Comments on "Expertise versus Diversity across Epistemic Landscapes"

Conor Mayo-Wilson

Imagine you're a track coach choosing a relay team. You can either pick the fastest runners, or you can select a random group from your squad. Whom do you pick? The fastest runners of course! Here's a general lesson that many draw from simple examples like these: when you form a team to accomplish some task, choose "the best" individuals. Call this *Trump's principle.*

Hong and Page's results show that Trump's principle isn't a conceptual truth: there is a way of modeling group "problem-solving" in which a diverse set of individuals will, with high probability, outperform a homogeneous group consisting of the "best" problem-solvers. Hong and Page's model has several idiosyncratic features, but many homely examples suggest a general strategy for developing a model in which "diversity trumps [lowercase] ability" (DTA). I conjecture that, in 2003, a random selection of NBA players would have beaten a team consisting of five clones of Kobe Bryant. Although Kobe was the best player in the NBA at that time, members of the randomly-formed team would likely pass the ball to one another to improve their chances of scoring; the random team would contain at least one person who played defense, etc.

In general, people have complementary skills. If a group's members have diverse skills, then they can harness their complementary skills, and they might be able to achieve more than they could individually. In contrast, in homogeneous groups, individuals have the same skills, and so the combined efforts of several people, even very talented people, need not be more effective than one group member alone. There are many ways in which one might turn this crude argument into a precise mathematical result. Hong and Page provide one way, but there are obviously many others.

Because diversity is an important issue, it's unsurprising that Hong and Page's result has been oversold. Every scientific model has some range of applicability, and Hong and Page's work doesn't indicate that diversity is always valuable. If Bennett and his coauthors showed only that Hong and

Page's results are not true in in all models, then their work wouldn't be particularly interesting. But they show something considerably stronger. Bennett and his coauthors show that Hong and Pages's results are sensitive to seemingly small changes in the original model's assumptions. Specifically, they show that whether Trump's principle or DTA holds depends upon (1) the "smoothness" of the landscape and (2) the variety of problem-solving strategies available to the agents.

I think both conclusions are true and the argument supporting each is valid. So I would like to invite to Bennett and is coauthors to say more about (i) the types of problems to which their model is appropriate and (ii) how they might go about applying their work in those cases. I'll motivate my questions by analogy.

Suppose it's the late 18th century, and the geography and topography of the Western United States is still unknown to Europeans. You are in charge of forming a group to explore the West, and you ask a group of modelers, like Bennett, to help you decide which explorers to hire. Here's what the modelers tell you. If the Western terrain turns out to be very rugged and inhospitable, you'll need various specialists for your journey. For example, you'll need expert climbers, skilled hunters for tracking prey in foreign contexts, botanists for determining which plants are edible, and more. If the the landscape isn't particularly rugged, you'll want to bring the best mapmakers who can hike. The modelers also tell you that these recommendations are sensitive to the skill-sets of European explorers currently in America; if there are people (perhaps Natives) who have different ways of exploring that they haven't anticipated, then their recommendations might change.

You immediately recognize two problems with the modelers' suggestions. First, you don't know how rugged the landscape is. That is, in part, what you're trying to learn. Second, you don't know how diverse the set of exploring techniques are: there's no 18th century *Linked-In* in which aspiring explorers post detailed resumes discussing their approaches to navigating unknown lands.

In trying to apply their model to "epistemic" problems, Bennett and his coauthors seem to face two difficulties similar to those you would face in forming your group of 18th century explorers. First, the topography of the "epistemic" landscape is unknown, and second, it's not clear how diverse, in the relevant sense for Bennett and his co-authors, the set of scientific methodologies in use is.

Statisticians and empirical scientists routinely overcome problems analogous to the first difficulty. For example, suppose you're interested in the functional relationship between dosage of some new cancer drug and a patient's survival chances. In this case, which statistical estimators are "best" (e.g., in minimizing predictor error) will depend upon your model (here, the set of possible dosage response curves) and the best fitting member (here, the "best" curve). Empirical scientists, therefore, have domain-specific knowledge that suggest when particular types of statistical models are most plausible; dose-response curves are, for example, typically sigmoidal and not linear. Once a model is fixed, one can appeal to statistical criteria (e.g., consistency, unbiasedness, minimization of mean-squared error, etc.) for choosing among estimators, even though the performance of estimators depends upon the "true" member of the underlying model. Finally, statisticians also have developed methods for testing the appropriateness of the original model choice as more data is acquired.

Can Bennett and his coauthors do something similar? Can they give guidance about what types of real world problems are adequately represented by which types of landscapes? Once the set of relevant landscapes is fixed, can they provide criteria for choosing among different ensembles of problem-solving methodologies, given that the "true" landscape is unknown? Finally, can they provide guidance about how to test, as one watches various individuals' performances, whether or not they should expand the set of landscapes that one considers possible?

The second difficulty faced by Bennett and his coauthors is, I believe, importantly new, in that it has no analog in statistics or the sciences. In the Hong-and-Page-style models, the set of problem-solving techniques is just a set S of mathematical functions, of a particular type. So one can assess whether Trump's Principle or DTA is true by investigating what happens when functions/techniques are pulled at random from S. But there is little reason to expect that the methods that real people use, in any domain, populates the entire space of mathematical functions that one might use to model problem-solving techniques. Commonalities in education, professional training, and social background likely produce similar approaches to problem-solving in wide swaths of the Earth's population. Is there an application that Bennett and his coauthors have in mind in which there is a way of estimating the actual diversity of problem-solving techniques in use?

The final question I have for Bennett and his co-authors pulls in a different direction. Thus far, I've asked, in a nutshell, is there any way of measuring the factors that you've identified as relevant to group problem-solving ability, and if not, can you make informed recommendations about how to form groups in the absence of such information? The final question concerns whether the two factors that Bennett and his co-authors have identified are worth the effort I've discussed.

As we've said, Bennett and his co-authors show that whether DTA depends upon (1) the smoothness of the landscape and (2) the variety of available heuristics. But dozens of other factors might also be relevant. Do we have good reasons to think that, in certain empirical applications, Bennett and his coauthors have identified the *important* factors for assessing the value of diversity? Here's an analogy. Suppose you take the bus to work each day. You can then either walk slowly or briskly from the bus stop to your office. Clearly, your walking speed affects your commute time. But if your bus ride is an hour long and you could save 25 minutes by taking a train, then your walking speed is not the most important variable in deciding how you should get to work.

By analogy, there are lots of variables that affect group problem-solving ability. One that Bennett and his co-authors discuss only briefly (in the online supplement) is the size of the group. In reality, too many cooks can spoil the broth. But that's mathematically impossible in the models developed by Hong and Page and Bennett and his co-authors. In these models, a group can never get worse by adding a team member. Is there any way that we can have some confidence when, other variables are added to these models, the importance of the factors that Bennett and colleagues have identified is not swamped by new variables?

I want to conclude by noting the difference between questions/difficulties and criticisms. I don't see how to address that questions I've raised, and in part, that's because I'm not an economist, psychologist, or a sociologist who studies group dynamics in the real world. My lack of imagination and training does not constitute an objection to a paper. I hope that, using their combined skills, Bennett and his co-authors might be able to educate me or help me find people who could answer my questions.