build anything interesting. With both, we can build homes, tree forts, and fences. As the management guru Peter Drucker wrote, "Effective work is actually done in and by teams of people of diverse knowledge and skills."[1]

In this chapter, we put the toolbox framework through its paces, contrasting it with measuring stick approaches such as IQ scores and multidimensional intelligence scores. Even the multidimensional measures understate the amount of diversity by projecting cognitive differences onto multidimensional spaces. In the toolbox framework, people differ in their capacities and these capacities translate into crude rankings at best. We see why we might be able to rank mathematicians but why we cannot say whether Tolstoy was smarter than Newton. We can rank within domains, but not across them. And we can rank within only some domains—ranking physicists is easier than ranking authors. We cannot say whether Joyce was better than Austen.

A Choice of Metaphors: Measuring Sticks or Toolboxes

Americans love to rank things: cities, schools, cars, airlines, breeds of dogs, and movie stars. We also like to think that we can rank people by their intelligence. But many of us believe that our cognitive abilities cannot be summarized in a single number or vector of numbers. Our analysis of toolboxes argues against complete rankings, but simply to say that everyone differs may be too extreme, too. At a minimum, we may be able to classify people in categories and to make comparisons across those categories. The great novelist or nuclear physicist is, in a real sense, smarter than the rest of us—but one may not be "smarter" than the other.

Different cognitive skills, like the varied attractions of a city, prove beneficial in different contexts. Cognitive skills, like physical skills, apply in some domains and not in others. Anyone who has trouble remembering that physical skills are context dependent need only pick up a Michael Jordan baseball card. Though Jordan was a great baseball player by ordinary standards, he was not up to major league level, even though he was certainly a better all around athlete than almost all major league players. Just as the physical qualities that make for a good figure skater are different from those that make for a good sumo wrestler, the cognitive skills required to be brilliant at one task often overlap little with the skills required for another.

To measure what's under the hood, so to speak, psychologists use general intelligence tests. These tests map the human mind onto a single dimension. That single dimension invites a conceptualization as ability. We think we are smarter than everyone who gets a lower score than we do and dumber than everyone who gets a higher score. We cannot help but do this. But in light of the complexity and diversity of our brains, this ranking of individual people or groups of people on a single measure seems a stretch. Any mapping from something as complex as intelligence into a single number condenses a lot of information. By way of comparison, think of giving cities a quality score. Can we reduce Paris or New York or even Tulsa to a single number? Of course not.

And still, IQ tests measure something meaningful. They capture someone's ability to exhibit a range of cognitive skills in a relatively short time span. People with high IQs should, on average, have more tools, especially tools that allow a person to answer questions quickly and accurately. Thus, the people who score highly may not be diverse.[2]

The toolbox framework fundamentally differs from measuring stick frameworks in how it captures intelligence. IQs and toolboxes are both interpretations. Each interprets a person as a set. In the case of IQs, the sets have numbers. In the case of toolboxes, the sets are combinations of tools. A person's IQ can take at most one of two hundred or so values. By contrast, the number of sets of unique toolboxes can be enormous, as we will see in a few pages.[3] Thus, the toolbox framework embraces our differences more than do measuring stick approaches.

IQ tests determine a person's score by the number of questions she correctly answers. Think back to the city analogy. To assign a single number to Chicago or Boston, we could ask the city questions. Do you have museums? Do you have parks? Is your air clean? Do you have a symphony? To get a high score, a city would have to give the correct answers. The city that scores highest depends on the questions asked. So we cannot say that Chicago is better than Boston or that Boston is better than Chicago. Depending on the questions we ask, either might do better. However, given almost any set of questions, Paris scores better (on average) than Tulsa or Grand Rapids. But, on some questions (are the locals friendly?) Tulsa and Grand Rapids give better answers than Paris. So we cannot say that Paris is better on every dimension.

