Diversity and Optimality

Lu Hong* and Scott E. Page†‡

October 22, 1998

Abstract

In this paper, we construct a general model of diverse problem solvers of limited abilities. We use this model to derive two main results: (1) a collection of diverse, bounded problem solvers can locate optimal solutions to difficult problems and (2) a collection of problem solvers of diverse abilities tends to jointly outperform a collection of high ability problem solvers, where a problem solver's ability equals her expected individual performance.

1 Introduction

Humans have limited abilities in solving problems. We suffer from numerous biases, shortcomings, and constraints (Dawes 1988 and Nisbett and Ross 1980). Yet, we often locate good, or even optimal decisions to difficult, i.e. multi–dimensional and highly nonlinear, problems. Whether designing products, constructing welfare policies, re–organizing corporations, performing scientific research, or proving mathematical theorems, people collectively, either in groups or over time, perform amazingly well. This performance appears to contradict reason: How can people of bounded ability find good, and often optimal solutions to difficult problems?

In this paper, we demonstrate in a formal model that diversity can resolve the apparent contradiction between boundedly rational problem solvers and optimal decision making. We construct a model of diverse, boundedly rational problem solvers and analyze this model both computationally and mathematically, deriving two main results. First, with sufficient diversity, a collection of bounded problem solvers can locate the optimal solution to difficult problems. Second, we develop a general theorem providing sufficient conditions for a group

*Department of Economics, Syracuse University. lhong@mailbox.syr.edu

†Department of Economics, University of Iowa. scotte@page.biz.uiowa.edu

‡The authors would like to thank Ken Arrow, Jenna Bednar, Jonathon Bendor, Steven Durlauf, John Ledyard, Chuck Manski, Dierdre McClosky, Jeff Polzer, and Stan Reiter for comments on earlier versions of this paper.

of the best problem solvers as ranked by individual performance to be collectively outperformed by a group of randomly selected problem solvers. This rather surprising result has an intuitive explanation. If several thousand bounded problem solvers with diverse problem solving approaches are ranked by their individual abilities, the best problem solvers tend to take similar approaches. Therefore, the best ten, twenty, or forty problem solvers may not be effective collectively. In contrast, a collection of randomly selected problem solvers more likely contains a diversity of approaches, enabling it to locate better solutions.

By a collection of problem solvers, we do not necessarily mean a group of people sitting in a room together making a joint decision. The problem solvers might also operate within a hierarchy, where each person works on the problem and passes his solution to the person above him. We can even interpret the collective performance to be that which would occur in a market, where problem solving activities are not explicitly coordinated. Whoever discovers how to improve the steam engine earns economic rents. The ultimate product, whether it be an automobile, a microwave oven, a movie, or a piece of software, embodies the efforts of many individuals. Though it is likely that teams, firms, and markets differ in how they encourage people to locate solutions to problems, we emphasize here that, all else equal, firms, teams, and markets perform better when they consist of diverse problem solvers.

What do we mean by diversity? Do we then mean race, profession, gender, or ideology? We mean all of these and yet none of these. Zenisms aside, to us diversity means differences in problem solvers' perspectives and heuristics — variations in how people encode and search for solutions to problems.[1] These differences could result from disparate identities or ideologies, but they need not. And although we distinguish our approach from the idea that diversity refers to identity differences, i.e. racial, gender, or cultural diversity, we acknowledge and have sympathy for the idea that these more familiar notions of diversity may correlate with our formal definition. A Korean woman trained as a biologist probably frames and approaches a problem differently than an American man trained in materials science. Our approach extends beyond this, allowing diversity to arise from life experiences. Two white males from Oconomowoc, Wisconsin trained as engineers may think about how to design an engine very differently. In sum, diverse agents in our sense probably are more likely to be diverse in identity as well, though for we consider that to be an empirical question.

Interestingly, empirical research on racially, culturally, and gender diverse work forces suggests that they may be effective when these traditional identity notions of diversity correlate with diversity as we mean it here. In a recent article on diversity in the workplace, Thomas and Ely (1996) write that "Diversity should be understood as *the varied perspectives and approaches to work* that members of different identity groups bring."[2]

Formulating human capital as perspectives and heuristics extends the standard unidimensional conception of ability employed in many human capital models (Becker 1973),

[1]Economists often use the term diversity to describe heterogeneous preferences. She prefers chicken to fish. He prefers fish to chicken. Our diversity does not refer to preferences. Similar to McCloud (1996), we are advocating driving a wedge between behavior and preferences. We borrow the image of the "wedge" from McCloud.

[2]The *italics* are theirs.

and matching/marriage models (Mortensen 1988). These models focus on how a single variable, IQ for example, determines success (Hernsteinn and Murray 1994). We take a broader view. We think that economic agents create high value added by being diverse, by bringing new perspectives and heuristics to a task, not just by possessing raw intelligence. To give just one example, by focusing on economic incentives, Charles A. Beard, in *The Economic Basis of Politics* (1957), led to a new and better understanding of the motivation's of our country's founders. The brilliance of that work stems from its taking a different perspective, not from its undertaking a deeper, but similar analysis. This example suggests that the unidimensional ordering of abilities, though useful as a first cut, does not create a sufficiently rich instrument to measure potential contribution. Take, for instance, marriage and matching models. Ideal partners should possess complementary skills, not merely similar abilities. Two lawyers with identical perspectives and heuristics may not be of much use to one another as partners. Neither could ever offer a strategy that the other had not already considered.

The idea that diverse groups of people find better solutions than do homogeneous groups appears to have wide acceptance among people who study organizational behavior, with the caveat that heterogeneous groups perform less effectively at first due to communication problems (Watson, Kumar, and Michaelsen 1993). The following excerpt was taken from a well known organizational behavior textbook by Robbins (1994):

> *When a group is heterogeneous in terms of gender, personalities, opinions, abilities, skills, and perspectives, there is an increased probability that the group will possess the needed characteristics to complete its tasks effectively. The group may be more conflict laden and less expedient as diverse positions are introduced and assimilated, but the evidence generally supports the conclusion that heterogeneous groups perform more effectively than do those that are homogeneous. (p 261)*

The lack of expediency and potential for conflict lead many participants to think that groups make worse decisions than individuals. Compounding this impression, in many groups, participants value outcomes differently, often creating group decisions that appear suboptimal to group members. Hiring committees often seem to make silly decisions. The potential for and prevalence of poor decisions by groups may lead some to think that we have misstated a stylized fact, and that instead groups perform worse than individuals. Regardless of these preconceptions, the facts speak to the contrary: *groups on average outperform individuals on difficult problems.*

The main results of this paper add theoretical foundations for these empirical findings. Suprisingly, the results were derived analytically prior to us having any knowledge of the organizational theory literature on diverse groups. The formal model borrows earlier work on diverse agents (Hong and Page 1997). As mentioned, problem Solvers differ along two dimensions: their encodings of problems, *perspectives*, and the algorithms they apply in searching for solutions, *heuristics*. Confronted with a problem, a problem solver first encodes the space of possible solutions, and then applies the heuristics she has acquired during her lifetime to locate a local optimum. The use of heuristics is widespread in economics. Several papers define heuristics for game playing automata (Rubenstein 1986 and Kalai and Stanford

1988).[3] These automata may evolve (Miller 1996, Arifovic 1994, or Marimon, McGrattan and Sargent 1990) or be rationally selected. In our model, a heuristic defines a search procedure: climb along the gradient, flip a binary variable, etc. Finding a strategy in a game and attempting to locate a good solution to a problem are similar enterprises. Each requires search over a multi–variable domain for a good solution to a nonlinear problem. An agent of bounded abilities, or limited search time, searches by using a heuristic. These heuristics can become stuck on local optima.

In contrast, our inclusion of diverse perspectives is novel. Standard automata models assume exogenous and identical perspectives for all agents (Gilboa 1988). Yet in practice, people differ in how they encode problems. And game theorist vary in how they encode strategies. Some use neurel nets, some perceptrons, and some Moore machines. These differences in perspectives can simplify problem solving or make it more difficult. We all know that switching from Cartesian to polar coordinates can simplify integration in multiple dimensions. Polar and Cartesian coordinate systems are exactly what we mean by perspectives.

