CHAPTER 6

Diversity and Problem Solving

DARWIN'S BRASS TACKS

Talent hits a target that no one else can hit; genius hits a target no one else can see.
— attributed to Arthur Schopenhauer

NOW that we've covered what cognitive diversity is—we've defined perspectives, heuristics, and so on—we get to the meat of the book: the demonstration that diversity produces benefits. We see why diversity may be as important as ability in some contexts, and how it can be even more important in others. In short, we show that diversity creates benefits. The proofs rely on the toolbox framework. We'll see how perspectives, heuristics, interpretations, and predictive models aggregate. Throughout, I avoid making blanket statements that diversity is always good or always bad. Blankets cover things, and I do not want to cover the particulars. Whether and how diversity improves performance on a task depends on the type of diversity considered and the type of task. We should not expect any one type of diversity to be beneficial in all contexts any more than we should expect friction to hinder performance in all contexts. When we want to stop the car, we like friction.

In these next three chapters, we see why diversity is such a powerful force. We see how our differences contribute as much to our collective performance as our individual abilities. We will take an abstract, logical approach. By being abstract, we can apply our logic across a variety of contexts—to everything from identity group politics to stock market predictions to interdisciplinary science. Some of what follows may take a little time to absorb, but it's all there for a reason and it all has a big payoff. If you've wondered if, how, and why diversity creates benefits, you're about to find out.

Problem Solving

We begin with problem solving. We restrict attention to difficult problems. We don't need diverse perspectives and heuristics to help us with easy ones. Two plus two is easy. Protein folding is difficult. We face no shortage of difficult problems—finding renewable sources of energy, designing health policies, and managing ecosystems receive the most attention, but difficult problems also include designing buildings, writing examinations, and producing movies. Difficult problems are nothing new. Animal domestication and developing the steam engine do not fall into anyone's "easy" category. But many people believe that the dimensionality of current problems and the linkages between them result in greater problem difficulty than existed in the past.[1]

Our analysis highlights the contextual nature of an individual's contribution. How much a person improves a solution depends on how her tools combine with and differ from the tools of the other problem solvers. Yes, her contribution also depends on her ability to find good solutions on her own, but we should not equate intelligence with individual contributions. That's a mistake. Actually, it's two mistakes. First, doing so ignores the difficulty of the problem. We don't want to reward people who happened to have picked or been assigned relatively easy problems, and to underestimate the abilities of people who worked on problems that lie beyond our current collective abilities. Few know the names of the many scientists who unsuccessfully attempted to develop a workable form of fusion in the late 1970s and early 1980s. They've slipped into scientific oblivion because fusion proved too difficult. Had fusion been an easier problem given the perspectives and tools of the day, several people who had modest scientific careers might now be hailed as geniuses. The same is true of all those alchemists who tried to turn lead into gold, and of those people who have tried to build perpetual motion machines. Good try, but hard (in fact, impossible) problems. Second, we assume that the person who makes the discovery has higher intelligence, when that person may just be different. Recall that breakthroughs often come from diverse perspectives. Having a diverse perspective is not the same as being smart. Someone with very few tools—but the right ones for the task—may make a breakthrough.

Our investigation of how individual diversity aggregates culminates in the *Diversity Trumps Ability Theorem*. This theorem provides conditions under which collections of diverse individuals outperform collections of more individually capable individuals. As mentioned in the prologue, this result was not something expected or desired. It just popped out of some experiments with agent-based models that I ran as an assistant professor at Caltech.

A Little Old Logic from Pasadena

I now fill in more of the details of those agent-based models. This detailed analysis helps us see why the theorem holds and what conditions are needed. First, we need some methodological background. An agent-based model consists of artificial agents—computer-based objects that interact in time and space according to rules described in computer code.[2] These agents can represent almost anything: viruses, nations, birds, fish, teenagers, firms, or politicians. The behavioral repertoires of ants and teenagers differ—teenagers are a bit more sophisticated, although not by as much as you might think—but you can model either. In a well-constructed agent-based model, the interactions among the agents and between the agents and the environment tell us something about the real world: how prices emerge in a market or even how riots start. The agent-based models that make the pages of science magazines generate beautiful, structured patterns, like flocking birds, but more often than not these models generate unintelligible muddles.

My model included agents who tried to solve hard problems. We can think of them as making advances on fusion or as trying to figure out how to make every popcorn kernel pop. The objective for the agents in my model was to improve on the existing solutions to a set of predefined problems. The larger the improvement an agent found, the more money the agent received.[3] In the first models, I endowed agents with random perspectives and heuristics. These random assignments created cognitively diverse agents. I did not differentiate the agents by identities. I didn't paint them different colors. Such was my intelligent design.

I had planned to allow the agents to learn from one another and to experiment with new perspectives and heuristics. I hoped to explore the trade-off between exploitation (copying and learning from others) and exploration (searching for new representations and search algorithms). That tension exists in many contexts. Evolution confronts it. Organizations confront it. Individuals confront it. The tension is this. Sure, exploration can be risky (think Magellan), but the benefits can be huge (think Cortez). However, if we explore too much, we never take advantage of what we learn. Exploitation produces guaranteed benefits, but if everyone exploits, they have no new ideas to exploit.

When writing an agent-based model or when proving a mathematical theorem, catching errors is crucial. Computer programs contain hundreds, if not thousands of lines of code. One misplaced semicolon or bracket can alter the model's performance. After writing the initial code, I ran some experiments to check for errors. If I gave an agent greater cognitive ability than the other agents (I did this in the form of more perspectives and heuristics), did that first agent tend to make more money? Yes. If I gave all of the agents more perspectives and heuristics, did the average performance on the problems increase? Yes. If I created more agents, did average performance increase? Again, yes. Finally, I created two economies, one consisting only of agents who performed best individually and another with random, but smart, agents. Did the agents in the first economy, those more able as individuals, on average find better solutions to the problems than did the agents in the second economy?

No.

A Cup of Joe

This last finding, the one I mentioned in the prologue, ran counter to my intuition. I assumed a coding error. I checked and found none, so I rewrote the same model in a different computer language. Same result: the society composed of the better individual agents performed worse. I decided to strip my model down to its core. I created an economy with a single problem. I decided that problem should be something of great importance—making the perfect cup of coffee for my wife. Not being a coffee drinker, I had been greatly vexed by this problem. The coffee problem had only two dimensions to it. Agents tried to find the ideal amount of cream (dimension one) and sugar (dimension two).

This construction allowed me to represent the set of possible solutions on a square. Each point on this square has a horizontal displacement (the infamous x from seventh-grade algebra) and a vertical displacement (x's partner in crime, y). The x value represents the amount of cream in the coffee and the y value as the amount of sugar. In this way, each point within the square represents a unique combination of cream and sugar (see figure 6.1).

Each point in this coffee space also can be assigned a quality—how much my wife liked the coffee. In figure 6.1 quality is shown as the third dimension. I assumed that qualities varied between zero and one hundred: the more she liked the cup of coffee, the higher the value. The problem my agents faced was to find the highest point. Recall from our earlier discussion of perspectives that we can think of the value of each point in the square as an elevation. High-quality cups have high elevations, and low-value cups have low elevations. This representation creates a rugged landscape.

