

forthcoming in PSA2020/2021

---

# The diversity–ability trade-off in scientific problem solving

---

*Authors:*

Samuli REIJULA

Theoretical Philosophy, University of Helsinki

Jaakko KUORIKOSKI

Practical Philosophy, University of Helsinki

15 Jan 2021

# The diversity-ability trade-off in scientific problem solving

Reijula, Samuli<sup>1</sup> and Kuorikoski, Jaakko<sup>2</sup>

<sup>1</sup>Theoretical Philosophy, University of Helsinki

<sup>2</sup>Practical Philosophy, University of Helsinki

15 Jan 2021

## Abstract

*According to the diversity-beats-ability theorem, groups of diverse problem solvers can outperform groups of high-ability problem solvers. We argue that the model introduced by Lu Hong and Scott Page (2004; see also Grim et al. 2019) is inadequate for exploring the trade-off between diversity and ability. This is because the model employs an impoverished implementation of the problem-solving task. We present a new version of the model which captures the role of ‘ability’ in a meaningful way, and use it to explore the trade-offs between diversity and ability in scientific problem solving.*

**Keywords**— social epistemology of science; group problem solving; cognitive diversity; agent-based modeling; distributed cognition

## 1. Introduction

Modern science is a deeply collaborative enterprise. Most genuinely important intellectual challenges cannot be tackled by a single scientific discipline, let alone by individual researchers.

Science needs diversity – solving scientific research problems requires attaining specialized expertise and resources from a variety of perspectives.

Problem-solving groups in general are taken to benefit from diversity (Reagans and Zuckerman 2001; Mannix and Neale 2005; Jeppesen and Lakhani 2010; Steel et al. 2019). Among other important benefits, it is assumed that differences in how members of a group see a problem, in the cognitive resources they have at their disposal, and in the kind of heuristics they use, make it more likely that the the group as a whole has the resources to solve the problem. An important question, therefore, is whether the diversity of a group is in itself epistemically valuable, over and above the epistemic abilities of the group members.

Besides the empirical evidence cited above, a particularly influential argument in favor of diversity has been presented in the form of a mathematical theorem and an agent-based simulation. According to the *diversity-beats-ability (DAB) theorem*, groups of diverse problem solvers can outperform groups of high-ability problem solvers. This means that in assembling problem-solving teams, functional group diversity should sometimes be prioritized over selecting the most able individual members. Although they originate in computational social science, in management and organization studies, the DAB results have recently been also discussed in the philosophy of science (Grim et al. 2019; Singer 2019; Holman et al. 2018).

We argue that the "can" in the DAB theorem is ambiguous between several different modalities: in some of its uses, it is only a claim about conceptual possibility, whereas in its much advertised practical applications, it is clearly regarded as a more substantial possibility. This raises the question of when and under which exact conditions diversity really beats ability. We examine whether the original model by Hong and Page, and its further developments by Grim and associates, actually support the existence of the diversity-beats-ability phenomenon.

We show that due to their *impoverished task implementation*, these models cannot capture

interesting trade-offs between functional diversity and individual ability: the problem-solving tasks portrayed in the models are too difficult (i.e., random noise) for ability to make any difference to the outcomes. We develop a new version of the model with an improved problem-solving task. The new task representation allows our model to capture the role of individual ability in problem solving. Only when both diversity and ability really affect the outcome can the trade-off between them be studied.

We start by briefly presenting the DAB theorem and the associated simulation models, focusing on the latter. In Section 2, we highlight the "bait-and-switch" argumentative strategy used by Page to argue for DAB, showing that many of the modeling results supposed to support the theorem are problematic and do not replicate well. In Sections 3 and 4, we present our main argument: the model template used by Hong and Page as well as Grim and colleagues is ill-suited for exploring the trade-off between diversity and ability, because the problem-solving task is computationally implemented in a way that does not afford any advantage to individual ability or expertise. We introduce our version of the model, the stairway landscape, and demonstrate how it captures a substantial trade-off between diversity and ability. We draw two potentially interesting conclusions concerning the trade-off.