Central to our analysis is the idea that all people do not perform identically on difficult problems. On simple problems, they may. But once a problem becomes hard, diversity appears, creating a link between bounded rationality and diversity. Ironically, critics of bounded rationality models often cite the plethora of suboptimal, though sensible, rules which could be applied to a given situation as evidence against a particular rule. They dismiss bounded rationality models as either ad hoc or unstable. We agree that diversity abounds, but consider the diversity to be a blessing rather than a curse and explicitly model problem solvers with a variety of encodings and search rules.

Diversity is a subtle concept. Two problem solvers may have distinct perspectives and heuristics, yet perform identically on a problem (Hong and Page 1997). Therefore, perspective–heuristic pairs can be partitioned using an equivalence relation. Diversity, then, must mean variations in movements in the space of solutions. Diversity of perspectives and heuristics, though necessary, is not sufficient for differences in final solutions. This begs the question, why not then simplify the model and consider just movements in solution space? In fact, in our mathematical analysis, we do just that, abandoning the perspective–heuristic framework. However, the general model still plays an important role by providing underpinnings for diversity. Using the more general model, we can attribute diversity to distinct encodings of information and unique life experiences that lead to distinct accumulations of rules of thumb. Unable to rely on diverse perspectives and heuristics, we would be making ad hoc assumptions about diversity in solution space.

We have two additional motivations for including the full perspective–heuristic model. First, the mathematical theorems began as conjectures derived from computational experiments. These experiments relied on problem solvers with diverse perspectives and heuristics. We include two computational experiments in this paper as they provide a more transparent demonstration of the theory. They also substantiate the mathematical results. The theorems state that there exist an N and an N_1 such that the best N_1 of N problem solvers do not perform as well as a random N_1 problem solvers. The computational experiments use

[3] In addition to modeling heuristics with automata, perceptrons have been used (Cho 1993).

rather small N and N_1, in one case $N = 60$, suggesting that theorems with even weaker assumptions may be obtainable. In our opinion, the computational experiments should be interpreted as separate and corroborating support for the intuition that diversity improves outcomes.

Our second reason for including the full perspective–heuristic model is that it allows us to distinguish human problem solving diversity from computational problem solving diversity. Computers and people differ in their abilities to exploit diverse perspectives and diverse heuristics. Computers have a difficult time communicating across perspectives, something people can learn to do effectively. In sum, while abandoning the perspective–heuristic model may be mathematically convenient, doing so would sacrifice intuition and realism.

Any weighty interpretations of our analysis rest upon the existence and relevance of difficult problems. By a difficult problem, we mean a nonlinear, isolated problem with many variables (Page 1996).[4] The term isolated implies that a problem can be solved independently of other problems, that the ordinal ranking of a solution does not change when new solutions to other problems are located. For example, developing an efficient gasoline–combustion engine is a difficult problem. An engine's efficiency does not vary depending upon the pace of progress in other technologies. Similarly, solving a traveling salesperson problem is difficult, as are developing computer software, producing a movie, and setting airline prices in some instances.[5]

The remainder of the paper is organized into four sections. In the next section, we describe a general model of diverse problem solvers who rely on perspectives and heuristics. The following section contains two computational models that demonstrate the main formal results of the paper: that the collective performance of bounded problem solvers can be optimal and that a group of randomly selected problem solvers can outperform a group of the best performing problem solvers. The models also highlight the distinction between diverse perspectives and diverse heuristics. Section 4 contains a finite version of our mathematical results. The general version is in the appendix. The final section includes applications and a discussion of possible extensions of our model.

2 A General Model

We assume a finite number of problem solvers of limited ability who attempt to maximize a value function defined over a set of objects X. The set X can be finite or infinite. All problem solvers assign the same values to objects as determined by a value function $V : X \rightarrow \Re$. We assume that each problem solver has an internal language in which she perceives the objects. This internal language may be interpreted either at the neurological level—our brains perceive and store information, and these perceptions differ across individuals—or at the metaphorical level—we interpret problems based on our training as economists, lawyers, etc. We call the representation of objects in the problem solver's internal language her

[4]Information processing problems of the type studied by Radner (1993) and Radner and VanZandt (1995) are not, in our classification, difficult. They are just big.

[5]When the ordinal ranking of values can change with the actions of others the problem is complex (Page and Ryall 1998).

perspective.

Def'n: *A* **perspective** $M : R \rightarrow \Gamma$, *where* Γ *is the internal language, and* R *is a subset of* X.

If a perspective is both one to one and defined for all of X, i.e. $R = X$, and M is a one to one mapping, then we say that it is a *complete perspective.* In general there is no reason to assume that a problem solver's perspective is complete. A perspective may not be defined over all of X, i.e. R can be a proper subset of X, in which case an agent has no internal representation of some objects. Alternatively, a perspective may be many to one, in which case more than one object are mapped to the same representation in the internal language.

A problem solver's heuristic, denoted by A, is a mapping from elements of $M(R)$ in her internal language to subsets of $M(R)$. Given a $\gamma \in M(R)$, $A(\gamma) \subseteq M(R)$ is interpreted as the set of neighboring objects in the internal representation of the problem solver that she would check to find an improvement. Let $S = M(R)$. We restrict attention to a class of heuristics that consists of a collection of functions defined on S. For any $j = 1$ to m, let $f_j : S \rightarrow S$ be a function. We then define a heuristic $A = \{f_1, ..., f_m\}$, where $A(\gamma) = \{f_1(\gamma), ..., f_m(\gamma)\}$.

Def'n: *A* **heuristic** $A = \{f_1, ..., f_m\}$ *where* $f_j : S \rightarrow S$ *for* $j = 1$ *to* m.

For the moment, we refrain from presenting a precise description of how problem solvers apply heuristics to problems. We only assume that the heuristic determines those objects which a problem solver would evaluate given a status quo object. A problem solver tries each function f_j in her heuristic until she can no longer find an improvement. When she arrives at an object from which she can not find another improvement by applying her heuristic, this object is her solution to the problem and it is called her local optimum.

Def'n: *Given a problem solver* (M, A) *where* $A = \{f_1, ..., f_m\}$, *an object* x *is a* **local optimum with respect to** (M, A), $x \in L((M, A), V)$ *if and only if* $V(x) \geq V(M^{-1}(f_j(M(x))))$ *for all* $f_j \in A$

Clearly, the set of local optima of an agent depends on both her perspective and her heuristic.

We have yet to describe how the collection of problem solvers attacks a problem. They can approach the problem sequentially or simultaneously. For the mathematical results that we derive later, the precise rule does not matter so long as the final solution lies in the intersection of the local optima of all the problem solvers. However, in performing simulations, a fair comparison of groups of problem solvers requires that the two groups proceed similarly. In the two computational models that we describe in the next section, the problem solvers attack the problem sequentially. The first problem solver searches until she attains a local optimum. The second problem solver begins her search at that point. After all problem solvers have attempted to locate higher valued objects, the first problem solver searches again. Search stops only when no problem solver can locate an improvement, i.e. until the object lies in the intersection of the problem solvers' local optima.

3 Two Computational Models

We construct two computational models that support the main findings of our mathematical model: diverse, bounded problem solvers can collectively find good solutions to difficult problems and groups of random problem solvers outperform groups of the best problem solvers—those who working alone have the highest expected value from search. The first model considers problem solvers with diverse perspectives. The second model considers problem solvers with identical perspectives but diverse heuristics.

3.1 Model 1: Diverse Perspectives in R^n

We assume that there are k problem solvers. These problem solvers are confronted with a multidimensional problem whose domain is the unit ball in R^n.

$$f(x) = \sum_{i=1}^{n} \alpha_i \cdot x_i + \sum_{i=j+1}^{n} \sum_{j=1}^{n-1} \beta_{ij} \cdot x_i \cdot x_j \quad \text{where } \alpha_i, \beta_{ij} \in [-1, 1]$$

Each problem solver has a perspective: $(x^1, x^2, ..., x^n)$ where $x^i \in B^n$, the n dimensional Euclidean ball. A perspective is nothing more than a basis. All problem solvers use the same heuristic. Each problem solver sequentially searches along the vectors that define her perspective. Formally, a problem solver's heuristic consists of movements along each of these vectors a distance of 0.1 units. Perspectives generate all of the diversity.

In the simulation data shown below, we set $n = 30$. We performed similar tests for n varying between ten and one hundred and found similar results. Notice first that the individuals are not especially good at solving the problems and that even the best individual performs poorly relative to a simple hillclmbing algorithm or a genetic algorithm. Collectively, the group of all problem solvers outperforms the genetic algorithm and the hill climbing algorithm. The computational model demonstrates how a group of bounded, diverse problem solvers can perform effectively on a difficult problem.