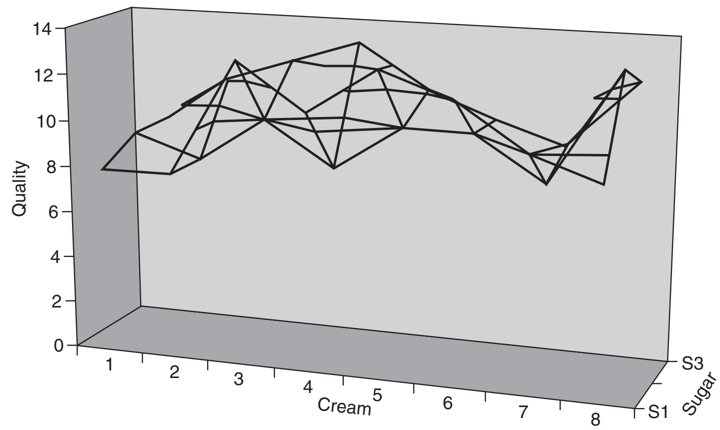

Figure 6.1 The Coffee Space

My agents used search algorithms to roam around this two-dimensional landscape.[4] Placed at any point, an agent would look in two directions. I assigned these directions randomly. If an agent saw a point with a higher value, it would move to that point. An agent followed this search rule until it could not see any cups of higher value. The points at which they got stuck would not necessarily be the best cup, owing to the ruggedness of the landscape. In fact, they rarely were.

To calculate the performance for a group, I allowed one agent in the group to search for a good cup of coffee. That agent continued until it reached a point where it could find no improvement. I then let the next agent start at that point (that cup of coffee) and try to find an even better one. If the second agent found an improvement, the first agent would then try to find further improvements. I continued this process until no agent in the group could find a better cup of coffee.

The model's simplicity made the results more transparent. I could represent an agent graphically by the two directions it searched. When I plotted the two search directions of the best-performing agent, I saw that both directions pointed toward the upper right-hand corner of the square. I then plotted the search directions for each of the best ten agents. I saw a similar pattern in their search directions. They all searched toward the upper right-hand corner of the square. They were all heading to the same place to find coffee. When I plotted the search direction vectors for the random agents, I saw a different pattern. They pointed every which way, like the spokes of a bicycle wheel.

What had been a counterintuitive result now had a logical explanation: The best problem solvers tend to be similar; therefore, a collection of the best problem solvers performs little better than any one of them individually. A collection of random, but intelligent, problem solvers tends to be diverse. This diversity allows them to be collectively better. Or to put it more provocatively: *diversity trumps ability.*

Green Eggs and Ham

I wondered, had I reached into the haystack of possible models and pricked myself on a needle? How general was this finding? To see, I constructed several models of problem solving agents. I varied the space of possible solutions from squares, to lines, to trees, to spheres. And I varied the sophistication of the agents. Most of the time, I found the same result: diversity trumped ability. There did seem to be some necessary conditions, though. I had to create enough agents, I had to assume groups of moderate size, I had to make the problem difficult, and I had to make all of the problem solvers pretty smart. These conditions all made intuitive sense. A collection of third graders probably has little to add to the study of global warming.

These computational experiments were suggestive but not definitive proof that diversity trumped ability. I took to calling this a "green eggs and ham" result in homage to Dr. Seuss's book in which Sam tries to convince his friend to try green eggs and ham in a variety of locales. The friend, as it turns out, will not eat them on a bus or in a box or on a train or with a fox. The same was true of this result. It was true in a box or on a tree or on a graph. It was true anywhere.

But I had a problem, one related to my profession. Although computational experiments persuade most physicists, mathematicians, and biologists, they are not the coin of the realm for economists, and I was paid to be an economist. Economists believe that the computer experiments are correct, but they prefer formal theorems and proofs. Formal logic helps us understand exactly when and why a result holds. To go about constructing a proof, I contacted Lu Hong and pleaded for help. Lu and I hammered out a mathematical proof that provides sufficient conditions for the green eggs and ham result. In 2004, we published these results in the *Proceedings of the National Academy of Sciences*.[5]

The Crowded Chess Table

Could this be true in the real world? Could diverse teams outperform teams of high-ability people? There is one famous example involving the game of chess. On June 21, 1999, Garry Kasparov began a game of chess against approximately 50,000 other players. At the time, Kasparov was the reigning world champion. Some of these people in the group that he was playing were rank amateurs. MSN.com sponsored the game to show the power of the Internet in performing collective problem solving. Kasparov moved first. Each subsequent move took place forty-eight hours later. To determine their move, the people playing Kasparov posted notes to a bulletin board and voted on what to do.

These people had some help. Each time the crowd had to make a move, a team of young chess masters, none older than twenty, proposed an initial set of possible moves. Though great young chess players, they were not of the caliber of Kasparov. At the end of the forty-eight hours, a vote was held. The move that received the most votes was then played. Kasparov then had forty-eight hours to make his next move. After sixty-two moves, Kasparov won. In chess, sixty-two moves suggests a close game. The collection of people performed far better than would be expected of any of its members individually.

Individual Diversity and Collective Problem Solving

It's worth our time and effort to see how a diverse group can outperform a group of more able individuals in solving a problem. To show this result, we need to use perspectives and heuristics. We'll begin with a model of a problem solver with perspectives and heuristics that can be applied to the problem at hand.[6] We consider perspectives first, then heuristics, and then compare the two types of diversity. To keep things as simple as possible, when we study diverse perspectives, we assume that everyone uses the same heuristic, and when we study diverse heuristics, we assume that everyone uses the same perspective.

In undertaking this analysis, we see differences and similarities between diverse perspectives and diverse heuristics. We even see how these two types of diversity can be equivalent. This equivalence can be overemphasized. In a person, perspectives and heuristics operate differently. Diverse perspectives are more likely to lead to breakthroughs and to create communication problems. Diverse heuristics are more likely to lead to smaller, more iterative improvements.

Diverse Perspectives and Problem Solving

Remember, a perspective is a "one-to-one" mapping of reality into some internal language. Two people have diverse perspectives if they map reality into different internal languages, or if they map the reality differently into the same internal language. To capture diverse perspectives and common heuristics, we consider an example in which multiple perspectives map the solutions onto a line. Suppose that we have a lot containing a thousand cars, and the problem is to find the car that gets the best gas mileage. Each car has a fact sheet that contains all relevant information *except* its miles per gallon. Determining gas mileage is costly; it requires taking a car out for a long drive.

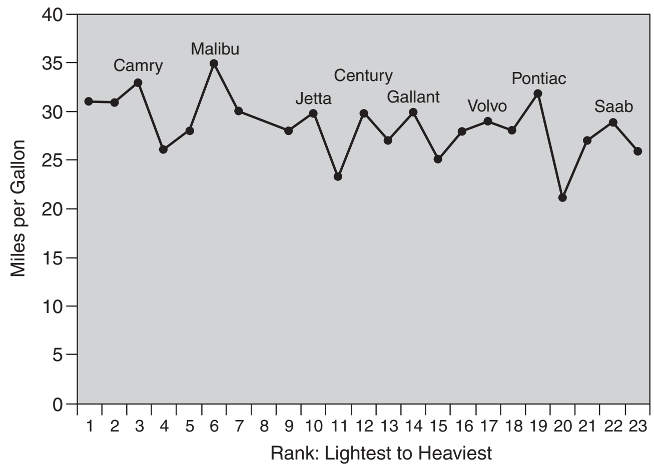

Figure 6.2 MPG in Curb Weight Perspective

Many possible perspectives could be used to arrange these cars. Someone might arrange them by weight, thinking that the heavier the car, the worse the mileage. Someone else might arrange them by height, thinking that aerodynamics matter. A third person might arrange them by their wheelbase, thinking that it is good proxy for the overall size of the car. Each of these perspectives embeds knowledge of the causes of better gas mileage, but none perfectly predicts gas mileage. Therefore, the resulting one-dimensional landscape will be rugged, but not too rugged. Other perspectives that fail to embed understanding—say, arranging the cars by their color or by the diameter of their headlights—would result in very rugged landscapes.

