“I’m going to email him,” I tell my friend. She looks at me like I am crazy. I am starting to get used to this reaction. “Scott, you can’t just randomly email the head of the Cambridge Psychology Department! Who does that?” I sit down and think this through. I really want to step into the lion’s den and learn everything there is to know about intelligence. I need a mentor. I also am rather scared to stay in the country at the moment due to recent attacks. Which is why I searched online for intelligence experts in England. I came across Professor Nicholas J. Mackintosh, an expert on IQ and human intelligence. Since he was at Cambridge University, I dreamed of spires, ivy, and stained glass windows. Images of the prep school I was denied access to as a child came flooding back. “What do I have to lose?” I told my friend. “I’m gonna do it.”

To my surprise, Professor Mackintosh is very welcoming, and happy for me to work with him. Carnegie Mellon grants me a one-semester, six-month leave of absence. To work out the details, I fly to England to have a chat with the professor. In one of the most surreal moments of my life, I find myself sitting in the waiting area to see the head of the Cambridge University Department of Experimental Psychology. My heart and mind race. What if I’m not smart enough for this? This is Cambridge, for heaven’s sake. What if he gives me an IQ test? Then, the door opens.

“Mr. Kaufman?” I stand up and find myself face to face with a distinguished-looking elderly man. His smile puts me immediately at ease. “Come inside.” I walk into his office and look around the room. Lots of books with the word “intelligence” on the cover. I feel a tingle. Am I finally going to be able to solve the mysteries of intelligence that brought me so much pain and anxiety as a child?
“So let me get this straight,” he says, scratching his gray hair. “You’d like to take six months off from your American University and learn everything there is to know about intelligence from me?” I nod, just as a tall, middle-aged woman enters the room. “This is Sheila,” he says, beckoning her to sit down next to me. “She teaches at a local Sixth Form College, which is the equivalent of an American High School. We’re working on a study investigating differences in spatial intelligence. We would more than welcome your help. Is this the sort of thing you’d like to be involved with?”

I can barely contain my excitement. The room starts spinning. He continues: “I think we’d be able to set you up in King’s College for the duration of your stay as a Visiting Scholar. You could engage in the social life of the College. It’s a wonderful place. I’ll give you a tour later today.”

Now I can barely breathe. Cambridge consists of thirty-one colleges, in which the students live and work. King’s is one of the oldest (founded in 1441) and arguably the most beautiful, most noted for its magnificent late-Gothic perpendicular chapel. I look up at Professor Mackintosh, then over to Sheila. Feeling as though the room is spinning, I muster just enough strength to respond: “I’m in.”

So my journey to master the science of intelligence began. As I learned everything I could about measuring intelligence—keeping my own past a secret—I encountered many surprises. What surprised me most was the disconnect between two worlds—clinicians interested in the practical application of IQ and research scientists interested in understanding the nature of human intelligence. Over the past 100 years, the majority of scientists who studied intelligence weren’t interested in IQ, per se. For them, IQ was just a proxy for something called $g$ (general intelligence factor)—which many of them conceptualized as the essence of human intelligence. In fact, the first scientific study to discover $g$ was published a year before the first IQ test appeared.

I realized that if I truly wanted to understand the foundation upon which modern IQ tests were constructed and eventually applied, I had to tame $g$.

At the turn of the twentieth century, the British psychologist Charles Spearman made a startling discovery. He went to a few local British schools in the nearby village and collected student grades, teacher ratings, and scores on tests of sensory discrimination. He then ranked each child’s performance relative to his or her peers on this assortment of outcomes, and calculated by
hand the strength of the relationship among the different outcomes. This is what he found among six areas of school performance:

<table>
<thead>
<tr>
<th></th>
<th>Classics</th>
<th>French</th>
<th>English</th>
<th>Math</th>
<th>Pitch</th>
<th>Music</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classics</td>
<td>—</td>
<td>.83</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>French</td>
<td>.78</td>
<td>—</td>
<td>.67</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>English</td>
<td>—</td>
<td>.70</td>
<td>—</td>
<td>.64</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Math</td>
<td>.66</td>
<td>.67</td>
<td>.54</td>
<td>.45</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Pitch</td>
<td>.63</td>
<td>.57</td>
<td>.51</td>
<td>.51</td>
<td>.40</td>
<td>—</td>
</tr>
<tr>
<td>Music</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Each number represents the strength of the correlation between each set of tests, ranging from .40 to .83. These results surprised Spearman. Some relationships made good sense, such as the high correlation between classics and French (.83). They both, after all, involved facility with languages. But some correlations had a less obvious explanation, such as the relation between math grades and pitch discrimination, as rated by their school music teacher (.45).

