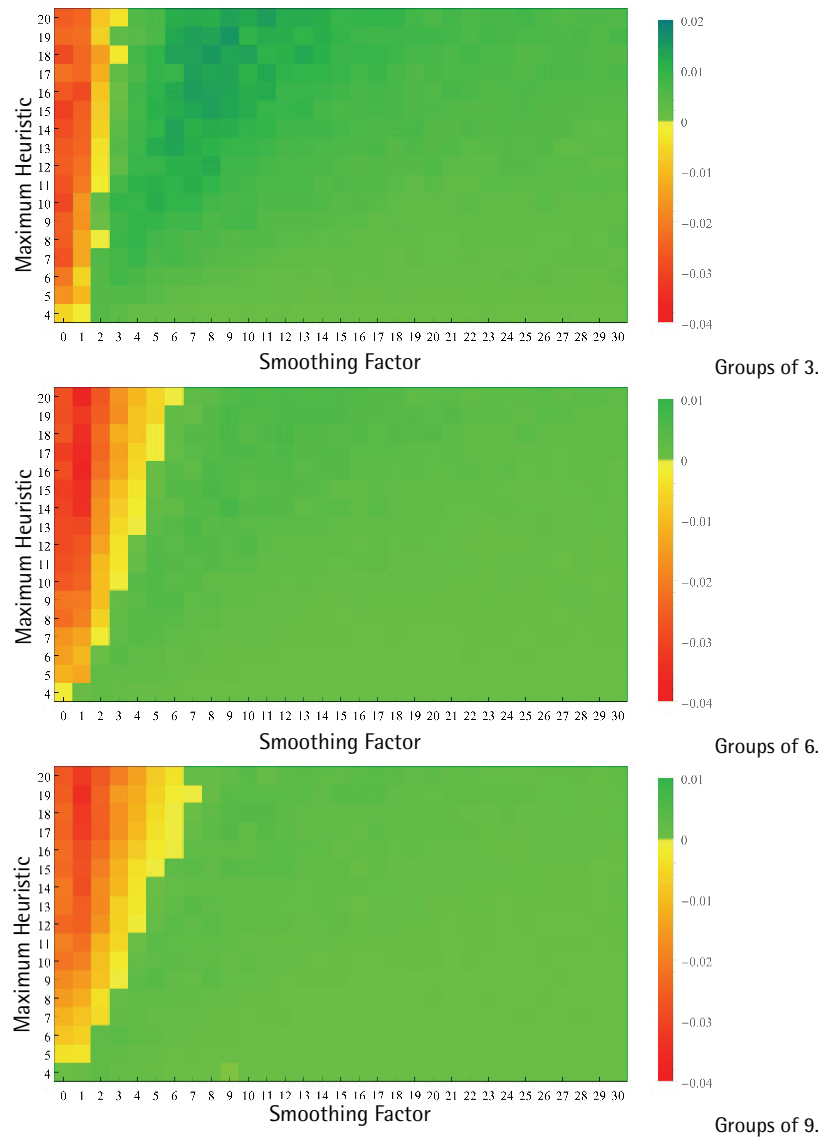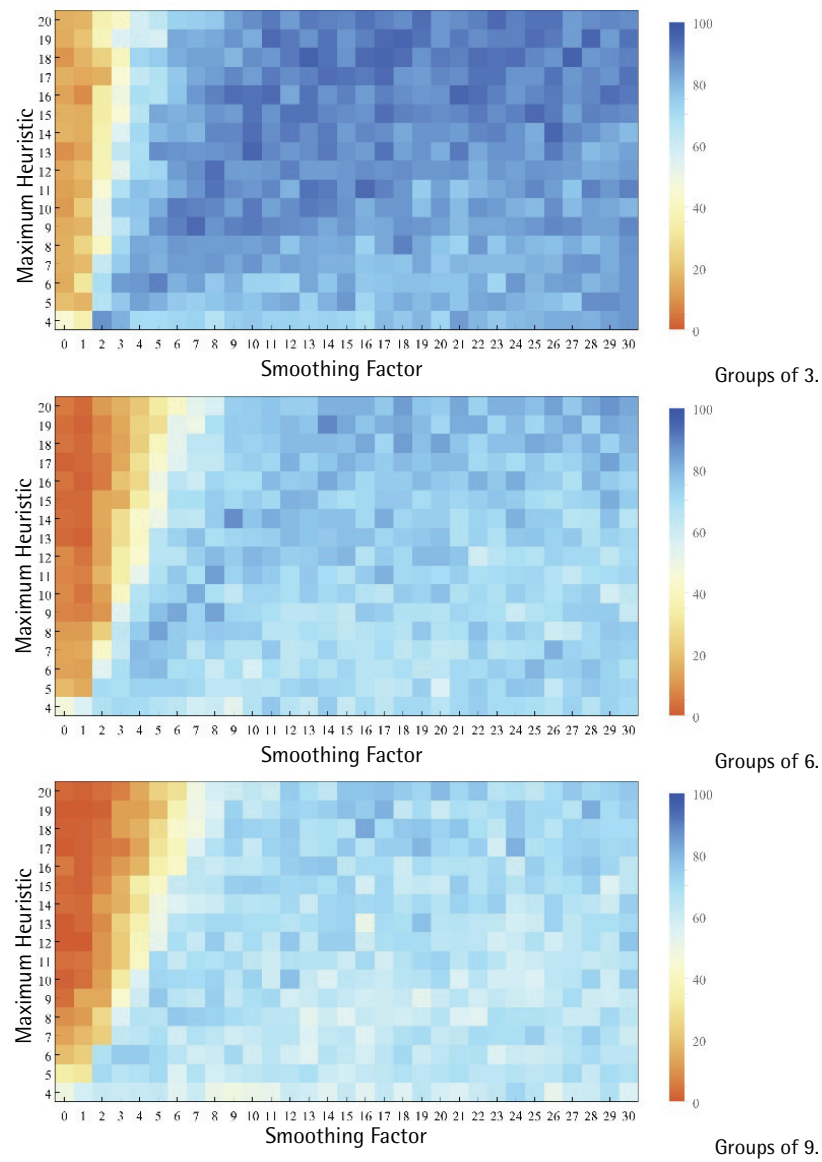

Groups of 3.

Groups of 6.

Groups of 9.

Figure 4: Relay: Percentages of Runs in Which Groups of the Best-Performing