To push this analogy with cities just a little bit further, suppose that Sarah, who lives in Chicago with her sister Kelly, is thinking of moving to Boston. Sarah knows nothing about Boston except that the magazine *Condé Nast* rates Chicago at 84 and Boston at 85 out of a possible 100. Should Sarah conclude that Boston is a better place to live? Of course not. She should instead conclude that each city has its advantages relative to the other and that on balance they're about the same. If we pull this analogy back to IQ differences, we're drawn to the conclusion that large differences in IQ may signify meaningful differences in intelligence but small differences in IQ do not. And focusing on differences obscures relevant cognitive differences. Two people with identical IQs can make vastly different contributions to society.

Nevertheless, once we have these single numbers, we start making comparisons. Those comparisons create stress and tension. Here's a scenario that gets played out in homes across America every year. David, all six feet two inches of him, receives his SAT scores and compares them to those of his older sister, Jackie, who stands five feet six. His scores are lower. His parents try to calm him by saying, "We're all different. No one is smarter than anyone else." He's thinking, "Yeah, right, I believe that." He can see the scores. His parents may as well have told him that he and his sister are the same height. Yet if we accept the toolbox metaphor, we recognize that David's parents make sense. He's taking these scores too seriously.[4] We can line people up along the wall and make little pencil marks to determine relative heights, but we cannot do the same for intelligence.

Information Loss and Multidimensional Measures

If one number won't do, perhaps we can get closer to capturing intelligence by increasing the number of dimensions. As we will see, this idea moves in the direction of the toolbox approach but doesn't get us all the way there. Nevertheless, it's worth exploring.

As already mentioned, the best-known multidimensional measures are Howard Gardner's seven dimensions of intelligence: linguistic, logical, musical, spatial, kinesthetic, interpersonal, and intrapersonal.[5] Gardner's choices of dimensions hardly can be classified as ad hoc. Each dimension satisfies seven criteria: evidence that the intelligence is located in a specific part of the brain, the existence of prodigies, stages of mastery, and so on.

Robert Sternberg offers a second multidimensional measure of intelligence, one with three dimensions: analytic intelligence, creative intelligence, and practical intelligence.[6] Analytic intelligence can loosely be translated as IQ; it emphasizes the ability to solve test problems. Creative intelligence captures someone's ability to apply past experiences to new problems and to combine ideas, which resonates with our idea of toolboxes. Creative intelligence partially tests the ability to combine tools. Practical intelligence captures a person's ability to apply scholarly knowledge to real-world situations. A person of high practical intelligence can apply her tools when confronted with how much wood to buy to build a deck, but may perform poorly on a math problem. A person with low practical intelligence may be able to solve calculus problems for the area under a graph and then buy five times as much paint as needed when redecorating a room. These two people may marry. If so, all will be fine.

Rankings and Settling Things on the Field

While single dimensional measures of intelligence create rankings, multidimensional measures need not. We can say that

Table 5.1:
Sternberg intelligence scores

Person	*Analytic*	*Creative*	*Practical*
Kathleen	60	95	80
Patrick	90	55	85
Paul	55	70	70

someone with an SAT score of 700 verbal and 700 math did better than someone who scored 600 on both sections, but we cannot say whether she did better than someone who scored 800 on the math section but only 600 on the verbal section. Without placing weights on the two sections of the exam, we cannot even say that she did better than someone who scored 710 on the math section and 600 on the verbal section. We could weight the sections to create a combined score—and a one-dimensional ranking—but when we do, we make an implicit value judgment about the parts' relative merit. Even more problematically, if we allow ourselves to vary the weights, we can change the rankings among a group of people.

Suppose that we want to rank Kathleen, Patrick, and Paul based on their test scores. An ordering of Kathleen, Patrick, and Paul is a list with a greater-than relation between their names. The expression

$$\text{Kathleen} > \text{Patrick} > \text{Paul}$$

represents Kathleen coming before Patrick and Patrick coming before Paul. Orderings are transitive so Kathleen must also come before Paul. In this case, the relation > could signify age. Kathleen might be older than Patrick, and Patrick older than Paul. Transitivity implies that Kathleen must also then be older than Paul. Such orderings by age are possible because age is one-dimensional.[7] Let's assume that Kathleen, Patrick, and Paul all take Sternberg's intelligence test and receive the scores shown in table 5.1.