Approach	**n = 30 Value**	**(s.d)**
Individuals	8.1	(0.14)
Best	10.8	(0.18)
Group	14.5	(0.22)
9000 Random	9.7	(0.28)
Hill Climbing	13.3	(0.24)
Genetic Algorithm	14.4	(0.21)

We next turn to how well collections of the best problem solvers, the worst problem solvers, and random problem solvers perform. Here, we obtain two surprising findings. First, on average, the worst twenty find nearly as good of a solution as the best twenty collectively. Second, twenty random problem solvers outperform the best twenty problem solvers. One explanation, the one that we propose in this paper, is that the twenty random problem solvers have more diverse perspectives than do the best twenty problem solvers.

Approach	**n = 30 Value**	**(s.d)**
Individuals	8.1	(0.14)
Best 20	9.5	(0.17)
Worst 20	6.7	(0.14)
Group	14.5	(0.22)
Group Best 20	14.2	(0.23)
Group Worst 20	13.8	(0.22)
Group Random 20	14.3	(0.22)

To test the claim that diversity explains the increase in performance, we could measure the diversity of the perspectives, but the calculations of diversity become messy, so we construct a simple model of diverse heuristics that allows for a cleaner analysis.

3.2 Model 2: Diverse Heuristics on a Circle

In this model, a finite set of n *objects* are encoded as n points on a circle. All problem solvers use the same perspective, i.e., they have the same encoding of the objects. We assume a random value function mapping $\{1, 2, ..., n\}$ into the real numbers. The value of each of the n points is independently uniformly drawn from the interval $[0, 100]$. We number the points consecutively from 1 to n on a circle clockwise, so that point n is to the left of and next to point 1. The heuristic that a problem solver uses allows her to check k $(1 \leq k < n)$ positions that lie within l $(1 \leq l < n)$ points to the right of the status quo point on the circle.

Example: *Let* $n = 200$, $k = 3$ and $l = 12$. *A problem solver with heuristic* $(1, 4, 11)$ *starting at the point 194 would first evaluate point 195 (194+1) and compare it to 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 keeps evaluating until none of her three searches locates a point with a higher value.*

Def'n: *The stopping point of a* **heuristic** $\phi = (\phi_1, \phi_2, ..., \phi_k)$ *applied to object* m*, where* $\phi_i \in \{1, 2, ..., l\}$ *and* $\phi_i \neq \phi_j$ *for* $i \neq j$*, denoted by* $\phi(m)$*, is defined as follows*:

Step 1: $a = 0$, $t = 0$, $\hat{s} = m$

Step 2: $t = t + 1$, $s^t = \hat{s}$

Step 3: If $t > k$ *and* $s^{t-k} = s^t$ *then go to Step 7 else go to Step 4*

Step 4: Let $a = a + 1$. *If* $a > k$, *then let* $a = 1$

Step 5 : Let $s' = \phi_a + s^t$. *If* $s' > n$, *then let* $s' = \phi_a + s^t - n$

Step 6: If $V(s') > V(s^t)$, *then* $\hat{s} = s'$, *otherwise* $\hat{s} = s^t$. *Goto Step 2*

Step 7: $\phi(m) = \hat{s}$. *End*

A heuristic, then, maps a point into a point with a weakly higher value. To compute the expected value for a problem solver, we start the problem solver at each of the n points on the circle and compute the average value of the points where search stops.

Def'n: *The* **expected value** *of a heuristic* ϕ *given* V,

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

In these computational experiments, we evaluate all heuristics within a well defined class instead of randomly generating a set of perspectives as we did in the first model. We restrict the set of heuristics to k movements to the right, where each movement has a maximum length l. The order that a problem solver applies these movements may matter. The heuristics $(5, 6, 9)$ and $(9, 5, 6)$ typically have different values. Therefore, we consider these to be distinct heuristics.[6] The total number of unique heuristics equals $l \cdot (l-1) \cdot \cdot (l-k+1)$. For example, if $l = 12$ and $k = 3$, then the total number of heuristics equals 1320. The 1320 heuristics can be ranked by their expected values.

The diversity of two heuristics ϕ^a and ϕ^b can be measured either with respect to order or not. In the first case, we calculate the percentage of ϕ_i^a that equal ϕ_i^b. In the latter case, we calculate the percentage of ϕ_i^a that equal ϕ_j^b for some j. The definitions are formalized below:

Def'n: *The* **ordered diversity** *of* ϕ^a *and* ϕ^b,

$$O\Delta(\phi^a, \phi^b) = \frac{k - \sum_{i=1}^{k} \delta(\phi_i^a, \phi_i^b)}{k} \qquad \text{where } \delta(\phi_i^a, \phi_i^b) = 1 \text{ if } \phi_i^a = \phi_i^b \text{ and } 0 \text{ else}$$

[6] In simulations where we only create one heuristic instead of six for each triple of numbers, we find identical qualitative results.

Def'n: *The* **diversity** *of* ϕ^a *and* ϕ^b,

$$\Delta(\phi^a, \phi^b) = \frac{k - \sum_{j=1}^{k} \sum_{i=1}^{k} \delta(\phi_i^a, \phi_j^b)}{k} \qquad \text{where } \delta(\phi_i^a, \phi_j^b) = 1 \text{ if } \phi_i^a = \phi_j^b \text{ and } 0 \text{ else}$$

For example, let $\phi^a = (5, 6, 9)$ and $\phi^b = (9, 5, 6)$. Then $O\Delta(\phi^a, \phi^b) = 1$ since for any $i \in \{1, 2, 3\}$, $\phi_i^a \neq \phi_i^b$. However, $\Delta(\phi^a, \phi^b) = 0$ since for any $i \in \{1, 2, 3\}$, $\phi_i^a = \phi_j^b$ for some $j \in \{1, 2, 3\}$. It is easy to see that for any two heuristics, the ordered diversity weakly exceeds the diversity.

In the computational data we report, we set l equal to either twelve or twenty and set k equal to three. The number of points on the circle n equals two thousand. We experimented with l varying between six and twenty, k varying between two and seven, and n varying between two-hundred and ten thousand. Within these parameter ranges, we found qualitatively similar phenomena.[7]

The values of each of the two thousand points on the circle were randomly distributed uniformly in the interval $[0, 100]$. We ranked all of the possible problem solvers by their expected values and created two groups, one consisting of the, say ten best problem solvers—the problem solvers with the highest expected values—and one consisting of ten randomly chosen problem solvers. The results from a representative single run looked as follows: The best problem solver had a score of 87.3. The worst problem solver had a score of 84.3. The average score of the ten best problem solvers was 87.1, and the average score of the ten randomly selected problem solvers was 85.6. The group performance of the ten best problem solvers had a value of 93.2, their average diversity was 0.45, and their average ordered diversity was 0.72. The randomly selected group's performance was 94.7, their average diversity was 0.76 and their ordered diversity was 0.92. As in the previous computational model, the group of random problem solvers collectively performed better. More importantly, we can now verify that the random group contains more diverse members.[8]

Below we present data averaged over fifty trials.

Ten Problem Solvers $l = 12$

Group Composition	**Group Performance**	**Ordered Diversity**	**Diversity**
Best Problem Solvers	92.56 (0.020)	70.98 (0.798)	38.77 (1.59)
Random Problem Solvers	94.53 (0.007)	90.99 (0.232)	75.13 (0.204)

[7] As the group size becomes large relative to the number of possible problem solvers, the group of the best agents can outperform a group of randomly selected agents.

[8] Mathematically, the expected ordered diversity of two randomly selected problem solvers equals $\frac{11}{12} = 0.9183333$. A more elaborate calculation shows that the expected diversity of two randomly selected problem solvers equals $\frac{3}{4} = 0.75$.

On average, the group of the random problem solvers significantly outperforms the group of the best problem solvers. In a typical run, the best problem solver in the population had an average value of between 86 and 88 in each set of fifty trials and a random problem solver had a value around two to three points lower. 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. Two questions immediately come to mind. First, what if we enlarge the group size, and second, what if we increase the the set of possible heuristics.