In this article, we are only concerned with the purely instrumental value of cognitive diversity; we are not arguing against the DAB phenomenon as such. We only ask whether the particular models we discuss are an informative and reliable way of exploring the possible trade-off, and provide what we regard as a better alternative way for doing so.

## **2. The diversity–ability trade-off in group problem solving**

Consider design tasks such as designing an automobile, a space shuttle, or a piece of software, or scientific tasks such as measuring the mass of an elementary particle or discovering the structure of a macromolecule. Heterogeneous cognitive and material resources need to be applied to solve all these problems, and as the set of solution candidates is not known beforehand, a search for solutions is needed. Simon (1989) suggested viewing the scientific research process through the lens of heuristic search. For instance, scientists search for formulations of problems, experimental designs, patterns in data, mechanisms behind data, and implications of their theories. On some occasions, these multi-dimensional search trajectories result in beneficial epistemic design; in other cases, they yield research approaches of little cognitive value. Importantly, most scientific problems worth solving lie beyond the capacities of a single knower, and scientific progress relies on a successful division of labor and collaboration between researchers, research groups, and sometimes even between scientific disciplines. Hence, scientific research should be understood as a socially distributed problem-solving process.

Such a picture of collective search immediately suggests a possible trade-off. On the one hand, as Newell and Simon (1972) suggested, expert performance often relies on highly specific search heuristics. On the other hand, more diversity in the group’s cognitive resources is beneficial, all other things being equal, as more varied resources provide access to larger portions of the solution space. Diversity may, however, conflict with individual ability. Experts are often more alike (in the relevant respects) than non-experts. Herein lies the trade-off: individual ability and group diversity both contribute to group performance, but, at least in some circumstances, the two factors may be in conflict.

Explicit modeling of the epistemic benefits of diversity in collective problem solving is

needed, because the phenomenon involves multiple group-level mechanisms as well as possible interactions between different epistemic, processual, and social factors. Therefore, purely verbal and conceptual theorizing is not a reliable tool for drawing out the implications of theoretical assumptions, and empirical (experimental or case-based) evidence does not usually unambiguously discriminate between alternative mechanistic explanations for why, in any particular case, diversity may or may not facilitate successful problem solving. Group problem solving has proved challenging to model, however. The computational implementation of the problem (task), cognitive resources (and differences therein), problem-solving behavior and cognition, and interaction between the group members all present difficult methodological and theoretical choices for the modeler, easily resulting in complex and intractable models with too many methodological degrees of freedom. Such models yield results which are hard to interpret. We believe that the heuristic-search paradigm proposed by Newell and Simon (1972) still provides the most promising approach for addressing these modeling challenges (see also Kauffman and Levin 1987; March 1991; Darden 1997). The models discussed and developed in this article join this tradition.

In a series of articles and books, Lu Hong and Scott Page have provided model-based evidence for the existence of the diversity-ability trade-off (Hong and Page 2001, 2004; Page 2008). They, in fact, use two distinct models to investigate diversity. The first model, introduced in Hong and Page (2001) and described in length by Page (2008) in the context of the diversity theorem, represents the problem to be solved as a binary string of finite length, where each bit could be seen as portraying a yes–no decision regarding a solution to a particular sub-problem (Kauffman and Levin 1987). A group of problem solvers of limited ability attempts to maximize a value function defined over the possible states of this string (potential solutions to the problem). Diversity is represented in the model by each agent having a different set of possible ways of

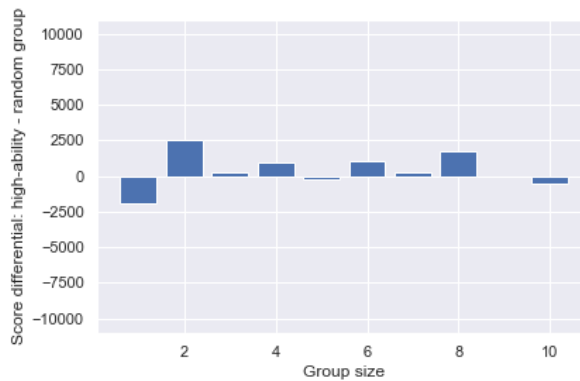


Figure 1: High-ability vs. random groups in the bit string model. The vertical axis represents the score differential between high-ability groups and random groups.

flipping the bits ("flipset heuristics") of the candidate solution string shared between the group members. Measures of problem difficulty can be assigned to alternative value functions (see Page 1996), and so the model can be used to represent a range of problems of different difficulty and complexity. This model template therefore corresponds well to pre-theoretic intuitions about how cognitive diversity can facilitate collective problem solving.

