Groups of diverse problem solvers can outperform groups of high-ability problem solvers

Lu Hong[†‡§] and Scott E. Page[¶]

[†]Michigan Business School and [¶]Complex Systems, University of Michigan, Ann Arbor, MI 48109-1234; and [‡]Department of Finance, Loyola University, Chicago, IL 60611

We introduce a general framework for modeling functionally diverse problem-solving agents. In this framework, problem-solving agents possess representations of problems and algorithms that they use to locate solutions. We use this framework to establish a result relevant to group composition. We find that when selecting a problem-solving team from a diverse population of intelligent agents, a team of randomly selected agents outperforms a team comprised of the best-performing agents. This result relies on the intuition that, as the initial pool of problem solvers becomes large, the best-performing agents necessarily become similar in the space of problem solvers. Their relatively greater ability is more than offset by their lack of problem-solving diversity.

A diverse society creates problems and opportunities. In the past, much of the public interest in diversity has focused on issues of fairness and representation. More recently, however, there has been a rising interest in the benefits of diversity. In the legal cases surrounding the University of Michigan's admissions policies and in efforts to curtail affirmative action in California, Texas, and elsewhere, there have been claims that diverse perspectives improve collective understanding and collective problem solving. Coincident with this political and legal wrangling has been an effort on the part of scholars to identify how to exploit this diversity both in solving hard computational problems (1, 2) and in human organizations (3).

In the common understanding, diversity in a group of people refers to differences in their demographic characteristics, cultural identities and ethnicity, and training and expertise. Advocates of diversity in problem-solving groups claim a linkage among these sorts of diversity (which we will refer to as identity diversity) and what we might call functional diversity, differences in how people represent problems and how they go about solving them. Given that linkage, they conclude that, because of their greater functional diversity, identity-diverse groups can outperform homogeneous groups (4–6).

Building on earlier ideas from the psychology and artificial intelligence literatures (7), we describe a mathematical framework for modeling problem solvers that captures the functional diversity that cognitive psychologists and organizational theorists claim is correlated with identity diversity. In our framework, agents possess internal representations of problems, which we call perspectives, and algorithms that they use to locate solutions, which we call heuristics. Together, a perspective-heuristic pair creates a mapping from the space of possible solutions to itself. A diverse group is one whose agents' mappings are diverse. Our perspective-heuristic framework is not minimal, because we show in an earlier paper (8) that two problem solvers with distinct perspectives and heuristics can act identically in the space of solutions. However, the advantage of the full framework is that it generalizes models in the computer science literature that focus on diverse heuristics (1, 2), and models in the organizational behavior and psychology literature, which often emphasize diverse perspectives (3, 4, 6).

The conclusion that identity-diverse groups can outperform homogeneous groups due to their greater functional diversity rests upon a well accepted claim that if agents across groups have equal ability, functionally diverse groups outperform homogeneous groups. It has also been shown that functionally diverse groups tend to outperform the best individual agents, provided that agents in the group are nearly as good (1). These results still leave open an important question: Can a functionally diverse group whose members have less ability outperform a group of people with high ability who may themselves be diverse? The main result of our paper addresses exactly this question.

Consider the following scenario: An organization wants to hire people to solve a hard problem. To make a more informed decision, the organization administers a test to 1,000 applicants that is designed to reflect their individual abilities in solving such a problem. Suppose the applicants receive scores ranging from 60% to 90%, so that they are all individually capable. Should the organization hire (*i*) the person with the highest score, (*ii*) 20 people with the next 20 highest scores, or (*iii*) 20 people randomly selected from the applicant pool? Ignoring possible problems of communication within a group, the existing literature would suggest that *ii* is better than *i*, because more people will search a larger space, but says little about *ii* vs. *iii*. The intuition that agents with the highest scores are smarter suggests that the organization should hire *ii*, the individually best-performing agents. The intuition that the randomly selected agents will be functionally diverse suggests that the organization should hire *iii*, the randomly selected ones. In this paper, we provide conditions under which *iii* is better than *ii*.