In fact, the most unexpected finding of them all was that every single correlation was positive. You would expect that the more time a student puts into one area of study, the more performance in other areas suffer. But that wasn’t the case. Students who performed well in one school subject tended to perform well in other school subjects. Spearman referred to this phenomenon as the positive manifold.

Based on the patterns of student performance, Spearman was able to rank all of the students on a single dimension, which he called the g factor, and showed that performance in each class was related to this factor to a varying degree. For instance, classics was correlated with this factor (.95) much more than music (.65), but all subjects were still positively related to this factor. Spearman also showed that each subject also had its own unique set of skills (for instance, classics involves learning how to conjugate Latin verbs), which he labeled s for “specific.” In his groundbreaking 1904 paper “General Intelligence, Objectively Determined and Measured,” Spearman proposed that the cause of the g factor (or g, for short) was a general cognitive ability he called “general intelligence.”

Spearman presumed that the g factor—which explained the largest source of differences in performance across all school subjects—represented something we would want to refer to as “general intelligence.” It is important to
recognize, however, that this was Spearman’s decision, to label the factor “general intelligence.” To derive \( g \), Spearman used a statistical method called factor analysis, which is a potentially useful way of telling a story with data. The technique simplifies a large number of correlations by finding common sources of variation. Using factor analysis, you can come up with a manageable number of factors instead of having to deal with the unwieldy correlations among the tests. But as is the case even to this day, it is entirely up to the psychologist to label the factors.

If Spearman wanted to be really descriptive, he could have just labeled the factor “general test-taking ability,” or “the general ability to quickly choose the response deemed correct by the psychologist,” but that wouldn’t have made for such a sexy journal article title. And it certainly wouldn’t have sounded as scientific—Spearman was proud that he “objectively” discovered “general intelligence,” as though he discovered a law of human nature akin to any other physical law of nature.

With that said, some frequent criticisms misrepresent Spearman’s position. In The Mismeasure of Man, the late Harvard paleontologist Stephen Jay Gould accused Spearman of reifying \( g \) (that is, treating \( g \) as though it really exists).\(^2\) While this is certainly a fair charge against some modern-day intelligence researchers (see Chapter 13), a close reading of Spearman suggests he knew better. He understood the difference between the \( g \) factor and “general intelligence,” which he proposed is the psychological ability that causes the \( g \) factor to emerge.

Remember, Spearman discovered the positive manifold. It didn’t have to be the case—and certainly wasn’t expected—that all his variables would positively correlate with one another. It’s the positive manifold that needs explaining: Why do people who do well on one cognitive test tend to perform well on all the others? Spearman argued that general intelligence was an innate general “mental energy” that pervaded performance on all tests of cognitive ability. This was just a hypothesis; he knew it would require further testing.

\( \sim \)

Spearman’s \( g \) didn’t go unchallenged.\(^3\) An American psychometrician, Louis Thurstone, showed that his battery of tests formed clear clusters of abilities.\(^4\) For example, his spatial tests were more strongly related to each other than to the verbal tests. Using a different method of factor analysis than Spearman, he argued that \( g \) doesn’t exist. Instead, he argued that we differ from each other on seven independent primary mental abilities: verbal comprehension,
verbal fluency, number facility, spatial visualization, associative memory, perceptual speed, and inductive reasoning.

Both Spearman’s and Thurstone’s theories were heavily influenced by their authors’ own psychological stances on pressing issues of the time. Spearman shunned “faculty” psychology that attempted to decompose the mind into distinct faculties. Thurstone, on the other hand, embraced the compartmentalization of the human mind and strongly believed that human intelligence is the result of a number of independent components.