It is therefore rather surprising that the influential diversity-beats-ability results are not derived from this model. Our replication of the model in Hong and Page (2001) did not provide evidence to support the diversity-beats-ability phenomenon (see figure 1).<sup>1</sup> As the figure illustrates, no systematic difference emerges between groups of high-ability problem-solvers and groups of randomly selected problem-solvers. A more careful look at Page's 2008 argument reveals that it is based on evidence for the diversity theorem from an altogether different model introduced in Hong and Page (2004). We refer to this simplified model as the *ringworld model*. In sum, the substantial intuitions about diversity and ability in collective problem solving are first

---

1. For details about the bit string model, see Hong and Page 2001. All program code for the simulations and the generated data sets are available for download at [https://osf.io/a6f5e/?view\\_only=fcee3f72db8643b9999ad19447f89886](https://osf.io/a6f5e/?view_only=fcee3f72db8643b9999ad19447f89886)

formalized in one model, but the results are derived from a different model based on assumptions which do not correspond as neatly to the original intuitions. We find such a "bait-and-switch" argumentative strategy confusing, and not appropriate for transparent and epistemically sustainable use of theoretical models.

The argument in Hong and Page (2004) has a two-pronged structure: The basic assumptions of the ringworld model are used to derive an analytical proof intended to provide support for the theorem. However, as argued by Thompson (2014), the implications of the proof are unclear: even after technical corrections, the theorem only provides a highly abstract proof of possibility, and its implications for a non-technical interpretation of diversity are difficult to judge. Although we agree with Singer (2019) that the proof does rely on diversity and not merely on randomness (see Thompson 2014), it still remains the case that as such, the proof tells us little about the conditions under which the trade-off between diversity and ability can be expected to be significant. Mere logical possibility is not enough for the far-reaching practical implications suggested by Hong and Page. Their more persuasive evidence for DAB and its relevance for real-world group problem solving are derived from their agent-based simulation of the ringworld model. It is to this simulation that we now turn.

### **3. Problems in the Ringworld**

The “computational experiment” used by Hong and Page to demonstrate DAB portrays a group of agents collectively searching for optimal solutions in a one-dimensional landscape. The discrete landscape consists of positions  $1 \dots n$  on the number line, wrapped as a circle.<sup>2</sup> Value function  $V$  defined over the set of positions assigns to each position a payoff value drawn from the uniform

---

2. It turns out that the circular topology of the landscape does not make a difference to the results, as the distance explored by the individual agents (and groups) typically does not exceed 20 steps along the 2000-step circle.



distribution  $[0,100]$ . The agents' goal is to find the largest possible values on this landscape. To do so, each agent employs a heuristic  $\phi$ . A heuristic is defined as consisting of  $k$  different jumps of length  $1 \dots l$  (e.g.,  $[1,5,11]$  and  $[3,4,12]$  are two examples of heuristics with parameters  $k = 3, l = 12$ ). Starting from its current position, an agent sequentially applies these jumps along the landscape, and moves to a new position along the circle if the payoff associated with that position is strictly larger than the current one. When no further improvement is possible, the agent stops. The performance of an agent is defined as the expected payoff of the stopping points over the different starting positions of the landscape, and over a set of landscapes.

Hong and Page implement group problem solving behavior as sequential, iterative search. First, one agent initiates the search. As its local maximum is found, the second agent in the group takes the baton, and applies the jumps included in its heuristic as long as they lead to improvements. After all group members have taken their turn, a new round begins. The collective search stops when no agent can make further progress. Group performance is defined as the expected value of the position at which the group search stops.