Thus, the focus of our analysis is on the tension between the individual abilities in a group and its functional diversity. Under the set of conditions we identify, as the initial pool of problem solvers becomes large, the functional diversity of the group of individually best-performing agents necessarily becomes very small. Ultimately, the gain in individual abilities is more than offset by the functional diversity of a group of randomly selected people. It is in this sense that we might say diversity trumps ability. This tension is established regardless of the precise nature of group cooperation. Complementary to our study, a computer science literature (2) has been addressing the questions of how to make the diverse group as effective as possible, and how and when the algorithms should share hints, information, and solutions, taking as a given that a diverse group does better than an individual. Organizational theorists have also focused on exploiting diversity. Their challenge has been how to encourage people with diverse identities and backgrounds to work together productively (3).

This paper focuses exclusively on functional diversity: differences in how people encode problems and attempt to solve them. The claim that perspectives and heuristics may be influenced by race, geography, gender, or age has much to recommend it, as does the claim that perspectives and tools are shaped by experiences, training, and preferences. However, even when applying our result to those cases when identity diversity has been shown

---

ECONOMIC SCIENCES

to correlate with functional diversity, we need to be acutely aware that identity-diverse groups often have more conflict, more problems with communication, and less mutual respect and trust among members (3, 9–11).

The next section presents the basic model of diverse problem-solving agents. *A Computational Experiment* reports simulation results establishing that a diverse group can often outperform a group of the best. *A Mathematical Theorem* explores the logic behind the simulation results and provides conditions under which diversity trumps ability. Some implications of our results are discussed in *Concluding Remarks.*

A Model of Diverse Problem Solvers

Our model consists of a population of problem solvers of limited ability who attempt to maximize a function V that maps a set of solutions X into real numbers. For example, the set of solutions could be the set of possible gasoline engine designs, with the value function generating the efficiency of various designs. Problem solvers have internal languages in which they encode solutions. This internal language can be interpreted at the neurological level, our brains perceive and store information, or metaphorically, we interpret problems based on our experience and training. The representation of solutions in the problem solver's internal language is called a perspective. Formally, a perspective is a mapping M from the set of solutions into the agent's internal language. A problem solver's heuristic is a mapping, denoted by A, from solutions in her internal language to subsets of solutions. It captures how she searches for solutions. Given a particular solution, the subset generated by the mapping A is the set of other solutions the agent considers.

In this way, the problem-solving ability of an agent is captured by her perspective and heuristic pair (M, A). Two agents can differ in either dimension or along both dimensions. Thus agents can have diverse perspectives (as psychologists assume), diverse heuristics (as computer scientists assume), or both. A solution is a local optimum for an individual agent if and only if when that agent encodes the problem and applies her heuristic, none of the other solutions she considers has a higher value. The set of local optima for an agent together with the size of their basins of attraction determines the agent's expected performance on the problem, or what we might call the agent's ability. It follows that a group of agents can get stuck only at a solution that lies in the intersection of the individual agents' local optima. This observation is independent of the procedure by which agents work together as a team. However, different procedures for interacting among the agents will generally result in different basins of attraction for those solutions that are local optima for all of the agents. Thus, how the team works together will matter for team performance (2).

A Computational Experiment

In a series of computational experiments we conducted based on this framework, we find that a collection of diverse agents can be highly effective collectively, locating good and often optimal solutions, confirming the widely accepted belief. More interestingly, we find that a random collection of agents drawn from a large set of limited-ability agents typically outperforms a collection of the very best agents from that same set. This result is because, with a large population of agents, the first group, although its members have more ability, is less diverse. To put it succinctly, diversity trumps ability.

Here, we report one such set of computational experiments where the second result is established and highlighted. We describe in detail how the general framework is applied, how individual and collective performances are measured, and how diversity is defined. We consider a random value function mapping the first n integers, $\{1, 2, \ldots, n\}$, into real numbers. The value of each of the n points is independently drawn

according to the uniform distribution on the interval [0, 100]. Agents try to find maximal values for this random function. In this set of experiments, we consider only agents who have identical perspectives but allow their heuristics to vary.[||] All agents encode n solutions as n points on a circle from 1 to n clockwise. The heuristic that an agent uses allows her to check k positions that lie within l points to the right of the status quo point on the circle. Here, $1 \le l < n$ and $1 \le k < l$. For example, consider $n = 200$, $k = 3$, and $l = 12$. A problem solver with the heuristic (1, 4, 11) starting at point 194 would first evaluate point 195 (194 + 1) and compare it with 194. If point 194 had a higher value, she would then evaluate point 198 (194 + 4). If point 198 had a higher value, she would then check point 9 (198 + 11−200). If that point had a higher value, she then would evaluate point 10 (9 + 1). She would keep evaluating until none of her three checks located a higher value. Therefore, a heuristic, denoted by $\phi = (\phi_1, \phi_2, \ldots, \phi_k)$, where each $\phi_i \in \{1, 2, \ldots, l\}$ specifies the position to check, naturally defines a stopping point for a search started at any initial point. Denote by $\phi(i)$ the stopping point of ϕ applied to initial point i. We measure the performance of an agent with a heuristic ϕ by the expected value of the stopping points, assuming that each point is equally likely to be the initial point,