The answer to the first question is that the same phenomenon occurs. The random group still does better, but with a less pronounced advantage. The group of the best problem solvers becomes more diverse. This occurs because the set of heuristics is finite and fixed. The Table below provides data from fifty simulations with groups of size twenty.

Twenty Problem Solvers ($l = 12$)

Group Composition	**Performance**	**Ordered Diversity**	**Diversity**
Best Problem Solvers	93.78 (0.015)	74.95 (0.425)	44.47 (0.612)
Random Problem Solvers	94.72 (0.005)	91.46 (0.066)	74.88 (0.080)

To answer the second question, we present data from computations where problem solvers can look up to twenty spots ahead on the circle. Now, the total number of problem solvers equals 6840. Three predictions come to mind. 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 random group's performance. And third, the increase in the number of problem solvers implies that the group of the best problem solvers should also find a better solution. We see, in fact, that all three occur. The best problem solvers do better. The random problem solvers do better. And the random problem solvers are more diverse. The Table below provides data from fifty simulations.

Ten Problem Solvers ($l = 20$)

Group Composition	**Performance**	**Ordered Diversity**	**Diversity**
Best Problem Solvers	93.52 (0.026)	73.69 (0.843)	44.53 (1.782)
Random Problem Solvers	96.08 (0.006)	94.31 (0.089)	85.17 (0.165)

4 The Mathematical Model

The two previous computational models demonstrate the benefit of diversity in group and collective problem solving. In both models, we found that a randomly formed group of problem solvers often jointly outperforms a group of the best individuals as ranked by expected value. The explanation substantiated in the second model is that the best problem solvers are less diverse in their problem solving approaches then are the members of a randomly selected group. We now construct a mathematical model that provides sufficient conditions for this result, for a group of randomly selected problem solvers to outperform a group consisting of the problem solvers who perform best individually. We begin with a set of objects X. The set can be finite, denumerable or a continuum. In the main body of the paper, we present a model where X is finite and leave the general model to the appendix. The finite model has been constructed so as to make the insight more obvious. The relationship between the assumptions and the result is much more subtle in the general proof.

Let X be a finite set of objects and $V : X \to [0, 1]$ be a given value function with a unique maximum at x^*, and $V(x^*) = 1$. The problem solvers try to locate a solution that maximizes V, but they have limited abilities. Each problem solver employs a search rule to search for the maximum but does not always end up at x^*. Suppressing the distinction between perspectives and heuristics, we characterize each problem solver by a mapping $\phi : X \to X$. We make several assumptions about problem solvers' mappings. The first is that the mappings locate higher valued solutions and that the problem is difficult.

Assumption 1

(a) $\forall x \in X$, $V(\phi(x)) \geq V(x)$

(b) $\phi(x^*) = x^*$

(c) (Difficulty) There exists $x \in X$, such that $\phi(x) \neq x^*$.

The mapping ϕ has the following interpretation: for each x, $\phi(x)$ denotes the local maximum if the agent starts search at x, that is, it is the stopping point of the search rule ϕ applied to x. In this interpretation, search is deterministic, an initial point uniquely determines a stopping point. The image of the mapping, $\phi(X)$, equals the set of local maxima for problem solver ϕ.

Next, we define ν to be an initial probability distribution on X that assigns a positive probability to each $x \in X$.

Assumption 2

$\nu : X \to [0, 1]$ such that (a) $\forall x \in X$, $\nu(x) > 0$ and (b) $\sum_{x \in X} \nu(x) = 1$

A problem solver ϕ begins search by drawing an initial point according to the probability distribution ν. If the initial point is x, then the search ends at $\phi(x)$. We call the expected value of the search the performance of ϕ, given ν and V. We denote this expected value as $E(\phi, V)$. $E(\phi, V) = \sum_{x \in X} V(\phi(x))\nu(x)$.

Let Φ denote a collection of problem solvers, a set of ϕ's that satisfy Assumption 1. Notice that because the sets of initial and final points are both finite, that Φ must also be finite. The next assumption guarantees a diversity of problem solvers.

Assumption 3 (Diversity)

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

This diversity assumption says that there are no solutions other than the global optimum that are local optima for all of the problem solvers. We next assume a probability distribution μ on Φ that assigns a positive probability to each ϕ in Φ.

Assumption 4

$\mu : \Phi \to [0,1]$ such that (a) $\forall \phi \in \Phi$, $\mu(\phi) > 0$ and (b) $\sum_{\phi \in \Phi} \mu(\phi) = 1$

¿From Φ, the set of problem solvers, we select a group of N agents, each agent is selected independently from Φ according to the probability distribution μ. These N agents are ordered by their individual performance, $E(\phi, V)$. 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 which is formed by selecting each from Φ independently according to μ.

We make the following uniqueness assumption before we present the theorem. We discuss the role this uniqueness assumption plays and its validity after stating and proving the theorem.

Assumption 5 (Uniqueness)

$argmax\{E(\phi, V) : \phi \in \Phi\}$ is unique.

Theorem 4.1 *Suppose V, Φ, ν, and μ satisfy Assumptions 1 - 5. Then, with probability 1, 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 following two ideas are used in the proof. First we show (Lemma 1 below) that for the first random event of drawing independently a group of problem solvers, with probability 1, the joint performance of the group will asymptotically converge to 1 — the best one can hope for. This is quite intuitive given that agents are drawn independently thus are very unlikely to have common local maxima. As the number of agents in the group grows, the probability of them having common local maxima converges to 0. The second idea uses the uniqueness assumption to show that in the second random event, with probability 1, asymptotically there exists a given size such that the subgroup of the best individuals of that size consists of one type of agents, namely, they are all ϕ^* — the unique problem solver of the highest individual ability in Φ. This establishes an upper bound of a value strictly less than 1 on the performance of the best group since ϕ^* can not always reach the global maximum (recall from Assumption 1 that we do not allow any individual problem solver to always locate the optimum.)

Consider the first random event of forming a group of problem solvers, each problem solver is independently selected from Φ according to the probability distribution μ. Fix a sample path of this random event, ω_1. Let $\phi^1(\omega_1), ..., \phi^{n_1}(\omega_1)$ denote the group of the first n_1 problem solvers selected on the sample path ω_1. The joint performance of these n_1 problem solvers is the expected value of $V(\widetilde{y})$ where $\widetilde{y}$ is a common local maximum of all n_1 agents. The distribution of $\widetilde{y}$ is induced by the probability distribution of the initial draw, ν, and a precise model of how agents work together. Here we want to point out that our theorem holds for any specifically given model of agents working together. The proof of the theorem that follows does not depend on a specific model. Without being explicit, we assume that $\widetilde{y}$ follows the probability distribution $\eta_{\omega_1}^{n_1} : X \to [0,1]$, i.e., for any $x \in X$, $\Pr(\widetilde{y} = x) = \eta_{\omega_1}^{n_1}(x)$.

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

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

$$\Pr(A_{n_1}) \leq \Pr\left\{\omega_1 : \phi^1(\omega_1), ..., \phi^{n_1}(\omega_1) \text{ have common local maxima other than } x^*\right\}.$$

Let $m = \min\{\mu(\phi) : \phi \in \Phi\}$. By Assumption 4, $m > 0$. For any $x \in X\backslash\{x^*\}$, we have $\mu(\{\phi \in \Phi : \phi(x) = x\}) \leq 1 - m$. This is because of Assumption 3 that there is at least one ϕ in Φ for whom x is not a local maximum.

By independence,

$$\begin{aligned}
&\Pr\{\omega_1 : \phi^1(\omega_1), ..., \phi^{n_1}(\omega_1) \text{ have common local maxima other than } x^*\} \\
&\leq \textstyle\sum_{x\in X\backslash\{x^*\}} \Pr\{\omega_1 : x \text{ is a common local max of } \phi^1(\omega_1), ..., \phi^{n_1}(\omega_1)\} \\
&\leq \textstyle\sum_{x\in X\backslash\{x^*\}} (1-m)^{n_1} \\
&\leq (|X| - 1)(1-m)^{n_1}
\end{aligned}$$

Therefore,

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

By the Borel-Cantelli Lemma, we have

$$\Pr\left\{\omega_1 : 1 - \sum_{x\in X} V(x)\eta_{\omega_1}^{n_1}(x) > \epsilon \text{ i.o.}\right\} = 0$$

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 the theorem.