In order to compare groups of high-ability problem solvers to more diverse ones, an exhaustive set of agents (with respect to possible heuristics) is first ranked according to their individual performance on a set of landscapes. A high-ability group of size  $g$  is constructed from the  $g$  highest performers in such a tournament, whereas the diverse group consists of  $g$  agents sampled randomly from the population.

In their model analysis, Hong and Page (2004) report results for various sets of parameter values. For example, for  $l = 12, k = 3, n = 2000$  they find that that the best individual agents scored 87.3 whereas the worst agent's score was 84.3. For groups of 10, the high-ability group scored 92.56 and the random group 94.53. This difference in favor of the random group is the diversity effect discovered in the simulation. Similar results were found by Grim and associates

(2019), and we were also able to replicate the findings.

Hong and Page suggest that there are reasons to believe that the random group scored higher due to its diversity. An alternative way to express this finding is in terms of *effective group size*. In our replication, we noticed that the difference in performance ('performance differential') between the random and the high-ability group was strongly correlated (.65) with the difference in effective group size between the two groups, where effective group size was defined as the size of the group heuristic from which overlapping elements had been removed. In other words, the similarity between the members of a high-ability group results in the group being functionally smaller (from the perspective of the problem-solving task). As the performance of a group generally increases as its effective group size gets larger, it is not surprising that smaller effective group size leads to worse performance.

Going back to the original DAB theorem, however, the explanation above seems to capture only one side of the diversity-ability trade-off. Although the correlation between effective group size and performance is an indication of the functioning of the "diversity mechanism," it is still unclear why that effect is stronger than the influence of the "ability mechanism," i.e., the fact that some heuristics should lead to higher performance than others, and that those high-performing heuristics should be more common in high-ability groups. A closer inspection of the model provides a solution to this puzzle.

Unlike Hong and Page, we regard the effect sizes from the simulation as remarkably small, given that they originate from theoretical modeling where the modeler is free to explore a broad range of hypothetical scenarios. One would expect a purely theoretical model, purpose-built to examine and demonstrate a specific mechanism using heavy idealizations, to reveal relatively unambiguous effects of the modeled mechanisms. As a matter of methodological principle, we believe that conclusions drawn from agent-based modeling would be strengthened by showing

how the effect size can be manipulated by changing model parameters. In other words, being able to "turn the dials" and observe how changes in model inputs result in systematic changes in the modeled effect suggests that we have reached understanding about the dependencies between model inputs and outputs (see Woodward 2003; Aydinonat, Reijula, and Ylikoski 2020).

Regarding the ringworld model, we argue there are two reasons to believe that the results reported by Hong and Page do not provide genuine insight into the diversity-ability trade-off.

First, with the parameter values studied by Hong and Page, in nearly half of the cases, the random group ends up with a *full heuristic*, that is, a heuristic consisting of all possible jumps  $[1, \dots, 12]$ . Furthermore, only 13% of the random groups have an effective group size smaller than 11. Hence, even if the agents in the high-ability group can make the jumps leading to high performance, it is highly likely that the same jumps will also be included in the heuristic of the random group – there is simply no way the high-ability group could systematically outperform the random one.

Secondly, as Grim and his colleagues (2019) also noted, the purely random landscapes studied by Hong and Page are simply not hospitable to anything that could be meaningfully interpreted as "ability" or "expertise." For heuristic search to be applicable, the task needs to have some structure or redundancy that the heuristic can exploit (Kahneman and Klein 2009; Kauffman and Levin 1987). Hence, aggregated over several random landscapes, no significant performance differences emerge between the different heuristics. This is seen in the very small performance differences between the best and worst performing individual agents (see above) in Hong and Page's simulations: the "ability mechanism" does not get any traction on the landscapes they studied. Therefore, we argue that the model does not appropriately capture the trade-off between diversity and ability.