$$E[V; \phi] = \frac{1}{n} \sum_{i=1}^{n} V[\phi(i)].$$

Given k and l, the set of heuristics is well defined. Because the order in which rules are applied matters, the total number of unique heuristics equals $l \times (l - 1) \times \cdots \times (l - k + 1)$. They can be ranked according to their expected values.

In our experiments, we considered environments in which a collection of agents attempt to find better solutions to the problem either sequentially or simultaneously. Our findings do not seem to depend on which structure was assumed. In the results we report here, agents approach the problem sequentially. The first agent searches until she attains a local optimum. The second agent begins her search at that point. After all agents have attempted to locate higher-valued solutions, the first agent searches again. The search stops only when no agent can locate an improvement, i.e., until the solution lies in the intersection of all agents' local optima. The collective performance of agents is then defined as the expected value of the stopping points, similar to the definition of performance of an individual agent.

The diversity of two heuristics ϕ^a and ϕ^b of the same size k, $\Delta(\phi^a, \phi^b)$, is defined by

$$\Delta(\phi^a, \phi^b) = \frac{k - \sum_{i=1}^{k} \delta(\phi_i^a, \phi_i^b)}{k},$$

where $\delta(\phi_i^a, \phi_i^b) = 1$ if $\phi_i^a = \phi_i^b$ and 0 else. For example, for $\phi^a = (5, 6, 9)$ and $\phi^b = (9, 5, 6)$, $\Delta(\phi^a, \phi^b) = 1$, because for any $i \in \{1, 2, 3\}$, $\phi_i^a \ne \phi_i^b$.

In the result we report, we set $l = 12$ or 20, $k = 3$, and the number of points on the circle $n = 2,000$. We experimented with l varying between 6 and 20, k varying between 2 and 7, and n varying between 200 and 10,000. Within these parameter ranges, we found qualitatively similar phenomena. For a given class of agents defined by k and l, we ranked all of the possible agents by their expected values and created two groups, one consisting of, say, the 10 best agents, the agents with the highest expected

Table 1. Result of computational experiments

Group composition	Performance	Diversity, %
Ten agents and $l = 12$		
Best agents	92.56 (0.020)	70.98 (0.798)
Random agents	94.53 (0.007)	90.99 (0.232)
Twenty agents and $l = 12$		
Best agents	93.78 (0.015)	74.95 (0.425)
Random agents	94.72 (0.005)	91.46 (0.066)
Ten agents and $l = 20$		
Best agents	93.52 (0.026)	73.69 (0.843)
Random agents	96.08 (0.006)	94.31 (0.089)

Numbers in parentheses are standard deviations.

values, and one consisting of 10 agents randomly chosen from the given class. For $l = 12$, the results from a representative single run were as follows: The best agent scored 87.3; the worst agent scored 84.3; the average score of the 10 best agents was 87.1, and the average score of the 10 randomly selected agents was 85.6. The collective performance of the 10 best agents had a value of 93.2; their average diversity (averaged over all possible pairs) was 0.72. The collective performance of the 10 randomly selected agents was 94.7; their average diversity was 0.92.** We present (Table 1) the results averaged over 50 trials. The data show that, on average, the collective performance of the randomly selected agents significantly outperforms the group of the best agents. Moreover, the diversity measures show a striking difference in the constituency of the two groups. The best group does not have nearly as much diversity as the random group. When we enlarged the group size from 10 to 20, the random group still did better, but with a less pronounced advantage. The group of the best agents became more diverse. This occurred because the set of heuristics was finite and fixed. Table 1 reports data with groups of 20, again averaged over 50 trials.