But who was right? No matter how hard Spearman tried to sweep the importance of Thurstone’s primary mental abilities under the rug, they kept cropping up whenever he analyzed sufficiently large batteries of tests. At the same time, no matter how hard Thurstone tried to make g disappear through his own special method of factor analysis (“rotation to simple structure”), individual cognitive abilities still tended to correlate positively.\(^5\)

Again, some critics miss this fact. Gould argued that whether or not g appears depends on your method of factor analysis. But this is not correct. Today there are formulas available to mathematically transform the g calculated by one method of factor analysis into the g calculated by a different method of factor analysis.\(^5\) In fact, this is a story that is repeated over and over again in every generation. In the 1960s, Raymond Cattell and his student John Horn argued that “fluid” and “crystallized” intelligence are independent abilities.\(^7\) They claimed that “fluid intelligence” is influenced by biology and independent of education and experience, whereas “crystallized intelligence” is influenced by learning and experience, and involves accumulated knowledge and skills. They quickly realized, however, that all of their tests positively correlated with each other. By the 1990s their model of intelligence included nine to ten different positively correlated cognitive abilities.\(^8\)

Eventually a compromise was made, based on a grand synthesis. In 1993 John Carroll published his monumental analysis of over fifty years of factor analytic research.\(^9\) He analyzed over 400 datasets—including some of the very same datasets used by Spearman and Thurstone. He looked at a wide (but certainly not all-inclusive) range of cognitive ability tests, including language, reasoning, memory, learning, visual perception, auditory perception, idea production, cognitive speed, and psychomotor abilities. Not all of the tests included abilities explicitly taught in school, not all were timed, and not all were administered in a paper-and-pencil format. Nonetheless, despite this diversity of cognitive ability tests and testing formats, g still emerged.
Based on a careful analysis, Carroll argued for a hierarchical model of intelligence, which he called the “Three Stratum Theory.” At the top of the hierarchy (“Stratum III”) is $g$—the common factor among all the diverse tests. This stratum is the most abstract, and doesn’t represent any actual cognitive ability. $g$ strips all of the specificity out of each test and represents solely what they all have in common. At “Stratum II” are eight broad abilities, each one varying in its relation to the $g$ factor at the topmost level: fluid intelligence, crystallized intelligence, general memory and learning, visual perception, auditory perception, knowledge retrieval ability, cognitive speediness, and decision speed.

At the lowest level of the hierarchy is “Stratum I”—sixty-nine narrow abilities that make up the eight broad abilities. For instance, “fluid intelligence” is made up of nonverbal, verbal-deductive, and quantitative forms of fluid reasoning, and “crystallized intelligence” is made up of a cluster of narrow abilities such as reading comprehension, inferring the meaning of peculiar words from their context, reading decoding, reading speed, spelling ability, writing ability, lexical knowledge, listening ability, and phonetic coding. As we saw in Chapter 2, Carroll’s model of cognitive abilities strongly influenced the CHC theory of cognitive abilities, which greatly informed the construction of numerous contemporary IQ tests (see the Appendix).

Rightly so, Carroll’s massive synthesis garnered immense respect from his colleagues. Horn referred to Carroll’s book as a “tour de force summary and integration” comparable to “Mendeleev’s first presentation of a periodic table of elements in chemistry.” Most can only dream of such recognition among colleagues!

At last, all seemed to be resolved. The field now had a road map—a taxonomy of patterns of covariation among a wide-ranging (but still limited) set of cognitive abilities. While there’s still some disagreement today as to the precise structuring of the levels of the hierarchy, new, more sophisticated statistical techniques have confirmed that the hierarchical structure fits the data the best. What’s more, you’d be hard-pressed to find a taxonomy of individual differences in cognitive abilities in which $g$ doesn’t sit securely at the top. But just how general is $g$?