To ground our logic, let's apply these perspectives to actual data on twenty-three 2005 model year midsize sedans (for which we know the official gas mileage).[7] The first graph (figure 6.2) shows the landscape created by curb weight. This perspective has eight local optima: Toyota Camry, Chevy Malibu, Volkswagen Jetta, Buick Century, Mitsubishi Gallant, Volvo S60, Pontiac G6, and Saab 9-5. Our next perspective (figure 6.3) considers the width of the wheelbase. This perspective has seven local

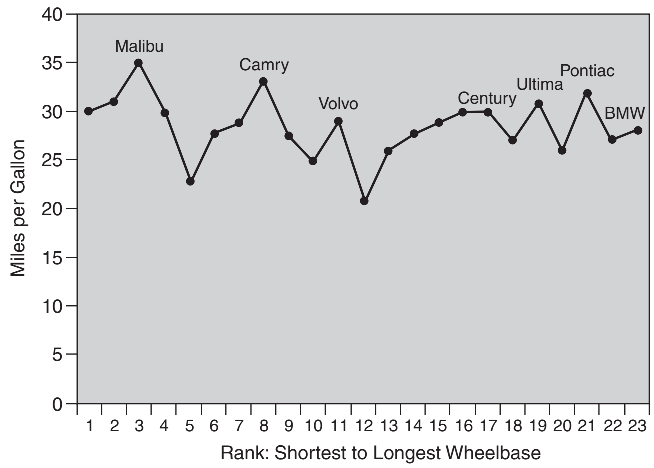

Figure 6.3 MPG in Wheelbase Perspective

optima: Chevy Malibu, Toyota Camry, Volvo S60, Buick Century, Nissan Ultima, Pontiac G6, and BMW 525. Our final perspective (figure 6.4), based on height, creates nine local optima: Dodge Stratus, Hyundai Sonata, Volvo S60 Mazda 6, Pontiac G6, Volkswagen Jetta, Suzuki Verona, Chevy Malibu, and Toyota Camry.

We assume that our testers start at some random point in their landscape (a random car) and test it to determine its miles per gallon. Each then tests a neighboring car. If that car gets better mileage, the problem solver continues in the same direction, testing the next car until finding a local peak. If the second car tested gets worse mileage than the first car tested, the problem solver tests in the opposite direction (if necessary) and searches until reaching a local optimum. Using this heuristic, each problem solver gets stuck on peaks in her landscape. Recall that test-driving these cars to determine their miles per gallon takes time and effort. That's why our problem solvers search locally given their perspectives.

Imagine these three people working as a group. Each landscape has a few local peaks, but a locally optimal solution for one person may not be a local optimum for the others. If they work together,

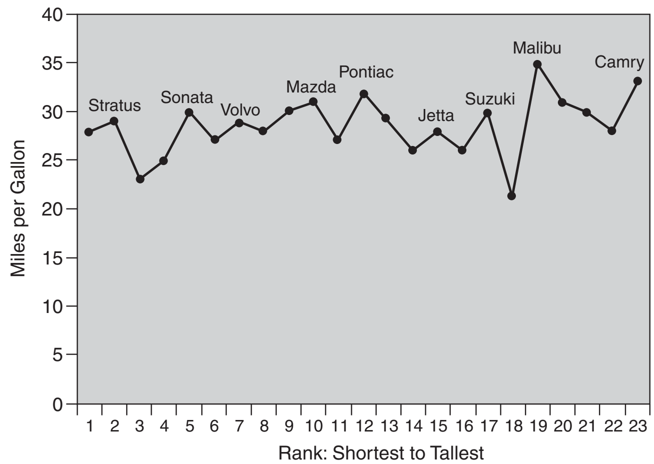

Figure 6.4 MPG in Height Perspective

when one person gets stuck at a local optimum, another person may be able to find a subsequent improvement. That second person's improved solution may then be further improved on by someone else.

For a car to be a local peak for the group, it must be a local peak for *every* member of the group. Putting the names of the local peaks in a table reveals that four cars: the Chevy Malibu (35 mpg), the Toyota Camry (33 mpg), the Pontiac G6 (32 mpg), and the Volvo S60 (29 mpg) are local peaks on all three landscapes. If the problem solvers know this information, they'd pick the best peak, but they don't. They get stuck on one of the peaks. The Malibu, Camry, and Pontiac are ranked first, second, and third in gas mileage, and the Volvo is ranked tied for tenth (see table 6.1).

Note several features of this example. First, any one of these perspectives might locate the Malibu, the car with the best mileage, but the ruggedness of their landscapes prevents them from always locating it. Second, the diverse perspectives create many possible proposed solutions. Of the twenty-three car models, thirteen are local optima for at least one of the three problem solvers. We

TABLE 6.1:
The group's locally optimal car models

Perspective	*Model Rankings*								
Weight	**Malibu**	**Camry**	**Pontiac**	**Volvo**	Century	Jetta	Gallant	Saab 9-5	
Wheelbase	**Malibu**	**Camry**	**Pontiac**	**Volvo**	Century	Ultima	BMW 525		
Height	**Malibu**	**Camry**	**Pontiac**	**Volvo**	Jetta	Suzuki	Stratus	Sonata	Madza6

might then expect the diverse group to generate lots of solutions but not expect all of those solutions to be good. Some might be pretty lousy. That's fine as long as the group can identify the best solution among those proposed. And in this case, they can.

Stringing Us Along

We've taken for granted that these diverse perspectives exist. We might ask why: Why wouldn't people all conform to a common perspective? One obvious reason for diverse perspectives is that each may apply to different parts of a problem. Physicists who study string theory rely on multiple perspectives because each perspective has a set of subproblems for which it is useful. Some people may be surprised to learn that for a long time, physicists didn't know that they were using multiple perspectives. We won't dig deeply into string theory, as it's a knotty subject. We need only these few facts: by the early 1990s, physicists had developed five distinct string theory models—creatively named the Type I, Type IIA, Type IIB, Heterotic-O, and Heterotic-E models. Each of these theories describes our universe as containing six extra dimensions—beyond space and time. These extra dimensions are all folded up so we don't notice them. Yes, string theorists believe that hidden inside that waffle you had for breakfast were six dimensions. Incredible if true, but irrelevant to our larger point. In the mid-1990s, Edward Witten proved that these theories were all the same theory, that they were distinct mathematical representations of the same thing—not competing, but complementary. In our language, they were five distinct perspectives of the same strings. Witten's remarkable discovery led string theorists to embrace all five models. Five perspectives on a set of problems create five landscapes. To quote Brian Greene's *Fabric of the Cosmos*,

> Theorists have found that for certain questions one of the five may give a transparent description of the physical implications, while the descriptions given by the other four are too mathematically complex to be useful. And therein lies the power of Witten's discovery. Prior to his breakthrough, string theory researchers who encountered intractably difficult equations would be stuck. But Witten's work showed that each such question admits four mathematical translations—four mathematical reformulations—and sometimes one of the reformulated questions proves far simpler to answer. Thus, *the dictionary for translating between the five theories can sometimes provide a means for translating impossibly difficult questions into comparatively simple ones.*[8]

Thus, string theory provides yet another example of how diverse perspectives improve problem solving: Different perspectives create different landscapes, and different landscapes have different peaks. Different peaks prevent people from getting stuck at the same point.

Diverse Heuristics and Problem Solving

We now consider how diverse heuristics can aid problem solving by applying some to a trading problem involving men and shoes. We assume that each of five men, Richard, Chauncey, Ben, Rasheed, and Tayshaun, wears size twelve shoes. Each grabs two random shoes from a collection of five pairs: loafers, tennis shoes, sandals, boots, and wingtips. These men don't care which shoes they have, so long as they have matched pairs. We can identify the shoes by the type and the foot so that loafer-L denotes the left loafer, wingtip-R denotes the right wingtip, and so on. The initial random allocation of shoes looks as shown in table 6.2.