Next, we increase the set of possible heuristics (or agents) by setting $l = 20$. Agents can now look up to 20 spots ahead on the circle, and the total number of agents equals 6,840. Intuitively, we can make the following predictions. First, the diversity of the random group should be greater as a result of the increase in the number of heuristics. Second, this increased diversity should improve the collective performance of the random group. And third, the increase in the number of agents implies that the collective performance of the best group should also improve. The results of this set of experiments are presented in Table 1. From the data, we see, in fact, that all three predictions occur. Once again, diversity is the key to collective performance.

A Mathematical Theorem

In this section, we develop a mathematical theorem that explains the logic behind our new result: that a random collection of intelligent agents outperforms the collection consisting of only the best agents. Following is a brief summary of the theorem. In the mathematical model, agents want to maximize a value function that is assumed to have a unique maximum. Consider a population of agents, denoted by Φ, that satisfy the following assumptions: (*i*) Agents are intelligent: given any starting point, an agent finds a weakly better solution, and the set of local optima can be enumerated. (*ii*) The problem is difficult: no agent can always find the optimal solution. (*iii*) Agents are diverse: for any potential solution that is not the optimum, there exists at least one agent who can find an improvement. (*iv*) The best agent is unique. Consider drawing agents independently from Φ according to some distribution. The theorem states that with

probability one, there exist sample sizes N_1 and N, $N_1 < N$, such that the collective performance of N_1 drawn agents exceeds the collective performance of the N_1 individually best agents in the group of N drawn agents.

To formulate the theorem precisely, we begin with a set of solutions X and a given value function $V: X \rightarrow [0, 1]$, which has a unique maximum at x^*, and $V(x^*) = 1$. The problem solvers try to locate the solution x^*, but they have limited abilities. Each problem solver uses a search rule to search for the maximum but does not always end up there. Suppressing the distinction between perspectives and heuristics, we characterize each problem solver by a mapping $\phi: X \rightarrow X$ and a probability distribution ν on X. A problem solver randomly selects according to distribution ν an initial point where the search starts. For each x, $\phi(x)$ denotes the local optimum if the agent starts the search at x, that is, $\phi(x)$ is the stopping point of the search rule ϕ applied to x. In this interpretation, the search is deterministic: an initial point uniquely determines a stopping point. The image of the mapping, $\phi(X)$, is the set of local optima for problem solver ϕ. Mathematically, the mapping ϕ of a problem solver has to satisfy the following assumption:

Assumption 0.

  (*i*) $\forall x \in X, V(\phi(x)) \geq V(x)$
  (*ii*) $\phi(\phi(x)) = \phi(x)$

In general, the set of solutions X can be finite, denumerable, or a continuum. However, to avoid measuring theoretical complications, we present a simpler version of our result where X is assumed to be finite. This finite version makes the insight more straightforward, although it comes at the cost of trivializing some intricate assumptions and arguments. For example, the group of the best-performing agents is proven below to be comprised of identical agents. This is an artifact of the finite version. In the general version[††] under reasonable conditions, the group of the best-performing agents can be shown to be similar, not necessarily the same. But this makes the proof more complicated.

For each problem solver (ϕ, ν), we measure her performance by the expected value of her search, which we denote as $E(V; \phi, \nu)$, i.e.,

$$E(V; \phi, \nu) = \sum_{x \in X} V(\phi(x))\nu(x).$$

For the purpose of our analysis, we assume that all agents have the same ν, and that ν has full support. This assumption does not diminish the power of our result, because all the problem-solving ability of an agent is supposedly captured by the mapping ϕ. We now define the set of problem solvers Φ we consider. First, the problem is difficult for all agents under our consideration.

Assumption 1 (Difficulty). $\forall \phi \in \Phi$, *there exists* $x \in X$, *such that* $\phi(x) \neq x^*$.

This assumption simply necessitates the group setting. Given that ν has full support, it implies that no single agent under our consideration can always find the optimum. Second, we formulate the idea of a diverse group.

Assumption 2 (Diversity). $\forall x \in X \backslash \{x^*\}$, $\exists \phi \in \Phi$ *such that* $\phi(x) \neq x$.

This assumption is a simple way to capture the essence of diverse problem-solving approaches. When one agent gets stuck, there is always another agent that can find an improvement due to a different approach. Our last assumption states that there is a unique best performer in the set of agents we consider.

**Mathematically, the expected diversity of two randomly selected agents equals $11/12 =$
  0.9183333.

[††]Hong, L. & Page, S. E. (2002) Working paper, *Diversity and Optimality* [Loyola Univ. (Chicago) and Univ. of Michigan (Ann Arbor)].