Alfred Binet succeeded in devising a test that is highly relevant to school performance: $g$ is substantially correlated with concurrent academic achievement, and also does a good job of statistically predicting future academic
performance. Ian Deary and colleagues gathered data on over 70,000 British school students who had taken the Cognitive Abilities Test (CAT), which measures verbal, quantitative, and nonverbal fluid reasoning abilities. They found positive correlations between the CAT’s g factor assessed at age 11 and standardized achievement test performance on all twenty-five individual school subjects at age 16, though the strength of those correlations varied.

The highest correlations were with mathematics (.77), science (.68), and English (.67), and the lowest was in art and design (.43). Nevertheless, all the correlations were medium to high. Even physical education (.55), drama (.47), religious education (.52), and business (.56) were moderately related to g. From a practical perspective, g accounted for about 65 percent of the total variation in total educational achievement, which leaves about 35 percent of the differences in student achievement unexplained. For some school subjects, such as art and design, over 80 percent of the variance was unexplained. This certainly leaves plenty of room for academic achievers who exceed statistical predictions based on their g ranking (see Chapter 3). This also leaves plenty of room for other factors to influence academic achievement, such as specific cognitive abilities, motivation, mindset, self-regulation, environmental support, and so on. Nevertheless, from a scientific perspective, these correlations are impressive, and do suggest that g is relevant to learning a wide range of material.

Not only does g pervade learning across a wide range of subjects, but many different standardized tests may be measuring the very same g—or at least very close. Wendy Johnson and her colleagues found that the g extracted from one diverse IQ test battery is almost perfectly correlated with the g extracted from a different IQ test battery. What’s more, two tests used in the United States for college admissions—the SAT and ACT—are correlated with g about .91, and with each other almost perfectly, once you correct for the reliability of each test.

Recently I teamed up with Matthew Reynolds, Xin Liu, Alan Kaufman, and Kevin McGrew to see whether the g factor extracted from IQ tests is the same as the g factor extracted from standardized tests of academic achievement. To find out, we looked at two large, nationally representative data sets with over 7,000 participants in total, and two independent individually administered sets of test batteries. Cognitive g was measured by the use of IQ test batteries that measure the CHC domains (see the Appendix), and academic achievement g was measured by standardized tests of reading and writing and quantitative knowledge. We found that cognitive g and academic
achievement $g$ were not identical (there was still a significant amount of variance unexplained), but the relationship was really close. While the correlations generally increased with age, the range of correlations across age groups (spanning ages 4 to 19) was .77 to .94, with an average correlation of .83.

$g$ also pervades reasoning about content that goes beyond school subjects. As part of my doctoral dissertation I teamed up with Colin DeYoung, Deidre Kolarick, and Jeremy Gray (my primary PhD advisor) to look at the role of $g$ in reasoning across different kinds of content. In particular, we looked at *contextualized* deductive reasoning, which involves if-then reasoning situated in real-world contexts. We looked at three types of problems: social exchange problems, precautionary problems, and arbitrary-rule problems. *Social exchange* problems concern the mutual exchange of goods or services between individuals, and involve detecting whether one party might be taking a benefit without fulfilling an obligation (“If you borrow my motorcycle, then you have to wash it”). *Precautionary problems* involve rules relating to avoiding potential physical danger (“If you surf in cold water, then you have to wear a wetsuit”). Finally, *arbitrary-rule problems* involve rules that are contextualized in realistic scenarios but are arbitrary (“If the soda is diet, then it has to be in a purple container”).

We found that precautionary and social exchange reasoning problems were solved more frequently and more quickly, on average, than reasoning about arbitrary rules. This suggests that reasoning about contextualized and familiar real-world scenarios does facilitate reasoning. Nevertheless, looking at individual differences in deductive reasoning, we found that all seventy problems—regardless of content—were substantially positively correlated with each other. In fact, the reliability of all seventy problems was .88. What’s more, this deductive reasoning common factor was significantly correlated (.45) with the $g$ factor extracted from other well-known tests of verbal, spatial, and fluid reasoning. This suggests that covariation between people on these three forms of reasoning—even though they were contextualized in real-world scenarios and involved reasoning on different content—were not completely independent of each other.