Do Better than Groups of Random Heuristics

Groups of 3.
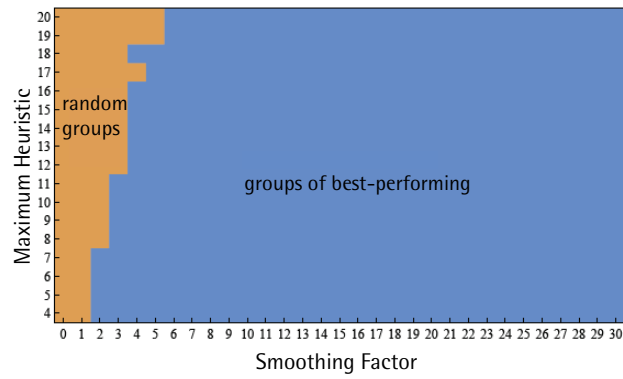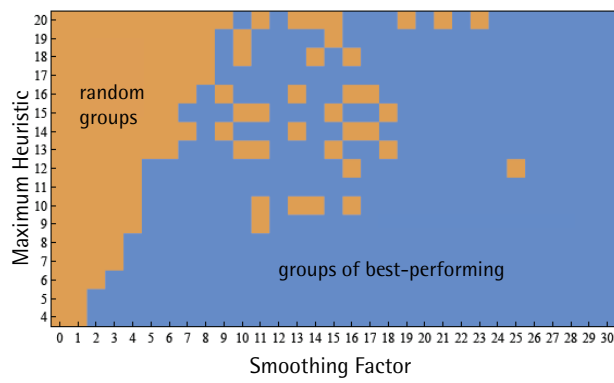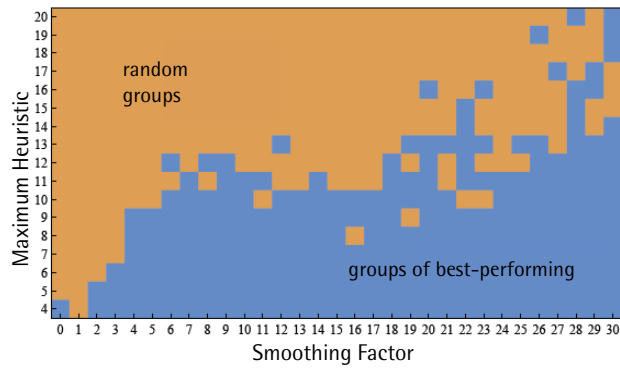
Groups of 6.