Grim and his coauthors (2019) propose to remedy this problem by partially smoothing out the

random landscape (by adding interpolated values between randomly generated values). They argue that such a task representation can better capture ability, because on smoothed out landscapes individual performance is more transportable to other landscapes of similar smoothness. Yet a closer numerical examination of the results of this remedy again reveals only small differences between diversity and ability. Even on smoothed random landscapes, the expected performance difference between best performing and random individuals is minute. This suggests that these landscapes simply do not represent a problem that is suitably complex for exploring trade-offs between ability and diversity.

#### **4. Modeling the diversity–ability trade-off on stairway landscapes**

In order to better understand the tension between diversity and ability, we need to portray scenarios where also ability plays a role. In our own simulations, we introduce a type of problem where high ability – either at individual or group level – leads to noticeably increased performance. In science, having the right methodology for the problem at hand often sharply increases the epistemic payoff. Our stairway model differs from the Hong and Page ringworld model only in problem structure. The specifications of agent and group behavior remain the same as in the ringworld model. In generating problem landscapes, we start from the uniform noise distribution employed by Hong and Page. On top of those landscapes, however, we superimpose an increasing sequence of values, where the positions of the values are separated by intervals drawn from a finite set of integers in  $1 \dots l$  (see figure 2). We call this set the *step set*.

For an agent to climb the increasing subsequence, the *stairway sequence*, it must possess the

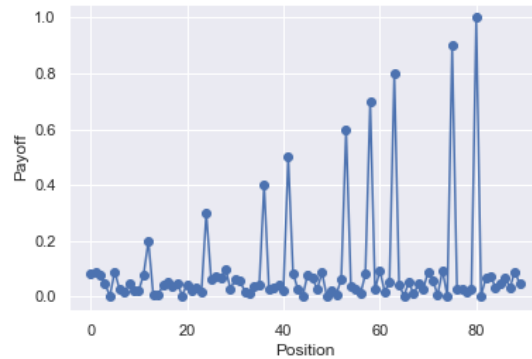


Figure 2: A stairway landscape with step set  $\{5, 12\}$ , and, therefore, step set size 2.

heuristic jumps corresponding to the steps used to generate the sequence (e.g.,  $[5, 12]$  in figure 2). This strongly favors some heuristics over others: whereas an agent who does not possess the full step set is bound to remain in the noise region of the landscape, a "high-ability" agent that has the necessary heuristic can climb through the whole sequence (and even reach the maximum payoff on the landscape, normalized to 1.0).

Figure 3 illustrates outcomes from our model with parameters values corresponding to those studied by Hong and Page (2004) and by Grim and his colleagues (2019). The left panel presents the difference between the performance of high-ability and random groups (positive values standing for high-ability group advantage, and negative values, for random group advantage). The results indicate that with these parameter values, stairway landscapes always favor high-ability groups. Especially when the group size is small, because it is made up of high-performing individuals (who typically possess valuable elements of the step set) the high-ability group performs significantly better than the random group. The right panel presents the difference between the redundancy of heuristics between the high-ability and random group (value 0 means that the overlap of heuristics in both groups is the same). As suggested by findings by Hong and

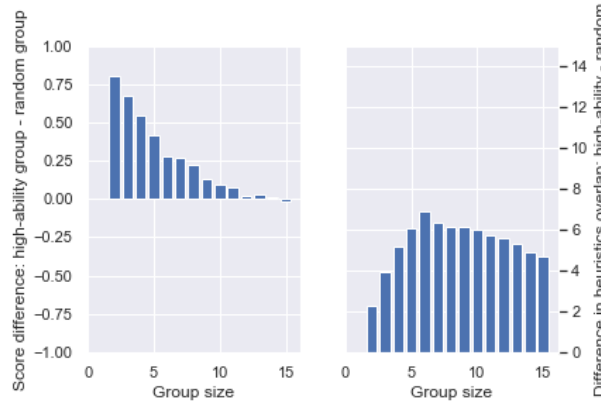


Figure 3: High-ability vs. random groups on a stairway landscape, step size 3. ( $k = 3, l = 12, n = 2000$ ; 100 repetitions over 100 landscapes)

Page (2004), random groups tend to have comparatively lower levels of overlap in their heuristics. As group size increases, the redundancy in the high-ability group increases more than in the random group. This suggests that when the group size is larger, random groups again begin to approach the full heuristic, which obviously is sufficient for climbing the stairway sequence. For this reason, at group sizes larger than 10, random groups catch up, and no significant performance difference is observed between high-ability and random groups (left panel).