Based on these scores, we cannot rank Kathleen and Patrick. She scores higher on creative intelligence. Patrick scores higher

TABLE 5.2:
Total Sternberg intelligence

Person	*Sum of Test Scores*
Kathleen	235
Patrick	230
Paul	195

on analytic and practical intelligences. Similarly, Patrick and Paul cannot be ranked. Patrick scores higher on analytic and practical intelligences. And Paul scores higher on creative intelligence. Kathleen and Paul, however, can be ranked (sorry Paul!). Kathleen scores higher on all three measures. These partial rankings invite confusion. Kathleen is more intelligent than Paul, while Patrick isn't. And yet Kathleen is not more intelligent than Patrick. These paradoxes arise because we are squeezing several dimensions down to one. In doing so, we're losing information and obscuring differences.

One way to get a complete ranking is to sum the intelligence scores on the three tests. Their total scores would be as shown in table 5.2.

Now we can rank them. However, this one-dimensional ranking denies the independence of the three types of intelligence. And it weights all three types of intelligence equally. This assumption is arbitrary. In some contexts, analytic intelligence may be more important than creative intelligence and practical intelligence. If we give analytic intelligence double the weight of creative intelligence to create a Double Math Sternberg Intelligence score, we then make Patrick the most intelligent.[8]

This ability to shift weights and alter rankings creates serious problems. Most ranking systems cope with multiple attributes by assigning weights to each attribute and summing them up. For example, *U.S. News and World Report* magazine assigns weights to attributes when ranking colleges and universities. Schools have attributes: test scores of students, graduation rates, faculty–student ratios, and so on. The magazine assigns each attribute a weight that determines the relative importance of the attribute. The score for a school equals the weighted sum of its attributes' values. These one-dimensional scores can then be arranged from highest to lowest. But what if you, as an applicant, would choose to weight various attributes differently than *U.S. News* does? Maybe you want a good school that's within one hundred miles of your house. You'd come up with a different ranking. But of course schools would lose their bragging rights. No school wants to take out an ad saying that it's "closest to Conrad's house for the fifth year in a row."

Granted, if one school outperformed another on every attribute, then regardless of the weights assigned to attributes it would be ranked higher, just as Kathleen would always be ranked above Paul regardless of the weights placed on the three types of intelligence. However, as the number of dimensions becomes large, the odds that one person or school scores higher than another on every dimension become small.

The arbitrariness of the weights makes us uneasy about these rankings. Witness the dissatisfaction with the Bowl Championship Series rankings in college football. These rankings are based on a score that combines polls, computer rankings, win–loss records, strength of schedule, and quality wins. In this particular case, lower scores are better. The two teams with the lowest scores earn the right to play for college football's national championship, a process many consider unfair. Fans want the best team to be determined on the field of play. They want a game to decide the best team.[9]

All this talk of football serves an important purpose. We can show that settling matters "on the field," so to speak, in head-to-head competition, may not work either. A best team may not exist. This same logic applies to people. If we had some way of having people compete to see who was the most intelligent, head-to-head (so to speak), we might have no clear winner for the same reason. With all due apologies to Marilyn Vos Savant, we probably could not find the smartest person in the world even if we produced a game show called *The World's Smartest Person* and determined a champion.

We can use something called the Colonel Blotto Game to call into question the notion that competition reveals the best team.

TABLE 5.3:
Colonel Blotto strategies

Player	*Door 1*	*Door 2*	*Door 3*
USC	40	20	40
Michigan	35	40	25
Florida	20	35	45
Oklahoma	33	33	34

In Colonel Blotto, each of two players has one hundred playing pieces. Each must array these pieces in front of three doors.[10] Whichever player has the most pieces in front of a door wins that door. The objective in Colonel Blotto is to win the most doors. Colonel Blotto has no single best course of action. Any placement of the pieces can be defeated. If one player places fifty pieces in front of the first and second doors and none in front of the third, a second player can place sixty pieces in front of the first and forty in front of the third and win two doors in the process.

To model football using Colonel Blotto, we can let door one represent one team on offense and the other on defense, door two the opposite combination, and door three special teams. To model lawyers, the doors could represent knowledge of the law, charisma, and recall of facts. The Colonel Blotto game doesn't capture competitive situations perfectly—no model does—but it's a decent approximation. Imagine four participants in a Colonel Blotto tournament. I'll call them USC, Michigan, Florida, and Oklahoma. They have the strategies shown in table 5.3.