TABLE 6.2:
The shoe problem

Person	*Shoe 1*	*Shoe 2*
Richard	loafer-L	boot-L
Chauncey	sandal-R	wingtip-R
Ben	tennis-L	sandal-L
Rasheed	loafer-R	boot-R
Tayshaun	wingtip-L	tennis-R

Our first heuristic allows trades between two players only if both players want to make the trade. This heuristic creates a trade between Richard and Rasheed. Richard has the left loafer and the left boot. Rasheed has the right loafer and the right boot. They don't care whether they trade loafers or boots. They just want pairs. Either trade gives each of them a pair. A trade between Richard and Rasheed is the only trade that this first heuristic can find. Any other trade benefits only one of the two people trading.

Our second heuristic allows for trades among exactly three people. As with the first heuristic, a trade takes place only if everyone involved in the trade is strictly happier. Using this heuristic, Chauncey can give the right sandal to Ben, Ben can give the left tennis shoe to Tayshaun, and Tayshaun can give the left wingtip to Chauncey. After this trade, all three have matched pairs. This heuristic would not have found the trade between Richard and Rasheed because that trade involved only two people. Thus, on their own, neither of these two heuristics locates a global optimum, yet together they do.

This example reiterates a point that we discussed earlier: local optima are defined relative to a perspective and the set of heuristics applied to that perspective. If a person has lots of heuristics, she probably is a good problem solver. Think back to the traveling salesperson problem. Some solutions were local optima for one heuristic but not for others. That example hinted at a general insight: *the more tools in our kits, the fewer places we get stuck*.

Interpreting Problem Solvers by Their Peaks

We've now hit upon an important idea: characterizing problem solvers by their local optima, the peaks in their landscapes. A problem solver typically has many local optima—and we know he always has one: namely, the best solution. Everyone agrees that Everest is a peak. A problem solver might not find Everest, but if put there he recognizes it as a peak. I next state three observations that connect characteristics of problem solvers to their local optima. The first describes a correlation between better problem solvers and the values of their local optima.

Better Individual Problem Solvers Have Better Local Optima: *Those problem solvers who perform better individually tend to get stuck at local optima that have relatively good values.*

This observation speaks to the advantage of having relatively high value local peaks. Again, think back to our traveling salesperson problem. If one problem solver has two local optima: one of length 1,400 miles (which is the global optimum) and one of length 1,500 miles, this person will probably do better than someone else with two local optima, one of which has length 1,800 miles (keep in mind that longer is worse). Note that we're assuming that the problem solvers are equally likely to find each of their local optima, an assumption we relax in a moment.

The second observation is that better problem solvers tend to have fewer local optima. They could use perspectives that create less rugged landscapes, which by definition have fewer local optima. Or they could possess more heuristics allowing them to move off from peaks in their landscapes. Either way, they get stuck at fewer points.

Better Problem Solvers Have Fewer Local Optima: *Those problem solvers with better individual performance tend to have fewer local optima.*

The intuition that explains this claim is also straightforward. Local optima are the points at which a problem solver gets stuck.

One of these local optima must be the best solution. The others have values that are not as good. Getting stuck at those other solutions results in worse outcomes. In general, the more local optima, the more likely the search gets stuck on one of them, and the worse the problem solver performs.

We can combine these first two observations. High-performing individual problem solvers should have relatively few local optima, and those local optima should have high values. In contrast, we should expect the poorest performing individual problem solvers to have lots of local optima and many—but not all—of those local optima to have low values.

As convincing as this sounds, the characterization of a problem solver by the number of local optima she has and the value of those local optima gets us only partway to where we need to go. The probability that a problem solver lands on each of those local optima also matters. A problem solver could have only two local optima—the global optimum and a local optimum of relatively low value—but if she almost always locates the low-value local optimum, her average performance will be poor. If she almost always locates the global optimum, her average performance will be great. To capture this insight formally, I introduce the concept of a *basin of attraction.* Loosely speaking, the size of a local optimum's basin of attraction equals the probability that the problem solver gets stuck at that peak.

The word *basin* comes from physics, but we can think of basins in our kitchens or bathrooms and get the idea. Imagine tossing a superball into a room filled with sinks of various sizes and shapes. The ball would ricochet from sink to sink until coming to rest at the bottom of one of the sink's basins. All else being equal, we might expect that the larger and deeper the sink's basin, the more likely the ball lands in that sink. We can apply this same intuition to hill climbing, but we have to invert the imagery. Physicists minimize. Our problem solvers maximize. So, for our problem solvers, the analogs of the depth and size of the basin are the height and width of a peak (see figure 6.5).

Even though we think of people and groups as trying to optimize, let's follow the convention and apply the basin terminology.

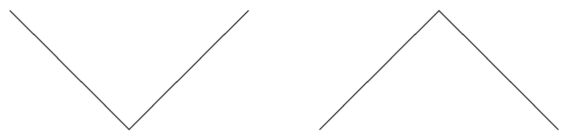

Figure 6.5 Basins and Peaks

Let's consider an example to get our bearings. Suppose that two South American banana growers, Karen and Paul, are attempting to breed bananas with a longer shelf life. We will assume that the best possible banana has a shelf life of thirty days after being picked. Karen uses old-fashioned genetic breeding techniques. These techniques are her heuristics, and we assume that they result in three local optima: the best solution, a second solution with a shelf life of twenty-four days, and a third solution with a shelf life of twelve days. Paul has abandoned old-fashioned breeding and relies on genetic modification. His heuristic also has three local optima. They produce bananas with shelf lives of thirty, twenty-five, and twenty days, respectively. If both problem solvers are equally likely to find each of their local optima, then Paul performs better than Karen as the calculations below show.

Shelf Lives of Karen and Paul's Bananas

$$\text{Karen's Solution's Expected Value: } 22 = \frac{1}{3}(30 + 24 + 12)$$

$$\text{Paul's Solution's Expected Value: } 25 = \frac{1}{3}(30 + 25 + 20)$$

To see the importance of basin size, suppose that Karen has a larger basin of attraction for the best solution—that she finds the best solution two-thirds of the time and that she finds the other solutions only one-sixth of the time each. Her bananas then have a longer expected shelf life. Suppose also that Paul remains equally likely to find each of his three solutions, so that his expected value remains the same. Given these assumptions, Karen performs better on average as the following calculations show.

New Shelf Lives of Karen and Paul's Bananas

$$\text{Karen's Solution's Expected Value: } 26 = \frac{2}{3}(30) + \frac{1}{6}(24 + 12)$$

$$\text{Paul's Solution's Expected Value: } 25 = \frac{1}{3}(30 + 25 + 20)$$

We can now make a third observation:

Basin Size Matters: *Better problem solvers tend to have larger basins of attraction for their better local optima.*

Based on what we've done so far, we can characterize a problem solver as a set of local optima together with a probability of getting stuck at each of those local optima. In our example, we can identify Karen as having three local optima given by the set $\{30, 24, 12\}$ with probabilities $(\frac{2}{3}, \frac{1}{6}, \frac{1}{6})$, and we can characterize Paul as also having three local optima. But his have values $\{30, 25, 20\}$, and have probabilities $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Note, of course, that these local optima both contain the global optimum. Everyone has to get stuck at the best solution.