Assumption 3 (Uniqueness). $argmax\{E(V; \phi, \nu): \phi \in \Phi\}$ *is unique.*

Let ν be the uniform distribution. If the value function V is one to one, then the uniqueness assumption is satisfied. Therefore, in the space of all value functions we consider, the uniqueness assumption is generically satisfied.

We do not make specific assumptions about how a group of problem solvers work together, other than requiring that search by a group can get stuck only at a point that is a local optimum for all agents in the group. An example of how this can be achieved is that agents approach the problem sequentially: wherever an agent gets stuck, the next person starts the search at that point.

Let Φ be a set of problem solvers that satisfy *Assumptions 1–3* above. Let μ be a probability distribution over Φ with full support. From Φ, we draw a group of N agents; each agent is drawn independently from Φ according to μ. These N agents are ordered by their individual performances, $E(V; \phi, \nu)$. Choose the best N_1 agents. We compare the joint performance of this group of N_1 agents with that of another group of N_1 agents that is formed by drawing each from Φ independently according to μ.

Theorem 1. *Let Φ be a set of problem solvers that satisfy* Assumptions 1–3 above. *Let μ be a probability distribution over Φ with full support. Then, with probability one, a sample path will have the following property: there exist positive integers N and N_1, $N > N_1$, such that the joint performance of the N_1 independently drawn problem solvers exceeds the joint performance of the N_1 individually best problem solvers among the group of N agents independently drawn from Φ according to μ.*

Here, there are in fact two independent random events: one is to independently draw a group of problem solvers, and the other is to independently draw a group of problem solvers and then select a subgroup according to their individual ability. The sample path we speak of in the theorem is the joint sample path of these two independent events.

The proof relies on two ideas. First we show (*Lemma 1* below) that the independently drawn collection of agents will, with probability one, find the optimal solution as the group becomes large. This lemma is intuitive, given that agents drawn independently thus are very unlikely to have common local optima. As the number of agents in the group grows, the probability of their having common local optima converges to zero. The second idea relies on the uniqueness assumption to show that, with probability one, as the number of agents becomes large, the best problem solvers all become similar and therefore do not do better than the single best problem solver, who by assumption cannot always find the optimal solution.

Consider the first random event of forming a group of problem solvers: each problem solver is independently drawn from Φ according to μ. Fix a sample path of this random event, ω_1. Let $\phi^1(\omega_1), \ldots, \phi^{n_1}(\omega_1)$ denote the first n_1 problem solvers. The joint performance of these n_1 problem solvers is the expected value of $V(\bar{y})$, where $\bar{y}$ is a common local optimum of all n_1 agents. The distribution of $\bar{y}$ is induced by the probability distribution of the initial draw, ν, and a precise model of how agents work together. The proof of the lemma that follows does not depend on what the specific model is. Without being explicit, we assume that $\bar{y}$ follows distribution $\eta_{\omega_1}^{n_1}: X \to [0, 1]$, i.e.,

$$\forall x \in X, \Pr(\bar{y} = x) = \eta_{\omega_1}^{n_1}(x).$$

Lemma 1. $\Pr\{\omega_1: \lim_{n_1 \to \infty} \Sigma_{x \in X} V(x)\eta_{\omega_1}^{n_1}(x) = 1\} = 1.$

Proof: Fix any $0 < \varepsilon < 1$. Define $A_{n_1} = \{\omega_1: 1 - \Sigma_{x \in X} V(x)\eta_{\omega_1}^{n_1}(x) > \varepsilon\}$. Obviously, $A_{n_1} \subseteq \{\omega_1: \phi^1(\omega_1), \ldots, \phi^{n_1}(\omega_1)$ have common local maxima other than $x^*\}$. Thus,

$$\Pr(A_{n_1}) \leq \Pr\{\omega_1: \phi^1(\omega_1), \ldots, \phi^{n_1}(\omega_1)$$

have common local maxima other than $x^*\}.$

Let $m = \min\{\mu(\phi): \phi \in \Phi\}$. Because μ has full support, $m > 0$. By *Assumption 2* (diversity), for any $x \in X\backslash\{x^*\}$, we have $\mu(\{\phi \in \Phi: \phi(x) = x\}) \leq 1 - m$. By independence,