Groups of 9.

Figure 5: Tournament: Areas in Which Groups of Random Heuristics Do Best ( ) and Areas in Which Groups of the Best-Performing Do Best ( ) across a Parameter Sweep of Landscape Smoothness and Max Heuristic for Groups of 3, 6, and 9

Groups of 3.

Groups of 6.

Groups of 9.

Figure 6: Tournament: Differences in Averages for Groups of Random Heuristics
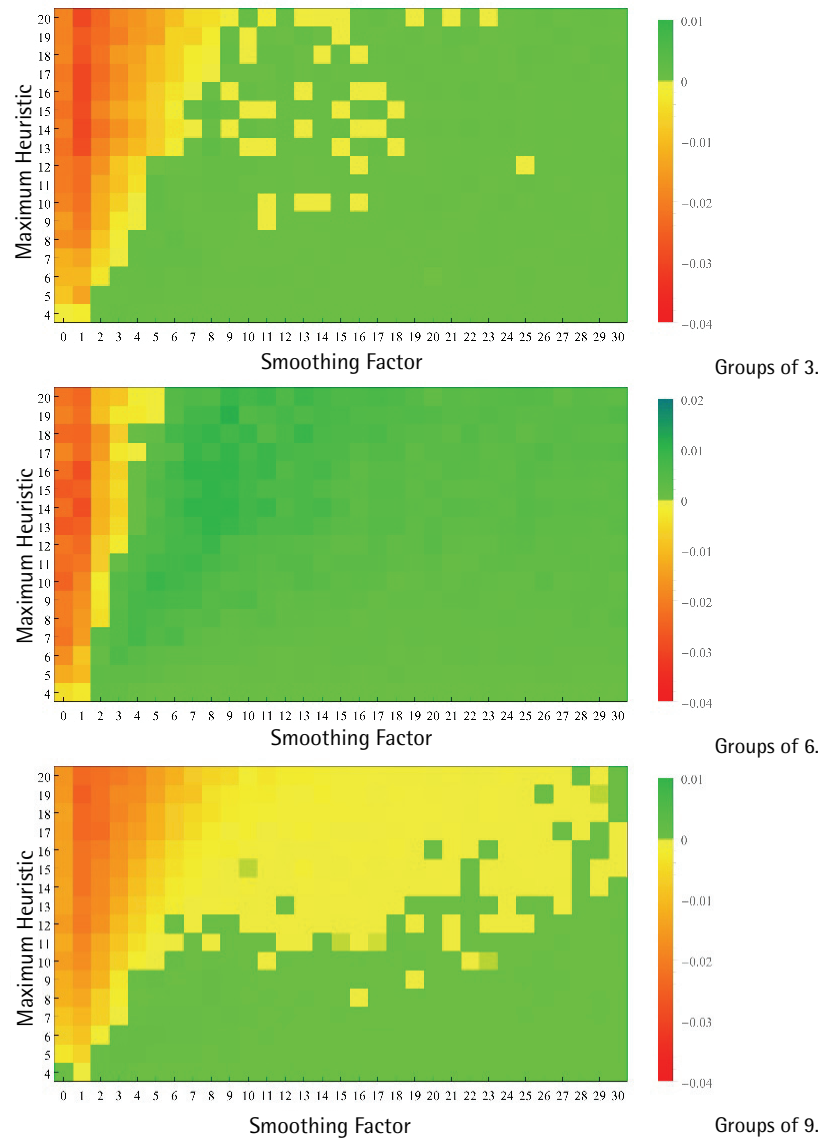and Groups of the Best-Performing for Groups of 3, 6, and 9

Groups of 3.

Groups of 6.

Groups of 9.

Figure 7: Tournament: Percentages of Runs in Which Groups of the Best–
Performing Do Better than Groups of Random Heuristics

Groups of 3.

Groups of 6.

Groups of 9.