This characterization of problem solvers as sets of local optima and the probability that they locate them proves useful, but it's not a perspective on the set of problem solvers. We next show that two problem solvers can have distinct perspectives and heuristics, yet produce identical sets of local optima and probabilities of attaining them. Thus, the mapping of a problem solver to a set of local optima with probabilities should be thought of as an *interpretation* in the formal sense of that word. It puts problem solvers into categories. This distinction may seem like hairsplitting, but it's not. It allows us to make an important distinction between internal and external problem solving diversity.

Internal and External Problem Solving Diversity

Before turning to that distinction, let's step back for a moment and take stock of what we've learned. A person's

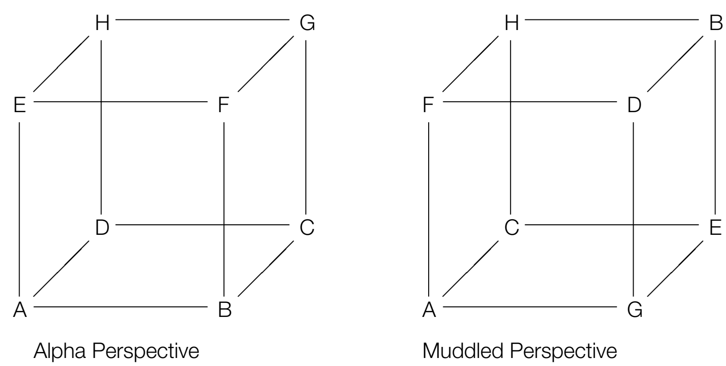

Figure 6.6 Two Perspectives on a Cube

perspective and heuristic define the neighbors of an existing solution—the solutions that a problem solver thinks to search (namely, the solutions, given her perspective, that her heuristics identify). Perspectives and heuristics are *internal* to a problem solver. An outside observer need not know them. What an observer can see, what is *external* to the problem solver, is how the problem solver maneuvers through the solutions. Even knowing this would require some effort on the part of an observer, but we ignore that for the moment.

Our next example requires slow and careful study. It's not difficult, but following it requires patience. Put yourself in the frame of mind you'd be in if looking at a map of a city's public transportation system: it looks like a mess, but there's an underlying logic. In the example, two problem solvers differ both in their perspectives and in their heuristics, and yet they are the same externally—they maneuver in the space of solutions in the exact same way. Let's start with the perspectives. Each of two perspectives shown in figure 6.6 organizes the eight possible solutions, lettered from A to H, on a cube in a different way. The *alpha perspective* organizes the solutions alphabetically moving counterclockwise around each level. The *muddled perspective* randomly arranges the letters on the cube.

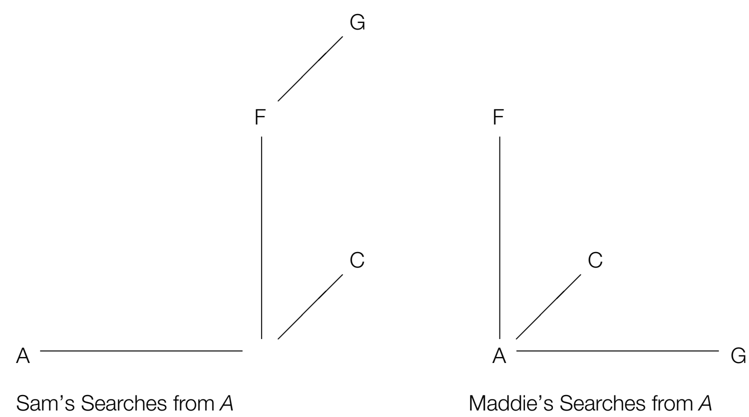

Figure 6.7 Internal Diversity and External Equivalence

We consider two problem solvers, Sam and Maddie. Sam uses the alpha perspective, and Maddie uses the muddled perspective. Sam and Maddie use different heuristics. Each has three. To define their heuristics, we rely on the edges of the cube. Maddie's three heuristics involve movements along these edges. Starting at solution *A* on the muddled perspective, Maddie checks the three solutions connected along edges of the cube. By inspection these are *G*, *C*, and *F*; *G* lies to the right of *A*, *C* lies behind it, and *F* lies above it.

Sam, in contrast to Maddie, uses sophisticated heuristics. Her first heuristic moves along both the left–right dimension and the front–back dimension. Starting from *A*, Sam checks *C*. Her second heuristic moves along both the left–right dimension and the top–bottom dimension. Again starting from *A*, Sam checks *F*. Finally, her third heuristic moves along all three dimensions. This heuristic moves to the far corner of the cube. It's the cubist version of "do the opposite." Using this heuristic, she jumps from *A* all the way to *G*.

A picture (figure 6.7) helps to make sense of Sam and Maddie's searches over their perspectives.

As this picture makes clear, Sam and Maddie check the same solutions from *A*. With a little effort, it can be shown that this equivalence holds regardless of where they start. Using her heuristic, if Maddie starts at solution *D*, then she checks *F*, *B*, and *G*. The same is true of Sam. Starting at *D*, she goes over and across to *B*, over and up to *G*, and jumps to the opposite corner, which is *F*.

Working through this example required some effort. People who like math probably found it cool, but should anyone else (any nongeeks) care?[9] Yes. They (you) should. This example shows that internal diversity (diverse perspectives and diverse heuristics) need not create external diversity (movements in the space of solutions). People can be different and yet search for solutions the same way. People with diverse perspectives *and* diverse heuristics can solve a problem similarly.[10] Put differently, diverse perspectives and diverse heuristics can cancel each other out. This cancellation may not be likely, but we have no guarantee that two people who use different representations and apply different problem-solving techniques search for solutions differently. The implication: *diverse people may not solve problems differently.*

Diversity Trumps Ability

With this background in place, I now turn to the claim that diversity can trump ability. I first show how diversity trumps homogeneity. Imagine two collections of people, one diverse and one homogenous. If the "ability" (I formalize this later) is the same for all of these problem solvers, then the diverse group finds better solutions. Though less surprising than the diversity-trumps-ability result, it helps us intuit for why diversity is beneficial.

To get started, we need a model of how a collection of people, diverse or not, solves a problem. We'll use the same process we described in the cup of coffee model: we order the members of the group and have them apply their problem-solving skills sequentially. One person searches until she gets stuck at a local optimum, then the next person searches beginning from that point. In this way, each person builds off the best solution found by the previous problem solvers. We have people literally standing on the shoulders (the solutions) of those who came before them. Only when no problem solver can find an improvement does the process stop.

Don't fret, the sequential search assumption is a convenience; it is not necessary for any of our results. We could also model all of the problem solvers working on the problem simultaneously and having any problem solver who locates a better solution post it to a common message board. The other problem solvers could then immediately begin searching for new solutions from this new, better solution.

Diversity Trumps Homogeneity

To see how diversity trumps homogeneity, we'll construct a model with two collections of problem solvers. Everyone in the first collection is unique. Each has a distinct perspective and set of heuristics. In light of the example that we just constructed with Sam and Maddie, we will assume that this internal diversity results in different approaches to solving the problem. Everyone in the second collection is identical. Every person uses the same perspective and the same heuristics. We further assume that all of the problem solvers in both collections have approximately the same individual ability, by which we mean that working alone on the problem, each does equally well.

If these assumptions hold, then the diverse collection generally outperforms the homogenous collection. The intuition behind this result should be clear: *the homogeneous collection may just as well contain only a single person*.[11] Every person in the homogenous collection possesses the same perspective and heuristics, so they all have the same set of local optima. Therefore, after the first person in the collection locates an optimum, no one else can improve on it. They would all look at the solution and say something like "looks good to me" and be done with it. Two heads are not only not better than one in this case—they *are* one.