Proof of the theorem: Consider the second random event where a group of n agents are drawn independently from Φ according to μ and then a subgroup of a given size that consists of problem solvers with the best individual abilities among n agents is formed. By Assumption 5, the uniqueness assumption, there is a unique problem solver in Φ with the highest individual ability. Call that agent ϕ^*. By the law of large numbers, we have

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

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

$$\Omega = \left\{\omega = (\omega_1,\omega_2) : \begin{array}{l} \lim_{n_1\to\infty} \sum_{x\in X} V(x)\eta_{\omega_1}^{n_1}(x) = 1 \\ \text{and } \lim_{n\to\infty} \frac{\#\left\{i\in\{1,...,n\}:\phi^i(\omega_2)=\phi^*\right\}}{n} = \mu(\phi^*) \end{array}\right\}$$

By Lemma 1, we have

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

Fix any $\omega \in \Omega$. Let $\epsilon_1 = 1 - E(\phi^*, V)$ which is positive since the perfect agent is excluded from our consideration and that ν has X as its support. From the first limit above, we know that there exists an integer $\overline{n}_1 > 0$ such that for any $n_1 \geq \overline{n}_1$,

$$\sum_{x\in X} V(x)\eta_{\omega_1}^{n_1}(x) > 1 - \epsilon_1 = E(\phi^*, V).$$

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

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

Let $N_1 = \overline{n}_1$ and $N = \max\left\{\frac{2\overline{n}_1}{\mu(\phi^*)}, \overline{n}\right\}$. Then

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

The left hand side of the above inequality is the joint performance of the group of N_1 agents independently selected 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, $N \geq \overline{n}$. Therefore,

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

That is,

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

since $N \geq \frac{2\overline{n}_1}{\mu(\phi^*)}$. This means that there are more than N_1 numbers of agents among the group of N agents that are the highest ability agent ϕ^*. Thus, the best N_1 agents among the N agents are all ϕ^*. Obviously their joint performance is exactly the same as the performance of ϕ^* which is $E(\phi^*, V)$. 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. Since the set Ω has probability 1, the theorem is proven.

Remark 4.1 *The uniqueness assumption plays an important role in the proof. It allows us to show that as the group size increases, the best subgroup becomes homogeneous, i.e., they are all the unique best and therefore, they are of no use to each other. This is a stronger assumption than we need. Suppose instead that there is a set of problem solvers that all have the highest expected value. All we need is that the intersection of the images of these problem solvers' mappings has a cardinality strictly greater than one. Thus, we can replace Assumption 5 with the following weaker assumption.*

Assumption $5'$

Problem solvers in $argmax\{E(\phi, V) : \phi \in \Phi\}$ have a common local maximum not equal to x^*.

Remark 4.2 *We might also ask what assumptions would lead to a violation of the uniqueness condition? Suppose the value function, V, does not assign unique values to each element of X. For simplicity, assume that there x' and x'' both have the second highest value under V. Suppose further that the probability distribution ν according to which the initial point of search is drawn, is the uniform distribution on X. Then, there could be two best problem solvers: $\phi^{*1} : X \to X$, who map x' to itself and maps everything else to x^*, the global optimum, and $\phi^{*2} : X \to X$, who maps x'' to itself and maps everything else to x^*. Working together these two agents would always locate the global optimum.*

Remark 4.3 *The previous example violates the spirit of the difficulty assumption. Difficult problems should have many, not just two, local optima for each problem solver. Moreover, the basin of attraction for the global optimum should be small. Problem solvers who are similar will have sets of local optima that do not differ by much. Diverse agents will tend to have less overlap in their sets of local optima. This suggests that the diversity assumption can also be weakened. And, in fact, in the general proof presented in the appendix, we do just that.*

5 Discussion and Extensions

The main results of this paper rely on straightforward logic. If people are bounded, they probably differ in how they solve difficult, i.e. multi-dimensional nonlinear problems. Differences in perspectives and heuristics enable collections of agents to design particle accelerators, pharmaceuticals, and basketball shoes. Being boundedly rational only stifles good decisions if we are boundedly rational in the same way. If the best problem solvers tend to think about a problem similarly, then it stands to reason that as a group they may not be very effective. Random groups may be better owing to their diversity. This paper investigates the strength of this logic. The computational experiments demonstrate the robustness of the argument, the finite model describes a set of sufficient conditions, and the general model provides much weaker sufficient conditions.

If the argument itself seems provocative, the implications are even more so. For problem solving firms, an employee's value depends upon her ability to improve decisions. The diversity of her problem solving approach, as embedded in her perspective–heuristic pair, relative to the other problem solvers appears to be an important predictor of her value. It may be even more relevant than her individual expected value on the problem. Though IQ tests, SAT scores, and college grades may be excellent predictors of raw problem solving ability, they may not be useful in determining the value added of an employee. Maintaining a diversity of problem solving approaches may be as, and possibly more important than hiring people of high individual ability when putting together a group, team, or collection of problem solvers. Therefore, employers might want to focus on the relative uniqueness of applicants' human capital, admittedly something that could prove hard to measure.

The need for a diversity of perspectives and heuristics for firms that solve problems may in part explain the increased prevalence of consultants (Frank and Cook 1996). Firms confronting difficult problems may benefit from bringing in outsiders especially if people within the firm think about problems similarly. Internal group thinking could arise from a corporate culture. Over time, employees may evolve common perspectives and heuristics. Group thinking could also arise if the firm hired only the best people because the best people may tend to think alike. Firms with only a few perspectives, or in extreme cases a single perspective at their disposal could easily become stuck on local optima, generating a need for outside consultants. The consultants need not be "smarter" than the firm to locate improvements. They only need to possess perspectives and heuristics that locate improvements over the status quo.

The results also speak to the power of markets. If the value of an object equals its market value and if rents from locating improvements are appropriable, then markets should lead to improving solutions to economic problems. An agent who locates an improvement to an economic problem receives a rent. In a market setting, anyone knowing of an improvement has an incentive to implement it or to sell their idea to someone who will. The diversity of human perspectives and heuristics implies that no local and non global optimum should be sustainable. Someone eventually builds a better mousetrap, not because that person is smarter than anyone who ever contemplated the mousetrap, but because that person sees or approaches the problem differently. While incentive effects may cause delays in innovations, eventually any improvement should either come to market or become obsolete.

In contrast, the incentives to implement an improvement may be stifled within a problem solving firm. A worker may know how to speed up an assembly line, but the decision may not lie in his problem domain, or the costs of mentioning the improvement may outweigh the benefits. In cases where the improvement is sufficiently large, the worker may have an incentive to create his own competing enterprise, but otherwise, the improvement may never be implemented. It follows that our theory of optimization based on a common value function and diverse problem solving approaches also has strong implications for organizational forms and management styles, especially for problem solving firms such as computer software companies. In an environment where competition depends on continuous innovation and introduction of new products, firms with organizational forms that are decentralized and consist of interdependent confederations of project teams linked by informal communications, and management styles based on teamwork, openness and participation, should flourish. All else equal, firms with these organizational forms and management styles take full advantage of the diverse problem solving approaches of their employees. In a recent book, Saxenian (1994) documents the superior performance of Silicon Valley firms relative to that of Route 128 firms which she attributes partly to the organizational forms and management styles that evolved through regional cultural influences. Our paper provides a theoretical underpinning for such observations.

The ideas put forth in this paper could also be adapted to the study of political decision making. Throughout, we have assumed that problem solvers have identical ordinal rankings of the outcomes. In political contexts this is decidedly not the case. Democrats and Republicans often profess opposing views on tax rates, environmental policy, and welfare plans. Each party probably has its own value function. Nevertheless, an extended version of our model would say that the party in power would still benefit from listening to proposals from the minority party *if the minority party has a different perspective and heuristic.* Of course, the potential for the diversity of perspectives and heuristics to be beneficial should vary with the level of correlation of the value functions.

The current model ignores several important features including incentives, communication, and learning. Incentives matter for two reasons. They could create diverse preferences over outcomes: One problem solver may prefer outcome A to outcome B because she obtains higher income under A, or because she does not have to work as hard to achieve A. And, as we mentioned earlier, incentives may temporarily lead a problem solver to not reveal an improvement because she wants a leg up in searching for subsequent improvements. Revealing a better solution may be informative to competitors. Second, our strong assumption of costless communication reveals another potential extension: the perspective–heuristic framework could be used to provide micro–foundations 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* heuristics and perspectives.