How does $g$ fare among the multiple intelligences that have been proposed in recent years? Have any of them dethroned $g$? Consider Howard Gardner’s theory of multiple intelligences (see Chapter 4). Gardner explicitly states that his eight intelligences are independent of each other. He argues that schools tend to overemphasize logical-mathematical and linguistic intelligence to the exclusion of spatial, bodily-kinesthetic, musical, interpersonal, intrapersonal, and naturalistic intelligence.
In a paper called “Beyond g: Putting Multiple Intelligences Theory to the Test,” Beth Visser, Michael Ashton, and Philip Vernon constructed a test to measure Gardner’s eight intelligences, selecting two paper-and-pencil tests for each intelligence.\(^\text{17}\) Lo and behold, a positive manifold emerged—all the tests were positively correlated with each other. But once again, not all tests were equally related to \(g\). Linguistic, spatial, and logical/mathematical reasoning tests were most strongly related to \(g\). Nevertheless, even measures of interpersonal and naturalistic intelligence (ability to classify one’s surroundings) were moderately related to \(g\).

Some tests, however, were much less related to \(g\): bodily-kinesthetic intelligence, tonal and rhythmic music ability, and intrapersonal accuracy (knowing thyself) were least related to \(g\). Visser and colleagues point out that some of the lower test correlations with \(g\) were most likely due to the unreliability of the tests. They also argue that the overall pattern of findings makes sense because they consider these tests less “cognitive” in nature.

In his response, Gardner argued that “while the intention is praiseworthy, the actual effort recreates the very conditions that I had sought to challenge.”\(^\text{18}\) While Gardner admitted that \(g\) “regularly emerges whenever a battery of tests is administered,” he argued that we still have little understanding of what the positive manifold actually is.\(^\text{19}\) Gardner went on to argue for the use of “intelligence-fair tests” that assess his multiple intelligences through methods other than a paper-and-pencil test and attempt to eliminate, as much as possible, linguistic or logical components.

Gardner gave an example of measuring spatial intelligence by seeing how people can navigate an unfamiliar terrain, or measuring interpersonal intelligence by examining how people negotiate with other people in the real world. These points are certainly worth considering, but they are problematic upon further examination. In reality, it’s very difficult to develop tasks that completely isolate a single cognitive process and entirely eliminate the use of logic and/or language. Also, while Gardner doesn’t mention it, Carroll did find that the ability to navigate unfamiliar terrains was significantly related to \(g\). So although Gardner’s proposal to construct fairer tests for people to fully express themselves is a worthy one, the positive manifold still exists. And from a scientific perspective, this still requires explaining.

What about other theories of intelligence that attempt to go beyond \(g\)? Again, the same story emerges. If you look closely at the research, you’ll see that Sternberg’s tests of his three so-called intelligences—analytical, creative, and practical—are moderately correlated with each other, and with traditional measures of intelligence.\(^\text{20}\) The same with emotional intelligence. Tests of the
four branches of “emotional intelligence”—perception, facilitation, understanding, and management—are also moderately correlated with each other and with traditional measures of intelligence. Although to be fair, neither Sternberg nor the originators of the theory of emotional intelligence—Peter Salovey and John Mayer—claimed that emotional intelligence is completely independent of g. Only Gardner has made that strong claim.

None of these findings suggest that attempts to go beyond g are fruitless. On the contrary, they highlight the diversity and richness of human intelligence. Each additional form of intelligence deserves appreciation, predicting important life outcomes above and beyond g. But out of the hundreds and hundreds of studies conducted on cognitive abilities measured by brief, reliable tests in a laboratory setting, it is difficult to find a single one where the positive manifold is completely absent. The positive correlations among diverse tests of cognitive abilities is such a robust finding that the psychologist Christopher Chabris referred to it as the “Law of General Intelligence.”