Next consider the diverse collection of problem solvers. The first problem solver applies his perspective and heuristic until finding a local optimum. The next problem solver then attempts to find an even better solution. As the second problem solver relies on different perspectives and heuristics, she might find a better solution. If, by chance, the first problem solver locates the best solution, the global optimum, then of course it must be a local optimum for the second problem solver as well. But if that occurs, the diverse collection has solved the problem perfectly. If the first problem solver fails to locate the global optimum, then the other one might improve on that solution. That possibility of improvement explains why diversity trumps homogeneity.

An example clarifies this logic. This example relies on only two problem solvers in each group. The problem is to figure out a way to get as many objects as possible into a box. If it helps, think of a shoe, a soda can, a book, a toy car, a desk telephone, and so on . . . that must be put in a box. Imagine two homogenous problem solvers, each of whom relies on the Steven Covey "deal with the big items first" heuristic. This heuristic might result in three local optima to this problem. One of these must be the global optimum in which ten objects fit into the box. Let's suppose that in the other two local optima, eight and nine objects fit into the box, respectively.

Each of these three local optima is equally likely, and which of these local optima is found depends on which objects are put in first. Assuming equal sized basins of attraction, the expected value of a local optimum (in this case, the expected number of objects) for each problem solver equals one-third times ten plus one-third times nine plus one-third times eight, for an average of nine. The expected value for the two of them working together sequentially is also nine. Once the first person proposes a solution, the second person will not see a way to improve on it—because she sees the problem in the same way.

In the diverse collection, the first problem solver, Blair, uses a perspective and heuristics that almost always lead her to fit nine objects in the box. The ten-object solution is a local optimum for her as well—it has to be—but she almost never finds it. Her expected value is approximately nine, the same as that of the problem solvers in the first collection. The second problem solver, Karl, uses a different perspective and heuristics. He gets stuck at two local optima. The first is the best solution. In the second, he can place only eight objects in the box. Let's assume that these two local optima are equally likely, so that his expected value is also nine. Here's the key insight: *if Blair and Karl work together on the problem, they always find the optimal solution*. Their expected value equals ten.

To see this, suppose that Blair goes first. She either finds the global optimum (which is unlikely) or finds her local optimum that puts nine objects in the box. Let's assume the latter; otherwise, they've found the global optimum and the proof is complete. By construction, her nine-object solution is not a local optimum for Karl. He gets stuck only at solutions with eight objects or ten objects. If he is not at a local optimum, he can find an improvement starting from that solution. Thus, if he starts at a solution with nine objects, he must be able to find the global optimum, since his only other local optimum has fewer than nine objects.

Suppose instead that Karl goes first. Here the logic is the same but it requires one more step. If Karl finds the solution with ten objects, then they've done it. They have found the optimal solution. So let's suppose that he finds the solution with only eight objects (see figure 6.8). Blair can then improve on this solution. Why? No solution with eight objects is a local optimum for her, so she must find either her solution with nine objects or find the global optimum. If she finds the former, then following the logic above, Karl can then find the ten-object solution. Thus, Karl and Blair always find the solution with ten objects.

The characterization of problem solvers by their local optima clarifies the intuition behind the finding that diversity trumps homogeneity. That clarification rests on the *Intersection Property*, which says that the only way for two people to be stuck on a local optimum is if both are stuck. Yes, that sounds circular, but it's deeper than it seems. It means that the only local optima for

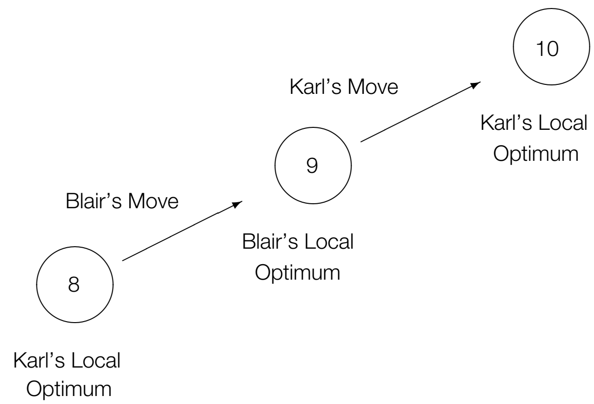

Figure 6.8 Blair and Karl Finding the Global Optimum

a collection of people are points that are local optima for *every* person in the collection. We previously saw this in the example of finding the car with the highest gas mileage.

The Intersection Property: *The local optimum for a collection of problem solvers equals the intersection of the individuals' local optima.*

We can use this property to recast our example. The analysis becomes a little more technical, but let's keep in mind that our goal is to move beyond metaphor to a logical understanding. Each of the two homogenous problem solvers has three local optima called X, Y, and B (for best). These have values eight, nine, and ten, respectively. The intersection of those sets of local optima remains the set X, Y, and B. Two heads are no better than one. In contrast, Blair has only two local optima, Z and B. Karl also has two local optima, W and B. The intersection of these two sets of local optima $\{W, B\}$ and $\{Z, B\}$ equals the single solution B. Regardless of how they interact, Karl and Blair must locate B. It is the only solution that is a local optimum for both problem solvers. Using the logic of the intersection property, we can see why diversity trumps homogeneity.

The Diversity Trumps Homogeneity Theorem: *If two collections of problem solvers contain problem solvers of equal individual ability, and if those problem solvers in the first collection are homogeneous and those in the second collection are diverse, that is, they have some differences in their local optima, then the collection of diverse problem solvers, on average, outperforms the collection of homogenous problem solvers.*

The proof of this claim is straightforward. The homogenous problem solvers all have the same local optima. Thus, the expected value of a solution located by the collection of agents cannot be better than that of any individual. This result does not hold for the diverse collection of problem solvers. Their local optima differ. In this way, just as in the example earlier, one problem solver gets stuck at a local optimum, and then another problem solver finds an improvement. The intersection property gives us a logic underpinning for the notion of standing on the shoulders of giants. What is a local optimum for one person's perspective and heuristics need not be for another's. As Emerson puts it,

> Every ultimate fact is only the first of a new series. Every general law only a particular fact of some more general law presently to disclose itself. There is no outside, no inclosing wall, no circumference to us. The man finishes his story, —how good! How final! How it puts a new face on all things! He fills the sky. Lo! On the other side rises also a man and draws a circle around the circle we had just pronounced the outline of the sphere.[12]

Emerson didn't believe that we ever got to the global optimum, whereas we allow for that possibility, but we can put such small quibbles aside. Emerson's prose and our logic agree: diversity creates iterative improvements. And these improvements continue until the collection finds a solution that lies in the intersection of the local optima for all of the problem solvers.

To some, the claim that diversity trumps homogeneity may not be surprising.[13] Fair enough, but we learn from dotting our *i*'s and crossing our *t*'s. Many intuitive insights turn out not to be true—Newtonian physics being just one example. So although this claim is not difficult to understand, it is important. It's better for two people to be diverse than for them to be the same when solving problems.

Diversity Trumps Ability

We now turn to the more surprising claim, namely that diversity trumps ability. This claim differs from the previous one in an important respect. This claim rests on different assumptions. The diversity-trumps-homogeneity claim assumes that all of the problem solvers have equal ability as measured by their expected value on the problem. The diversity-trumps-ability claim assumes lower average ability for the collection of diverse problem solvers. It also allows for variation among the collection of the best problem solvers. We're not assuming that every person in the group of the best is identical. We're only assuming that they're good.

We begin with an initial pool of potential problems solvers from which we draw the collections of problem solvers. We assume there are N of these people. N could equal everyone who works for a firm or every faculty member at a university. We then compare the collective performance of the M best problem solvers against the collective performance of M randomly selected problem solvers. The theorem provides sufficient conditions under which the random collection outperforms the collection of the best.