We argue that this tension between the "ability mechanism" and the "diversity mechanism" captures the trade-off addressed by the DAB theorem. What happens, however, when the level of ability or expertise required by the task changes? Different levels of task difficulty can be represented by stairway landscapes with different step set sizes. For example, landscapes with step set sizes up to three lie within the abilities of the individual agents studied in the simulation ( $k = 3$ ). Climbing the stairway for step sizes larger than 3 requires pooling heuristics from several agents.

Figure 4 summarizes tentative findings from our studies with landscapes of varying difficulty. In the figure, group size is represented on the horizontal axis, and step set size (complexity of the

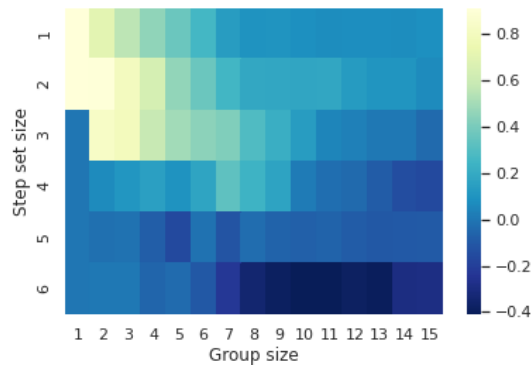


Figure 4: High-ability-vs-random group performance differential on stairway landscapes (50 repetitions, each over 50 landscapes).

problem) on the vertical axis. The color represents the performance differential between the high-ability group and the random group; lighter shades standing for high-ability group advantage. A genuine trade-off between diversity and ability can be seen. Observe the contrast between the upper-left quadrant, where ability dominates, and the lower-right, where random groups have a slight advantage over the high-ability groups; ability dominates when group size and step set size are small, whereas diversity leads to better performance when the group size and step set size are larger.

Finally, our results suggest a conceptual distinction between the complexity and difficulty of a problem: perhaps not surprisingly, ability dominates when the problem is simple in the specific sense that multiple cognitive resources do not need to be combined to solve it. Note that if the problem is simple in this sense, this does not necessarily mean that it is easy to solve. When the problem becomes complex, requiring efficient division of cognitive labor, the diversity effect begins to dominate over individual abilities. The results demonstrate how diversity and group size begin to outdo individual ability only when the problem complexity exceeds the cognitive resources of any single individual.

One could object to our stairway model on seemingly similar grounds to the ones on which we based our criticism of the original ringworld model. We questioned the DAB results on the basis that the model was built to favor diversity over ability. Why would our model fare any better, as it was clearly built to favor ability over diversity? This objection misses our point, however. Our argument is that the original model cannot be used to model the *trade-off* between diversity and ability, because it cannot be used to represent the gains from ability. Of course, we fully admit that the stairway landscape is built to favor ability, but the model nevertheless also retains the gains from diversity. Stairway landscapes give both ability and diversity their due, and, therefore, can illuminate the trade-off between them. This, we argue, was the original and interesting interpretation of the DAB results to begin with.

## 5. Conclusions

The original results by Hong and Page do not provide reliable evidence for the diversity-beats-ability theorem because the ringworld model, especially its task implementation, does not allow for ability to adequately influence individual or group performance. This one-sidedness implies that their model cannot be used to explore the possible trade-offs between diversity and ability in problem-solving groups. Our exploration of stairway landscapes illustrates how the results by Hong and Page (2004) rely on a problematic task structure to get their results.

Stairway landscapes provide a better model for "medium-hard" problems which require specialized abilities and true division of cognitive labor. Such landscapes can be used to model the interplay between diversity and ability relevant, and its effects on the division of cognitive labor in science.