USC, Michigan, and Florida all defeat Oklahoma because each of the first three schools has more than thirty-four pieces in front of two of the doors. If USC plays Michigan, USC wins by winning Doors 1 and 3. If Michigan plays Florida, Michigan wins the first two doors, and therefore wins. If Florida plays USC, Florida wins. Notice that USC beats Michigan, Michigan beats Florida, and Florida beats USC—a *cycle*. (Cycles reappear later when we cover diverse preference aggregation.) Given this cycle, none of these three teams should be thought of as better than any of the others. All are better than Oklahoma, however.

Suppose that we hold a tournament to determine a national champion from among these four teams. Of the three teams that form the cycle—USC, Michigan, and Florida—one plays Oklahoma in the first round. That team, let's suppose that it is Michigan, defeats Oklahoma. In the other game, Florida defeats USC (in a thriller). Michigan and Florida then play in the Tostitos–Subaru Fiesta Bowl.com, and Michigan wins. Thousands of fans parade in the streets. The next day, students and alumni run to stores to buy hats, T-shirts, and mugs commemorating the event. As they sing it in Ann Arbor, "Hail to the victors valiant, hail to the conquering heroes."

Hold on. If we look closely at the tournament, we see that Michigan won because they played Oklahoma in the first round. If we look at the matchups, we see that whichever teams draws Oklahoma in the first round wins the national championship.[11] So rather than sell shirts that say, "2008 National Champions," Moe's Sport Shop might more honestly sell shirts that say, "The team that got to play Oklahoma first." As titles go, it's not appealing, but at least it's accurate.

Is the Colonel Blotto model reasonable here? Absolutely. Cycles like the USC–Michigan–Florida cycle are widespread. A look at head-to-head competition in almost any conference in any year reveals these cycles, yet we delude ourselves by posting arbitrary rankings and by crowning champions. So long as we know that the title of champion means "winner of the tournament," we're on solid ground. But we should not confuse champion with best.[12]

Let's wrap this all up. With multiple dimensions, no clear winner may exist. Even attempts to settle it on the field won't work. What gets settled on the field depends on the tournament pairings. We cannot expect to find a best team, a best lawyer, or a smartest person. Those soft mumblings that ring hollow when people try to reassure children with lower aggregate SAT scores than their siblings, about everyone being different blah blah blah, should be spoken loudly and clearly. We *are* all different. We have different tools.

The Toolbox Framework

The toolbox framework is pretty simple. First, we think about the set of all possible tools, which consists of all possible knowledge, skills, abilities, heuristics, interpretations, and perspectives that a person might acquire. But an individual can't possibly acquire all of these, so a person's toolbox is the subset of tools that a person has acquired.

We often can apply tools across domains. We use basic arithmetic when balancing our checkbooks, counting change, and buying milk. A drug or surgical procedure developed in one domain often can be applied in others. Botox, a mild form of botulism, was developed to help solve a less-than-dire medical problem—wrinkles. Botox reduces facial wrinkles by numbing muscles—temporarily, so you do have to keep visiting your doctor (or host a Botox party). The ability to numb muscles is just a tool. So once Botox was developed, doctors could apply it to other problems: medical problems that involve overstimulated muscles, such as stuttering, some forms of ulcers, and even cerebral palsy. But then not all tools work on all problems. You can't fix a car with a blender. At least not very well.

That's it. That's the toolbox framework. Unpacking its implications leads to some surprises and insights.

Three Ways to Think about Tools

We now consider two people, Bobbi and Carl, and, using the toolbox framework, try to figure out who's smarter. We pick the smarter person at the end of the chapter. Bobbi and Carl possess distinct, large toolboxes, though Bobbi's contains a few more tools than Carl's. She has twenty and he has fifteen, so she must be smarter. But in the toolbox framework, to be smarter than Carl Bobbi would have to possess each of Carl's tools. If that's not true, Carl may make greater contributions. He may be more successful because he may have more appropriate tools.