To demonstrate the logic of this result, we'll characterize the problem solvers by the local optima they find and the probabilities that they attain them. This approach has advantages over looking at their perspectives and heuristics directly. The best problem solvers get stuck at fewer local optima, and these local optima tend have higher values. Less effective individual problem solvers have more local optima, and many of these may have low values.

Before we go too far, let's keep in mind that the diversity-trumps-ability result doesn't always hold. It holds given certain conditions. We take up these conditions one at a time. Then we look at the formal claim, and then go back over the conditions in a slightly different way to hammer home the intuition. What follows can't be skimmed and fully understood.

The first condition takes into account that if the problem is so easy that a problem solver (or several problem solvers) can always finds the best solution, then the collection of the best problem solvers (which necessarily contains this best problem solver) always locates the best solution. In contrast, the collection of randomly selected problem solvers need not contain anyone who always finds the optimal solution. So for diversity to trump ability, the problem must be difficult. For example, if we need to find the answer to a calculus problem, we can often just ask an expert. The expert can give us the correct answer. A random group might not. However, if we have a difficult, previously unsolved math problem, we would want to ask a diverse collection of mathamaticians. Diversity benefits to kick in when the problems we face are hard improving our educational system—designing products, curing diseases. We need not think of this as a limiting assumption.

Condition 1: The Problem Is Difficult. *No individual problem solver always locates the global optimum.*[14]

This condition requires some thought. We have assumed problems have an associated difficulty. This assumption seems to contradict our earlier discussion of perspectives in which we discussed how difficulty lies in the eye of the beholder. One person's rugged landscape is another's Mt Fuji. So, we need to be a bit more precise. What we're assuming is that the problem is difficult given any of the problem solvers' perspectives.

This second condition concerns the ability of the problem solvers. All of the possible problem solvers must have some ability to solve the problem. Their perspectives cannot create overly rugged landscapes. We cannot set loose a bunch of people from the humanities in the chemistry lab. We'll call this the *Calculus Condition* because people who know calculus can take derivatives. Derivatives tell the slope of a function. The slope of a mountain is either positive (uphill), negative (downhill), or zero (on a peak or a plateau). On a peak the derivative equals zero; the slope goes neither up nor down. So, people who know calculus can find peaks. People who do not know calculus could get stuck anywhere. They would not add much to a group trying to solve a problem. People who know calculus relative to the problem have perspectives that capture some of the structure of a problem. They have a reasonable number of peaks.[15] This condition trivially holds when the number of solutions is finite, such as in the traveling salesperson problem.

Condition 2: The Calculus Condition. *The local optima of every problem solver can be written down in a list. In other words, all problem solvers are smart.*

To see the importance of the Calculus Condition, suppose that it fails to hold. If a great many of the problem solvers had an infinite number of local optima, then the random collection could be just many monkeys on typewriters trying to peck out a little Shakespeare. They could not possibly do as well as a collection of the best problem solvers. To bring this more down to earth, a collection of random people would not outperform a collection of top statisticians on a statistical problem. Relative to the problem, most people would not satisfy the calculus assumption. They would get stuck on almost any solution.

The third condition requires that for any solution other than the global optimum, some problem solver can find an improvement on that solution. In other words, the intersection of all of the problem solvers' local optima contains only the global optimum. We'll call this the *Diversity Condition*, as it assumes diversity among the problem solvers.

Condition 3: The Diversity Condition. *Any solution other than the global optimum is not a local optimum for some nonzero percentage of problem solvers.*

This condition does *not* say that given any solution, some problem solver exists who immediately can jump to the global optimum. That assumption would be much stronger and would rarely be the case. Our assumption says, instead, that some problem solver exists who can find an improvement. That improvement may be small. In the example of the group of people playing Kasparov, this assumption may not have held. The collection of people may not have included someone to prevent a bad move. More than likely, though, Kasparov's advantage resulted from the collection's inability to evaluate moves properly, as they lacked experience in such high-level games—an observation we make more explicit in the next two chapters.

To see this in a more formal context, think back to the traveling salesperson problem. Imagine a collection of problem solvers each of whom has some collection of heuristics. Some of these heuristics might involve switching the order of four or five cities in the route. The Diversity Condition implies that given any nonoptimal route, there exists a problem solver who has a heuristic who can find an improvement in that route. This person doesn't have to be capable of finding the best route, just an improvement. That improvement could reduce the route by only a few miles or even just a few feet.

The final condition requires that the initial set of problem solvers must be reasonably large and that the set of problem solvers that form the collection must also not be too small. The logic behind this condition becomes clear when we consider extreme cases. If the initial set consists of only fifteen problem solvers, then the best ten should outperform a random ten. With so few problem solvers, the best ten cannot help but be diverse, that is, have different local optima. And as they individually perform best, they should do better collectively than the random ten problem solvers. At the same time, the collections that work together must be large enough that the random collection can be diverse. To see this, suppose that the collections contain only one problem solver. By definition, the collection with the best problem solver outperforms the collection with a random problem solver. Even with two problem solvers in the collections of problem solver, the collection of the best problem solvers would almost certainly do better. So we need to be picking from a big pool—firms, organizations, and universities do this, by the way—and we need to be picking collections of more than two or three people.

Condition 4: Good-Sized Collections Drawn from Lots of Potential Problem Solvers. *The initial population of problem solvers must be large and the collections of problem solvers working together must contain more than a handful of problem solvers.*

There is no explicit size that these collections have to be for the result to hold. The exact number depends on the difficulty of the problem and on the amount of diversity in the initial set of problem solvers. More difficult problems have more local optima and require more problem solvers to overcome the problem of overlapping local optima. The more diversity in the problem solvers, the smaller the collections can be. Diverse problem solvers get stuck at different local optima, so when working together on a problem they rarely get stuck at the same local optima.

These four conditions—the problem has to be hard, the people have to be smart, the people have to be diverse, and the group size has to be bigger than a handful and chosen from a large population—prove sufficient for diversity to trump ability. They're not the only conditions under which the result holds, but if they're satisfied, diversity trumps ability.

The Diversity Trumps Ability Theorem: *Given Conditions 1–4, a randomly selected collection of problem solvers outperforms a collection of the best individual problem solvers.*

This theorem is no mere metaphor or cute empirical anecdote that may or may not be true ten years from now. It's a logical truth.[16]

To see why it's true, let's go back to our beginning, a good place to start. Problem solvers have perspectives and heuristics. To make the logic as transparent as possible, we will assume that each person has a single perspective/heuristic pair. We can represent the space of perspectives and heuristics as being a box (figure 6.9). And as we have done with the problems themselves, we can create a landscape by setting the height of a perspective/heuristic pair equal to its average value when applied to our difficult problem. This landscape, the landscape of problem solvers, must have

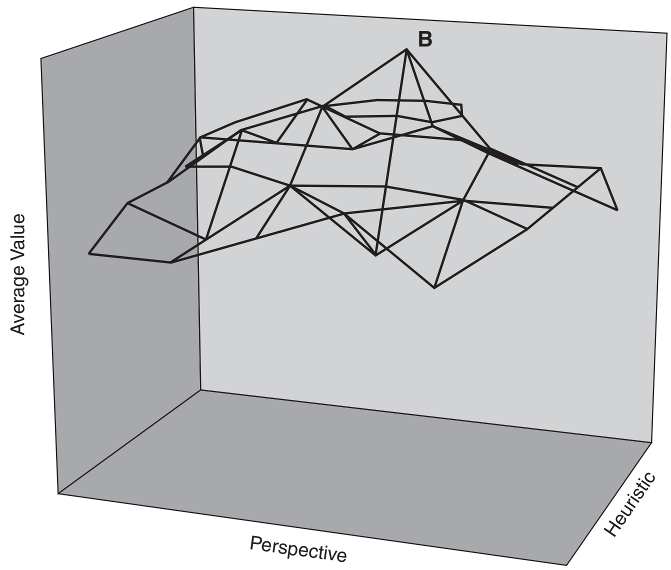

Figure 6.9 The Perspective/Heuristic Box

a global optimum.[17] That global optimum represents the best individual problem solver. We denote her by *B*. Note, though, that by Condition 1 (The Problem Is Difficult), no single perspective/heuristic pair locates the global optimum every time.