Our tentative modeling results suggest a trade-off between diversity and ability. Ability is



avored when the problem is moderately difficult, requiring only a few different expert heuristics, and when groups are small. Diversity is favored when the problem is complex, requiring multiple component solutions, and when the groups are large. A further qualitative effect can be observed at the point where problem complexity increases beyond the capacity of a single agent and necessitates division of cognitive labor: simple problems solvable by individuals favor ability regardless of group size.

## **Acknowledgments**

We thank Kristina Rolin, Inkeri Koskinen, Renne Pesonen, and the other members of the TINT group (University of Helsinki), as well as the participants of Diversity in Science workshop (Tampere University, 5 May 2019) and the poster sessions at EPSA 2019 (University of Geneva) for their helpful comments. Thanks to Kate Sotejeff-Wilson for editing the manuscript. This research was carried out as a part of the project "Social and Cognitive Diversity in Science" funded by the Academy of Finland.

## References

- Aydinonat, N. E., S. Reijula, and P. K. Ylikoski. 2020. “Argumentative landscapes: the function of models in social epistemology.” *Synthese*, forthcoming.
- Darden, L. 1997. “Recent work in computational scientific discovery.” In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*, 161–166. Mahwah, New Jersey: Lawrence Erlbaum.
- Grim, P., D. J. Singer, A. Bramson, B. Holman, S. McGeehan, and W. J. Berger. 2019. “Diversity, ability, and expertise in epistemic communities.” *Philosophy of Science* 86 (1): 98–123.
- Holman, B., W. J. Berger, D. J. Singer, P. Grim, and A. Bramson. 2018. “Diversity and democracy: Agent-based modeling in political philosophy” [in en]. *Historical Social Research* 43 (1): 259–284.
- Hong, L., and S. E. Page. 2004. “Groups of diverse problem solvers can outperform groups of high-ability problem solvers” [in en]. *Proceedings of the National Academy of Sciences of the United States of America* 101 (46): 16385–16389.
- . 2001. “Problem solving by heterogeneous agents.” *Journal of economic theory* 97 (1): 123–163.
- Jeppesen, L. B., and K. R. Lakhani. 2010. “Marginality and problem-solving effectiveness in broadcast search.” *Organization science* 21 (5): 1016–1033.

- Kahneman, D., and G. Klein. 2009. "Conditions for intuitive expertise: A failure to disagree." *American Psychologist* 64 (6): 515–526.
- Kauffman, S., and S. Levin. 1987. "Towards a general theory of adaptive walks on rugged landscapes." *Journal of Theoretical Biology* 128 (1): 11–45.
- Mannix, E., and M. A. Neale. 2005. "What differences make a difference? The promise and reality of diverse teams in organizations." *Psychological science in the public interest* 6 (2): 31–55.
- March, J. G. 1991. "Exploration and exploitation in organizational learning." *Organization science* 2 (1): 71–87.
- Newell, A., and H. A. Simon. 1972. *Human problem solving*. Vol. 104. 9. Prentice-Hall Englewood Cliffs, NJ.
- Page, S. E. 2008. *The difference: how the power of diversity creates better groups, firms, schools, and societies*. Paperback. Princeton, N.J. ;Woodstock: Princeton University Press,
- . 1996. "Two measures of difficulty." *Economic Theory* 8 (2): 321–346.
- Reagans, R., and E. W. Zuckerman. 2001. "Networks, diversity, and productivity: The social capital of corporate R&D teams." *Organization science* 12 (4): 502–517.
- Simon, H. A. 1989. "The scientist as problem solver." *Complex information processing: The impact of Herbert A. Simon*, 375–398.
- Singer, D. J. 2019. "Diversity, not randomness, trumps ability." *Philosophy of Science* 86 (1): 178–191.

Steel, D., S. Fazelpour, B. Crewe, and K. Gillette. 2019. “Information elaboration and epistemic effects of diversity.” *Synthese*, forthcoming.

Thompson, A. 2014. “Does diversity trump ability?” *Notices of the AMS* 61 (9): 1–024.

Woodward, J. 2003. *Making things happen : A theory of causal explanation*. Oxford studies in philosophy of science. New York: Oxford University Press.