Appendix

In the appendix, we present a general mathematical model that encompasses the finite model in the main body of the paper as a special case. Since the discussion in the appendix is self-contained, the numbering system should not be confused with the numbering system of the main body of the paper.

Consider a set of objects X. X can be a finite, a denumerable or a continuum set. Let $V : X \to [0,1]$ be a given value function which has a unique maximum at x^*, and $V(x^*) = 1$. Again, we consider problem solvers who have limited ability in trying to maximize V. Each problem solver employs some kind of search rule to search for the maximum but does not always end up at x^*. A problem solver is defined by a mapping $\phi : X \to X$ which satisfies the following assumptions:

Assumption 1

(a) $\forall x \in X$, $V(\phi(x)) \geq V(x)$

(b) $\phi(x^*) = x^*$

For each $x \in X$, $\phi(x)$ is the local maximum of the problem solver if the search starts at x. $\phi(X)$ is then the set of local optima for problem solver ϕ. Since X can be a continuum, we need some technical assumptions.

Let $\mathcal{F}$ be a σ-field of X. Let λ be a finite measure of $(X, \mathcal{F})$. Assume that a problem solver ϕ also satisfies the following assumption:

Assumption 2

(a) $\forall x \in X$, $\phi^{-1}(x) \in \mathcal{F}$

(b) $\phi(X)$ is countable.

Let Ψ be the set of all problem solvers that satisfy Assumptions 1 and 2. Then it is easy to show that for any $\phi^1, \phi^2 \in \Psi$, $\{x \in X : \phi^1(x) \neq \phi^2(x)\} \in \mathcal{F}$..

Problem solvers in Ψ may differ only on a set of λ-measure 0. We want to consider such problem solvers as the same.

Definition 1 *Two problem solvers ϕ^1, ϕ^2 from Ψ are called equivalent if*

$$\lambda\left(\left\{x \in X : \phi^1(x) \neq \phi^2(x)\right\}\right) = 0.$$

This defines an equivalence relation on Ψ. Consider the quotient space of Ψ with regard to this equivalence relation. For the rest of the discussion, we are only going to refer to this quotient space. We therefore still denote it by Ψ without confusion. Ψ refers to any set of all problem solvers that are not equivalent. We now define a metric on Ψ.

Definition 2 *Define* $d : \Psi \times \Psi \to \Re_+$ *such that*

$$d(\phi^1, \phi^2) = \lambda\left(\left\{x \in X : \phi^1(x) \neq \phi^2(x)\right\}\right).$$

Lemma 1 (Φ, d) is a metric space.

Proof: To show that the triangle inequality holds, we only need to notice that

$$\{x \in X : \phi^1(x) \neq \phi^2(x)\}$$
$$\subseteq \{x \in X : \phi^1(x) \neq \phi^3(x)\} \cup \{x \in X : \phi^3(x) \neq \phi^2(x)\}.$$

Let ν denote a probability measure on $(X, \mathcal{F})$. A problem solver ϕ starts her search for the maximum of V by first drawing an initial point according to the probability distribution ν. The expected value of the search by ϕ is then defined to be the performance of the problem solver ϕ. $E(\phi, V) = \int_X V(\phi(x))d\nu(x)$.

Assumption 3

ν and λ are mutually absolutely continuous.

Assumption 3 means that sets of ν-measure 0 and sets of λ-measure 0 coincide. Since both are finite measures, they will be treated exactly the same for our purposes without loss of generality. For the rest of the discussion, whenever λ appears, it will be replaced by ν.

Lemma 2 $E(\cdot, V) : (\Psi, d) \to [0, 1]$ is uniformly continuous.

Proof: For any $\phi^1, \phi^2 \in \Psi$,

$$|E(\phi^1, V) - E(\phi^2, V)|$$
$$\leq \int_X |V(\phi^1(x)) - V(\phi^2(x))|\, d\nu(x)$$
$$= \int_{\{x \in X : \phi^1(x) \neq \phi^2(x)\}} |V(\phi^1(x)) - V(\phi^2(x))|\, d\nu(x)$$
$$\leq \nu\left(\{x \in X : \phi^1(x) \neq \phi^2(x)\}\right)$$
$$= d(\phi^1, \phi^2)$$

The uniform continuity follows.

Let Φ be a compact subset of Ψ that satisfies the following assumption:

Assumption 4

(a) $a = \sup\{V(y) : \text{for } y \neq x^* \text{ and } y = \phi(y) \text{ for some } \phi \in \Phi\} < 1$

(b) (Difficulty) There exists a constant p, $0 < p < 1$, such that for any $\phi \in \Phi$, $\nu\left(\{x \in X : \phi(x) = x^*\}\right) <$ p.

Assumption 5 (Uniqueness)

There exist $\phi^* \in \Phi$ and $\overline{\delta} > 0$ such that for any $0 < \delta < \overline{\delta}$, $E(\phi'', V) > E(\phi', V)$ for any $\phi'' \in \{\phi \in \Phi : d(\phi^*, \phi) < \delta\}$ and $\phi' \in \{\phi \in \Phi : d(\phi^*, \phi) \geq \delta\}$.

Consider $E(\cdot, V)$. Since it is continuous on Ψ and Φ is a compact subset of Ψ, there is at least one maximum on Φ for $E(\cdot, V)$. We are actually assuming slightly more than this. We need that in addition that there is a small ball around ϕ^*, the unique maximum, that does not contain any other local maximum for the function. What we are ruling out is a sequence of mappings that have values converging to the value of ϕ^* that are all local optima.

Let μ be a probability measure on $(\Phi, \mathcal{B})$ where $\mathcal{B}$ is the set of Borel sets of Φ generated by the metric d. μ is a measure on the space of problem solvers. Assume

Assumption 6

(a) Every open set of (Φ, d) has positive μ-measure.

(b) (Diversity) There exists a q, $0 < q < 1$, and a finite partition of $X \backslash \{x^*\}$ denoted by $\{B_1, ..., B_k\}$ s.t. $\forall k = 1, ...K$,

$$\mu\left(\{\phi \in \Phi : \exists y \in B_k, \text{ s.t. } y = \phi(y)\}\right) < q.$$

The diversity assumption says that the problem space can be partitioned into a finite number of regions such that the probability of problem solvers who have local maxima in any given region is bounded away from probability 1.

With this general model, we still have the following theorem:

Theorem 5.1 *Suppose V, Φ, ν, and μ satisfy Assumptions 1 - 6. Then, with probability 1, 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 problem solvers independently drawn according to μ exceeds the joint performance of the N_1 individually best problem solvers among the group of N problem solvers independently drawn from Φ according to μ.*

The proof of the theorem follows ideas similar to the proof in the finite model. The details of the proof however are more elaborate given the general nature of the model. We prove the theorem with the help of two lemmas. Lemma 3 below shows when we draw problem solvers independently from Ω, that with probability 1, the joint performance of the problem solvers converges to the global optimum asymptotically. This lemma relies on the assumption of diversity. The diversity assumption guarantees that the intersection of the sets of local maxima of the problem solvers shrinks to a point: the global optimum. In Lemma 4, we establish an upper bound for the joint performance of a group of the best problem solvers. The proof of this lemma depends heavily on the uniqueness assumption. From a large enough sample of problem solvers, the best problem solvers all lie near the best problem solver in d-distance. Therefore, they tend to have similar sets of local maxima.

In the proof, we consider a randomly formed group of problem solvers and compare its performance to that of a group of the best solvers. Consider first the random event of forming a group of problem solvers, each problem solver is independently selected from Φ according to the probability distribution μ. Fix a sample path of this random event, ω_1. Let $\phi^1(\omega_1), ..., \phi^{n_1}(\omega_1)$ denote the group of the first n_1 problem solvers selected on the sample path ω_1. The joint performance of these n_1 problem solvers is the expected value of $V(\widetilde{y})$ where $\widetilde{y}$ is a common local maximum of all n_1 problem solvers. Assume that $\widetilde{y}$ follows the probability distribution $\eta_{\omega_1}^{n_1}$. The joint performance of $\phi^1(\omega_1), ..., \phi^{n_1}(\omega_1)$ is $\int_X V(x) d\eta_{\omega_1}^{n_1}(x)$.