$$\Pr\{\omega_1: \phi^1(\omega_1), \ldots, \phi^{n_1}(\omega^1)$$

have common local maxima other than $x^*\}$

$$\leq \Sigma_{x \in X\backslash\{x^*\}} \Pr\{\omega_1: x$$

is a common local maximum of $\phi^1(\omega_1), \ldots, \phi^{n_1}(\omega_1)\}$

$$\leq \Sigma_{x \in X\backslash\{x^*\}} (1 - m)^{n_1}$$

$$\leq (|X| - 1)(1 - m)^{n_1}.$$

Therefore,

$$\sum_{n_1=1}^{\infty} \Pr(A_{n_1}) \leq \frac{|X| - 1}{m} < \infty.$$

By the Borel–Cantelli Lemma, we have

$$\Pr\left\{ \omega_1: 1 - \sum_{x \in X} V(x)\eta_{\omega_1}^{n_1}(x) > \varepsilon \text{ infinitely often} \right\} = 0,$$

which implies

$$\Pr\left\{ \omega_1: \lim_{n_1 \to \infty} \sum_{x \in X} V(x)\eta_{\omega_1}^{n_1}(x) = 1 \right\} = 1.$$

We now prove *Theorem 1*.

Proof of Theorem 1: Consider the second random event, where a group of n agents are drawn independently from Φ according to μ, and then a subgroup of the best is formed. By *Assumption 3* (uniqueness), there is a unique problem solver in Φ with the highest individual performance. Call that agent ϕ^*. By the law of large numbers,

$$\Pr\left\{ \omega_2: \lim_{n \to \infty} \frac{\#\{i \in \{1, \ldots, n\}: \phi^i(\omega_2) = \phi^*\}}{n} = \mu(\phi^*) \right\}$$

$$= 1.$$

The fraction in the above expression is the frequency of ϕ^* in the draw. Let Ω be the set of sample paths $\omega = (\omega_1, \omega_2)$ that have both of the asymptotic properties above; i.e., define

$$\Omega = \left\{ \omega = (\omega_1, \omega_2): \right.$$

$$\lim_{n_1 \to \infty} \Sigma_{x \in X} V(x)\eta_{\omega_1}^{n_1}(x) = 1$$

$$\left. \text{and } \lim_{n \to \infty} \frac{\#\{i \in \{1, \ldots, n\}: \phi^i(\omega_2) = \phi^*\}}{n} = \mu(\phi^*) \right\}.$$

By *Lemma 1*,

$$\Pr(\Omega) = 1.$$

Fix any $\omega \in \Omega$. Let $\varepsilon_1 = 1 - E(V; \phi^*, \nu)$, which is positive by *Assumption 1* (difficulty). From the first limit above, we know there exists an integer $\bar{n}_1 > 0$, such that for any $n_1 \geq \bar{n}_1$,

$$\sum_{x \in X} V(x)\eta_{\omega_1}^{n_1}(x) > 1 - \varepsilon_1 = E(V; \phi^*, \nu).$$

From the second limit above, there exists an integer $\bar{n} > 0$, such that for any $n \geq \bar{n}$,

$$\frac{\#\{i \in \{1, \ldots, n\} : \phi^i(\omega_2) = \phi^*\}}{n} > \frac{\mu(\phi^*)}{2}.$$

Let $N_1 = \bar{n}_1$ and $N = \max\{2\bar{n}_1/\mu(\phi^*), \bar{n}\}$. Then

$$\sum_{x \in X} V(x)\eta_{\omega_1}^{N_1}(x) > E(V; \phi^*, \nu).$$

The left-hand side of the above inequality is the joint performance of the group of N_1 agents independently drawn according to μ. We now prove that the right-hand side term is the joint performance of the group of N_1 best agents from the group of N agents. By construction, $N \geq \bar{n}$. Therefore,

$$\frac{\#\{i \in \{1, \ldots, N\} : \phi^i(\omega_2) = \phi^*\}}{N} > \frac{\mu(\phi^*)}{2}.$$

That is,

$$\#\{i \in \{1, \ldots, N\} : \phi^i(\omega_2) = \phi^*\} > \frac{\mu(\phi^*)N}{2} \geq \bar{n}_1 = N_1,$$

because $N \geq 2\bar{n}_1/\mu(\phi^*)$. This means there are more than N_1 numbers of agents among the group of N agents that are the highest performing agent ϕ^*. Thus, the best N_1 agents among the N agents are all ϕ^*. Therefore, their joint performance is exactly the same as the performance of ϕ^*, which is $E(V; \phi^*, \nu)$. To summarize, for each $\omega \in \Omega$, there exist N_1 and N, $N > N_1$, such that the joint performance of the group of N_1 agents independently drawn according to μ is better than the joint performance of the N_1 best agents from the group of N agents independently drawn according to μ. Because the set Ω has probability 1, the theorem is proven.