Now let's generate lots and lots of problem solvers. We can represent these by dots in the figure. If we create enough problem solvers, most of the best ten or twenty will all lie near α, the global optimum on the problem solver landscape. Why? If we randomly tossed out billions of people in the Himalayas, most of the people at highest elevations would be on Everest (Everest in the space of problem solvers). However, even the best problem solver—the one atop Everest—cannot solve our problem alone. Why? On hard problems, even the globally optimal problem solve can't find the global optimum to the problem. Conversely, if we create only a few dozen problem solvers, the best ten or twenty are scattered throughout the perspective/heuristic box. Herein lies the key insight. When the initial set of problem solvers grows large, clustering of the best problem solvers near α becomes unavoidable, which is why we need Condition 4 (lots of initial problem solvers to choose from).

Because they are clustered near B, or at one or two other places, these best problem solvers all have similar perspectives and heuristics. And it follows that they do not perform much better as a collection than they do individually. In contrast, the random collection's members will most often lie all over the perspective/heuristic box. This diversity allows them to collectively perform much better than they do individually. That alone is not sufficient for them to outperform the best. They must be smart. Hence, the need for Condition 2 (the Calculus Condition). Finally, these diverse problem solvers cannot all get stuck at the same bunch of solutions with low values. Thus, we need Condition 3 (the Diversity Condition). And so we have it: Diversity trumps ability. Not always, but when the conditions hold.

How do we apply this in the real world? Simple. When picking two hundred employees from a pool of thousands, provided the people are all smart, we should keep the theorem in mind and not necessarily rank people by some crude ability score and pick the best. We should seek out difference. When picking four students from a class of sixty to work as a team to compete in a science competition, keep the theorem in mind as well, but pick the best four—unless they're demonstrably similar in skills.

Does the claim imply that a group of only chemists might benefit from having a poet join them? No. But perhaps they'd benefit from adding a physicist. Does it imply that a team of people with high test scores all trained at the same school in the same techniques might not do as well as a group that contains diversely trained people with diverse experiences and slightly lower average SAT, GRE, or IQ scores? Probably. If you work at a big firm filled with Harvard, Wharton, and Michigan Business School MBAs, you might question whether your firm shouldn't seek a little more diversity in educational background.

We should recall that Conditions 1 through 4 are sufficient but far from necessary for diversity to trump ability—just as ice cream is sufficient but not necessary to quiet my two boys. There are other conditions under which the results hold. The strongest of the conditions, the Diversity Condition, can often be relaxed. In fact, in the agent-based models that stimulated the formal analysis, it was almost never satisfied. So long as the solutions that lie in the intersection of the diverse agents' local optima do not have low values, the diverse groups perform better for the simple reason that they have fewer local optima.

Darwin's Brass Tacks

As I mentioned earlier, the veracity of the diversity trumps ability claim is not a matter of dispute. It's true, just as $1 + 1 = 2$ is true. However, the claim applies to mathematical objects and not to people directly. It is a claim about how diverse perspectives and heuristics aggregate. Returning once again to the traveling salesperson problem helps us understand the implications of this distinction. We consider two thought experiments. In the first, we gather a group of one thousand undergraduates, explain the traveling salesperson problem to them and ask each to create a problem-solving algorithm. An algorithm consists of an encoding of the problem (a perspective) and ways to manipulate solutions (heuristics), so this fits our model perfectly. We rank those algorithms from best to worst based on how they performed on a sample eighty-city problem. We might well find substantial diversity among the thousand algorithms and also find that the best algorithms were similar. If so, a random collection of twenty algorithms would probably outperform the collection of the best twenty algorithms.

In the second thought experiment, we take the same one thousand students and we give them a general intelligence test. We then take the best twenty and put them in a group, and we take a random twenty and put them in a group. Suppose we then ask these two groups to come up with a solution to our eighty-city traveling salesperson problem. Would we still expect the random group to do better? The answer is less clear. Scores on general intelligence tests correlate with factors such as ambition, effort, and concentration, suggesting that the high-IQ group would do better. Even more relevant to our discussion, people with similar IQs might go about solving the traveling salesperson problem differently. So the high-IQ group could well be diverse.[18]

These two thought experiments reveal subtleties in how we think about the perspectives and heuristics. In the first, we need not care whether people came with those perspectives and heuristics already in their heads or if they developed them on the spot. In the second, this distinction matters. If people create their heuristics on the fly, then group dynamics matter even more. For example, the high-IQ group might recognize that they need to have diverse ways of solving the problem. If so, they probably would do better than the random group.[19]

Returning to the toolbox model helps us to make sense of how these two scenarios differ. People who score high on general intelligence tests probably possess more tools. Either that or the people who don't score well just know lots of stuff that test makers think is unimportant, such as how to speak backward in Pig Latin. (You say, "ooh;" they say, "yahoo.") If we compare a collection of people with high general intelligence to a collection of people with random general intelligence on some problem, we should find that each member of the first group has more tools. We might also think that the number of unique tools in their combined collections of tools exceeds the number for the other group and that therefore the high-IQ group would do better.

We have to be careful not to jump to conclusions. Let's dump out the toolboxes and see what we get. Suppose that each high-IQ person has on average thirty tools and the members of the random collection have on average twenty-five tools. With twenty members per collection, the high-IQ group has six hundred tools (twenty people times thirty tools each), whereas the other group has on average about five hundred tools. The point is that not all of these tools are unique. The relevant comparison is between the numbers of *unique* tools in each set. Now, we might think that the set with more tools should have more unique tools as well. *But that thinking is flawed.* By construction, the tools of the people in the high-IQ group are not random. They are tools that allow a person to do well on an IQ test. And IQ tests choose questions that people with high IQs do well on; so do SAT tests and ACT tests. This rule for choosing questions makes the tools of high scorers more similar than they would be if questions were chosen randomly. Nothing in the construction of the random group makes their tools similar.

This effect becomes even more pronounced the more accurate our test. If we were to select the "best" group according to how they do on the traveling salesperson problem, its members would be even more likely to have similar perspectives and heuristics. Why? Because the heuristics that enable someone to do well on a traveling salesperson problem are things like flip pairs of cities or flip pairs of cities separated by a city. They're specific heuristics. A group of the best might well all be thinking about the problem in one or two ways. Their overlap in tools would work to their disadvantage.[20]

This phenomenon of diversity reduction created by picking the best on either the IQ test or the traveling salesperson problem should sound familiar. We can find it in Darwin: *selection reduces diversity.* Even though we might think that we are selecting on something called ability, ability depends on tools. So, ultimately, selection takes place over the tools. By selecting those who perform better, we get people with more tools. That helps collective performance. But we also get people who are less diverse, and that hurts collective performance. That's why groups of the best need not be the best groups.

The Contextual Problem of Valuing Diversity

The diversity-trumps-ability result also reveals the contextual nature of an individual's contribution. How much someone improves the current solution depends on her individual talents, to be sure, but her contribution also depends on her diversity relative to the other people working on the problem. She may be standing on the shoulders of the others and peering in a slightly different direction and therefore be able to make a small local