Lemma 3 $\Pr\left\{\omega_1 : \lim_{n_1\to\infty} \int_X V(x) d\eta_{\omega_1}^{n_1}(x) = 1\right\} = 1$

Proof: Fix any $0 < \epsilon < 1$. Define $A_{n_1} = \left\{\omega_1 : 1 - \int_X V(x) d\eta_{\omega_1}^{n_1}(x) > \epsilon\right\}$. Obviously, $A_{n_1} \subseteq \{\omega_1 : \phi^1(\omega_1), ..., \phi^{n_1}(\omega_1) \text{ have common local max other than } x^*\}$. Thus,

$\Pr(A_{n_1}) \leq \Pr\{\omega_1 : \phi^1(\omega_1), ..., \phi^{n_1}(\omega_1) \text{ have common local max other than } x^*\}$.

By diversity and independence,

$$\begin{aligned}
&\Pr\{\omega_1 : \phi^1(\omega_1), ..., \phi^{n_1}(\omega_1) \text{ have common local max other than } x^*\} \\
&\leq \textstyle\sum_{k=1}^{K} \Pr\{\omega_1 : \phi^1(\omega_1), ..., \phi^{n_1}(\omega_1) \text{ have common local max in } B_k\} \\
&\leq \textstyle\sum_{k=1}^{K} \Pr\{\omega_1 : \phi^1(\omega_1), ..., \phi^{n_1}(\omega_1) \text{ each has local max in } B_k\} \\
&\leq \textstyle\sum_{k=1}^{K} \left[\mu\left(\{\phi \in \Phi : \exists y \in B_k, \text{ s.t. } y = \phi(y)\}\right)\right]^{n_1} \\
&\leq K q^{n_1}
\end{aligned}$$

Therefore,

$$\sum_{n_1=1}^{\infty} \Pr(A_{n_1}) = \frac{K}{1-q} < \infty.$$

By Borel-Cantelli Lemma, we have

$$\Pr\left\{\omega_1 : 1 - \int_X V(x) d\eta_{\omega_1}^{n_1}(x) > \epsilon \text{ i.o.}\right\} = 0$$

which implies

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

In the next lemma, we give an upper bound for the joint performance of a group of n problem solvers who are close to each other in d.

Lemma 4 For any $\epsilon > 0$ and any positive integer n such that $(n-1)\epsilon < 1-p$ (Recall p appears in Assumption 4), consider any n problem solvers $\phi^1, ..., \phi^n$ that are within ϵ d-distance from each other, i.e. $d(\phi^i, \phi^j) < \epsilon$ for any $i, j \in \{1, ..., n\}$. Then the joint performance of these n problem solvers, denoted by $E(\phi^1, ..., \phi^n; V)$, is bounded by $1 - (1-a)[1 - (n-1)\epsilon - p]$. I.e.,

$$E(\phi^1, ..., \phi^n; V) < 1 - (1-a)[1 - (n-1)\epsilon - p].$$

Proof: Suppose the group starts their search at x such that $\phi^1(x) = ... = \phi^n(x) \neq x^*$. Then the search of the group will get stuck at $\phi^1(x) = ... = \phi^n(x)$ which is a common local max of all n problem solvers but is not a global max x^*. We consider the set of such x. If such set has a positive measure, then the joint performance of these n problem solvers will be less than 1. Notice that

$$\begin{aligned}&\{x \in X : \phi^1(x) = ... = \phi^n(x) \neq x^*\} \\ &= \{x \in X : \phi^1(x) = ... = \phi^n(x)\} \\ &\qquad \setminus \{x \in X : \phi^1(x) = ... = \phi^n(x) = x^*\}.\end{aligned}$$

First, since

$$\begin{aligned}&\{x \in X : \phi^1(x) = ... = \phi^n(x)\} \\ &= X \setminus \cup_{i=1}^{n-1} \{x \in X : \phi^1(x) = ... = \phi^i(x) \neq \phi^{i+1}(x)\}\end{aligned}$$

we have

$$\begin{aligned}&\nu(x \in X : \phi^1(x) = ... = \phi^n(x)) \\ &= 1 - \textstyle\sum_{i=1}^{n-1} \nu(\{x \in X : \phi^1(x) = ... = \phi^i(x) \neq \phi^{i+1}(x)\}) \\ &\geq 1 - \textstyle\sum_{i=1}^{n-1} \nu(\{x \in X : \phi^1(x) \neq \phi^{i+1}(x)\}) \\ &= 1 - \textstyle\sum_{i=1}^{n-1} d(\phi^1, \phi^{i+1}) \\ &> 1 - (n-1)\epsilon\end{aligned}$$

The last inequality holds since $d(\phi^i, \phi^j) < \epsilon$ for any $i, j \in \{1, ..., n\}$. Now consider $\{x \in X : \phi^1(x) = ... =$ Since

$$\{x \in X : \phi^1(x) = ... = \phi^n(x) = x^*\} \subseteq \{x \in X : \phi^1(x) = x^*\},$$

we have

$$\begin{aligned}&\nu(\{x \in X : \phi^1(x) = ... = \phi^n(x) = x^*\}) \\ &\leq \nu(\{x \in X : \phi^1(x) = x^*\}) < p.\end{aligned}$$

The last inequality is because of part (b) of Assumption 4. Thus we have

$$\begin{aligned}&\nu(\{x \in X : \phi^1(x) = ... = \phi^n(x) \neq x^*\}) \\ &= \nu(\{x \in X : \phi^1(x) = ... = \phi^n(x)\}) \\ &\qquad -\nu(\{x \in X : \phi^1(x) = ... = \phi^n(x) = x^*\}) \\ &> 1 - (n-1)\epsilon - p > 0\end{aligned}$$

given that $(n-1)\epsilon < 1 - p$.

Then, the joint performance of this group of n problem solvers has the following property:

$$\begin{aligned}&E(\phi^1, ..., \phi^n; V) \\ &\leq V(x^*)[1 - \nu(\{x \in X : \phi^1(x) = ... = \phi^n(x) \neq x^*\})] \\ &\qquad + a\nu(\{x \in X : \phi^1(x) = ... = \phi^n(x) \neq x^*\}) \\ &< 1 \cdot [1 - [1 - (n-1)\epsilon - p]] + a \cdot [1 - (n-1)\epsilon - p] \\ &= 1 - (1-a)[1 - (n-1)\epsilon - p].\end{aligned}$$

The first inequality holds because the right hand side is the joint performance if (1) whenever problem solvers end up with different $\phi^i(x)$, they are eventually taken to x^* (2) whenever problem solvers get stuck together, they achieve the highest value for all the local max. Both assume the most optimism.

Now we prove the theorem using Lemma 3 and Lemma 4.

Proof of the Theorem: Let $\Omega_1 = \left\{\omega_1 : \lim_{n_1\to\infty} \int_X V(x) d\eta_{\omega_1}^{n_1}(x) = 1\right\}$. Let δ be a positive number such that $0 < \delta < (1-p)(1-a)$. Then for any $\omega_1 \in \Omega_1$, there exists $\overline{n}_1(\omega_1) > 1$ such that for any $n_1 \geq \overline{n}_1(\omega_1)$, $\int_X V(x) d\eta_{\omega_1}^{n_1}(x) > 1 - \delta$.

Let $\epsilon(\overline{n}_1(\omega_1)) = \min\left\{\frac{1}{\overline{n}_1(\omega_1)-1}\left(1 - p - \frac{\delta}{1-a}\right),\ \overline{\delta}\right\}$. Recall that $\overline{\delta}$ is a parameter in the uniqueness assumption (Assumption 5). Since $\delta < (1-p)(1-a)$ and $\overline{\delta} > 0$, $\epsilon(\overline{n}_1(\omega_1)) > 0$.

For any $\epsilon > 0$, let $O(\phi^*, \epsilon) = \{\phi \in \Phi : d(\phi, \phi^*) < \epsilon\}$ where ϕ^* is the unique best problem solver in Φ. Since $O(\phi^*, \epsilon)$ is open, by part (a) of Assumption 6, $\mu\left(O(\phi^*, \epsilon)\right) > 0$. Consider $\epsilon = \frac{\epsilon(\overline{n}_1(\omega_1))}{2}$. Let $\overline{\mu}$ denote $\mu(O(\phi^*, \frac{\epsilon(\overline{n}_1(\omega_1))}{2}))$. Consider the second random event mentioned above. Define

$$\Omega_2(\overline{n}_1(\omega_1)) = \left\{\omega_2 : \lim_{n\to\infty} \frac{\#\left\{i \in \{1, ..., n\} : d(\phi^i(\omega_2), \phi^*) < \frac{\epsilon(\overline{n}_1(\omega_1))}{2}\right\}}{n} = \overline{\mu}\right\}.$$

By the law of large numbers, $\Pr(\Omega_2(\overline{n}_1(\omega_1))) = 1$. This is true for every $\omega_1 \in \Omega_1$.