Concluding Remarks

The main result of this paper provides conditions under which, in the limit, a random group of intelligent problem solvers will outperform a group of the best problem solvers. Our result provides insights into the trade-off between diversity and ability. An ideal group would contain high-ability problem solvers who are diverse. But, as we see in the proof of the result, as the pool of problem solvers grows larger, the very best problem solvers must become similar. In the limit, the highest-ability problem solvers cannot be diverse. The result also relies on the size of the random group becoming large. If not, the individual members of the random group may still have substantial overlap in their local optima and not perform well. At the same time, the group size

cannot be so large as to prevent the group of the best problem solvers from becoming similar. This effect can also be seen by comparing Table 1. As the group size becomes larger, the group of the best problem solvers becomes more diverse and, not surprisingly, the group performs relatively better.

A further implication of our result is that, in a problem-solving context, a person's value depends on her ability to improve the collective decision (8). A person's expected contribution is contextual, depending on the perspectives and heuristics of others who work on the problem. The diversity of an agent's problem-solving approach, as embedded in her perspective-heuristic pair, relative to the other problem solvers is an important predictor of her value and may be more relevant than her ability to solve the problem on her own. Thus, even if we were to accept the claim that IQ tests, Scholastic Aptitude Test scores, and college grades predict individual problem-solving ability, they may not be as important in determining a person's potential contribution as a problem solver as would be measures of how differently that person thinks.

Our result has implications for organizational forms and management styles, especially for problem-solving firms and organizations. In an environment where competition depends on continuous innovation and introduction of new products, firms with organizational forms that take advantage of the power of functional diversity should perform well. The research we cited earlier by computer scientists and organizational theorists who explore how to best exploit functional diversity becomes even more relevant. Most importantly, though, our result suggests that diversity in perspective and heuristic space should be encouraged. We should do more than just exploit our existing diversity. We may want to encourage even greater functional diversity, given its advantages.

The current model ignores several important features, including communication and learning. Our perspective-heuristic framework could be used to provide microfoundations for communication costs. Problem solvers with nearly identical perspectives but diverse heuristics should communicate with one another easily. But problem solvers with diverse perspectives may have trouble understanding solutions identified by other agents. Firms then may want to hire people with similar perspectives yet maintain a diversity of heuristics. In this way, the firm can exploit diversity while minimizing communication costs. Finally, our model also does not allow problem solvers to learn. Learning could be modeled as the acquisition of new perspectives and heuristics. Clearly, in a learning model, problem solvers would have incentives to acquire diverse perspectives and heuristics.

1. Huberman, B. (1990) *Physica D* **42**, 38–47.
2. Clearwater, S., Huberman, B. & Hogg, T. (1991) *Science* **254**, 1181–1183.
3. Polzer, J. T., Milton, L. P. & Swann, W. B., Jr. (2002) *Admin. Sci. Q.* **47**, 296–327.
4. Nisbett, R. & Ross, L. (1980) *Human Inference: Strategies and Shortcomings of Social Judgment* (Prentice–Hall, Englewood Cliffs, NJ).
5. Robbins, S. (1994) *Organizational Behavior* (Prentice–Hall, Saddle River, NJ).
6. Thomas, D. A. & Ely, R. J. (1996) *Harvard Bus. Rev.* September–October 1996, no. 96510.
7. Newell, A. & Simon, H. (1972) *Human Problem Solving* (Prentice–Hall, Englewood Cliffs, NJ).
8. Hong, L. & Page, S. E. (2001) *J. Econ. Theor.* **97,** 123–163.
9. MacLeod, B. (1996) *Can. J. Econ.* **29,** 788–810.
10. Ruderman, M., Hughes-James, M. & Jackson, S., eds. (1996) *Selected Research on Work Team Diversity* (Am. Psychol. Assoc., Washington, DC).
11. Watson, W., Kumar, K. & Michaelsen, L. (1993) *Acad. Manage. J.* **36,** 590–602.

ECONOMIC SCIENCES