Then $\forall \omega_2 \in \Omega_2(\overline{n}_1(\omega_1))$, $\exists \overline{n} > 1$ s.t. $\forall n \geq \overline{n}$,

$$\frac{\#\left\{i \in \{1, ..., n\} : d(\phi^i(\omega_2), \phi^*) < \frac{\epsilon(\overline{n}_1(\omega_1))}{2}\right\}}{n} > \frac{\overline{\mu}}{2}$$

or equivalently,

$$\#\left\{i \in \{1, ..., n\} : d(\phi^i(\omega_2), \phi^*) < \frac{\epsilon(\overline{n}_1(\omega_1))}{2}\right\} > \frac{\overline{\mu}}{2} \cdot n.$$

For any $\omega_1 \in \Omega_1$ and $\omega_2 \in \Omega_2(\overline{n}_1(\omega_1))$, let $N_1 = \overline{n}_1(\omega_1)$ and $N = \max\left\{\frac{2}{\overline{\mu}} \cdot \overline{n}_1(\omega_1),\ \overline{n}\right\}$. Since $N \geq \overline{n}$, we have

$$\begin{aligned}
&\#\left\{i \in \{1, ..., N\} : d(\phi^i(\omega_2), \phi^*) < \tfrac{\epsilon(\overline{n}_1(\omega_1))}{2}\right\} \\
&> \tfrac{\overline{\mu}}{2} \cdot N \\
&\geq \overline{n}_1(\omega_1) \\
&= N_1.
\end{aligned}$$

This means that among N problem solvers along the path ω_2, there are more than N_1 problem solvers who are within $\frac{\epsilon(\overline{n}_1(\omega_1))}{2}$ d-distance from ϕ^*. Since $\frac{\epsilon(\overline{n}_1(\omega_1))}{2} < \epsilon(\overline{n}_1(\omega_1)) \leq \overline{\delta}$, by the uniqueness assumption (Assumption 5), the individual performance of each such problem solver is strictly better than that of problem solvers who are more than $\frac{\epsilon(\overline{n}_1(\omega_1))}{2}$ d-distance away from ϕ^*. Therefore, the best N_1 problem solvers among the group of N problem solvers are all strictly within $\frac{\epsilon(\overline{n}_1(\omega_1))}{2}$ d-distance from ϕ^*. By the triangular property of d, these N_1 best problem solvers are strictly within $\epsilon(\overline{n}_1(\omega_1))$ distance from each other.

By definition, $\epsilon(\overline{n}_1(\omega_1)) \leq \frac{1}{N_1-1}\left(1 - p - \frac{\delta}{1-a}\right)$. Since $\delta > 0$, we have $(N_1-1)\epsilon(\overline{n}_1(\omega_1)) < 1 - p$. By Lemma 4, the joint performance of these N_1 best problem solvers among those

N problem solvers is strictly less than $1 - (1-a)[1 - (N_1 - 1)\epsilon(\overline{n}_1(\omega_1)) - p]$. But since $\epsilon(\overline{n}_1(\omega_1)) \leq \frac{1}{N_1 - 1}\left(1 - p - \frac{\delta}{1-a}\right)$, $1 - (1-a)[1 - (N_1 - 1)\epsilon(\overline{n}_1(\omega_1)) - p] \leq 1 - \delta$. Here notice that $N_1 = \overline{n}_1(\omega_1)$.

¿From the beginning of the proof, we have $\int_X V(x) d\eta_{\omega_1}^{N_1}(x) > 1 - \delta$. Thus the joint performance of the N_1 problem solvers along ω_1 is better than the joint performance of the N_1 best problem solvers among the group of N problem solvers along ω_2.

Let $\Omega_2 = \cap_{\overline{n}_1(\omega_1)} \Omega_2(\overline{n}_1(\omega_1))$. Since it is a countable intersection and $\Pr(\Omega_2(\overline{n}_1(\omega_1))) = 1$ for each $\overline{n}_1(\omega_1)$, $\Pr(\Omega_2) = 1$. Regarding Ω_1, by Lemma 3, we have $\Pr(\Omega_1) = 1$.

To summarize then, we have for any $\omega_1 \in \Omega_1$ and $\omega_2 \in \Omega_2$, there exist positive integers N_1 and N, $N_1 < N$, such that the joint performance of the N_1 problem solvers along ω_1 is better than the joint performance of the N_1 best problem solvers among the group of N problem solvers along ω_2. Since ω_1 and ω_2 are independent and $\Pr(\Omega_1) = 1$ and $\Pr(\Omega_2) = 1$, we have $\Pr(\Omega_1 \times \Omega_2) = 1$.

References

Arifovic, J. (1994). "Genetic algorithm learning and the cobweb model", *Journal of Economic Dynamics and Control* 18, 3-28.

Beard, Charles A. (1957) *The Economic Basis of Politics*

Becker, Gary (1973) "A Theory of Marriage Part 1:" *Journal of Political Economy* 81, 813–846.

Cho, I. (1993). "Perceptrons Play Repeated Prisoner's Dilemma" mimeo, University of Chicago.

Dawes, Robin (1988) *Rational Choice in an Uncertain World* Harcourt Brace Jovanovich, Orlando, Florida.

Frank, Robert H. and Phillip J. Cook (1996) *The Winner Take All Society* The Free Press, New York.

Gilboa, Itzhak, (1988) "The Complexity of Computing Best-Response Automata in Repeated Games," *Journal of Economic Theory* 45 342-352.

Hernnstein, Richard, and Charles Murray (1994) *The Bell Curve*, The Free Press, New York.

Hong, L. and S. Page (1997). "Problem Solving by Heterogeneous Agent" mimeo.

Kalai, E. and W. Stanford (1988). "Finite rationality and interpersonal complexity in repeated games", *Econometrica* 56, 397-410.

Marimon, R., E. McGrattan and T. Sargent (1990). "Money as a medium of exchange in an economy with artificially intelligent agents", *Journal of Economic Dynamics and Control* 14, 329-374.

McCloud, Bentley (1996) "Decision, Contract, and Emotion: Some Economics for a Complex and Confusing World" *Canadian Journal of Economics* November pp. 788–810.

Miller, John (1996) "The Coevolution of Automata in the Repeated Prisoner's Dilemma." *Journal of Economic Behavior and Organizations* 29: pp 87–113.

Mortensen, Dale T. (1988) "Matching: Finding a Partner For Life or Otherwise" *American Journal of Sociology* 94, S15–S240

Nisbett, Richard, and Lee Ross (1980) *Human Inference: Strategies and Shortcomings of Social Judgment*, Prentice–Hall, Englewood Cliffs, NJ.

Page, Scott E. (1996) "Two Measures of Difficulty" *Economic Theory* 8:321-346.

Page, Scott and Mike Ryall (1998) "Does Strategy Need Computer Experimentation?" forthcoming in *Advances in Strategic Management*, ed. Joel A. C. Baum, no. 15, Greenwich: JAI Press Inc..

Radner, R. (1993). "The organization of decentralized information processing", *Econometrica* 61, 1109-1146.

Radner, R. and T. Van Zandt (1995). "Information Processing in Firms and Returns to Scale", in *The Economics of Informational Decentralization: Complexity, Efficiency, and Stability,* ed. John Ledyard, Kluwer Academic Publishers.

Robbins, Stephen (1994) *Organizational Behavior* Prentice Hall, Saddle River New Jersey.

Rubinstein A. (1986). "Finite automata play the repeated prisoner's dilemma", *Journal of Economic Theory* 39, 83-96.

Saxenian, AnnaLee (1994) *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*, Harvard university Press.

Thomas, David A and Robin J Ely "Making Differences Matter: A New Paradigm for Managing Diversity" *Harvard Business Review* September–October 1996 #96510

Watson, W., K. Kumar, and L. Michaelsen (1993). "Cultural Diversity's Impact on Interaction Process and Performance: Comparing Homogeneous and Diverse Task Groups", *Academy of Management Journal* 36, 590-602.