No doubt, if Spearman were alive today, he’d be happy to hear his initial findings have become a law! More than 100 years ago, Spearman proposed that the precise makeup of the cognitive test battery shouldn’t matter, calling this the “indifference of the indicator.” He argued that you could substitute one cognitive test for another, and you would still find positive correlations among all the tests in the battery. According to Spearman, all that matters is that the substituted test has the same level of difficulty, and the battery is diverse enough that any characteristics unique to any one test will wash out. This is how David Wechsler, who spent a few months working with Spearman, conceptualized global IQ scores on his test (see Chapter 2). Although Wechsler created “verbal” and “performance” subscales, he believed that all “the subtests are different measures of intelligence, not measures of different kinds of intelligence.”

But what, really, is this elusive, seemingly pervasive g? Does it really represent a source of human variation we’d want to label “general intelligence”? Is Gardner right that we still have little understanding of what causes the positive manifold to emerge? Now that I’ve defended the existence of the positive manifold against nearly every attempt to challenge it over the past 100 years, it’s time to get to the bottom of g.

In The Mismeasure of Man, Gould rightly noted the limitation of correlations. Just because two tests are highly correlated—say, that between mental
rotation and vocabulary—doesn’t necessarily mean that the two tests are measuring the same psychological process. There are numerous alternative explanations for the positive manifold. For one thing, it’s possible that those who have a large vocabulary may also be good at mental rotation—not because those two tests are measuring the same mental process, but simply because we live in a society that encourages the development of vocabulary (English class) and spatial abilities (geometry class) in school.

Another possibility was pointed out to Spearman in 1916 by fellow Englishman Godfrey Thomson.25 Using keen mathematical proofs, Thomson showed that $g$ could theoretically arise through a large number of independent “bonds” or associations between basic processes. For instance, let’s assume that the mind consists of 500 basic processes. If two tasks each depend on 250 of those processes, there is a good chance that about half, say 125 of them, will be engaged by both tasks, producing a sizable correlation of .50. According to this account, a single psychological process common among all the tests is not necessary to produce $g$. All that’s needed is a partial overlap between processes engaged by different tests.

For example, consider the task of building a house. Even though making a bed, a wall, and a garden shed are vastly different tasks, people will generally most likely use a ruler in all three tasks, and a hammer in two of the tasks. So it’s almost impossible to study one cognitive process (“hammer”) in isolation, because whatever task you develop (building a shed) will also recruit other abilities shared in tasks where you aim to measure something else (a steady hand).

Thomson’s point still holds water. In their 2009 paper, David Bartholomew, Ian Deary, and Martin Lawn showed that there is no statistical way of deciding whether Spearman’s view or Thomson’s view is correct.26 So how are we ever to get at the truth? To attempt to get to the bottom of this mystery, we’ll investigate the issue at multiple levels of analysis, including the behavioral, biological, and cultural.

As a first pass to help us wrap our heads around $g$, let’s look at the features of cognitive tests that are most strongly correlated with $g$, and try to figure out what task demands they have in common. Imagine a bunch of cognitive tests spread around a circle. The closer to the center, the better the test measures $g$. The more the test falls off to the periphery, the less related it is to $g$. Figure 10.1 shows what that looks like in one study.27

What can we glean from this? One thing we notice is that processing speed tasks lie at the periphery of the circle. While still on the $g$ map, if we
want to understand g, sheer speed of basic information processing is not going to be the magic bullet. Let’s get closer. Surrounding the center of the circle are a bunch of tests that appear different on the surface. Each test differs in its content: verbal, quantitative, and spatial. Some involve rotating images in your mind, others involve mentally manipulating mathematical operations, and still other tests involve conceptual reasoning. As you can see, there are also quite a number of tests that involve analogical reasoning across different content (numbers, words, geometric forms). It’s almost as if figuring out what all of these seemingly diverse tests have in common is an IQ test in itself!

Let’s zoom in right to the very center, to the very heart of g. Here we find the Raven’s Progressive Matrices Test, or Raven’s for short. On each Raven’s
question, you are presented with a 3x3 matrix and you have to identify the missing piece that completes the pattern (see Figure 10.2 for an example).

What does it take to do well on this test? It turns out there are only a handful of rules required to solve all the items on this test. The easier problems require you to apply a single rule—such as adding or subtracting a single attribute (such as a line). But the harder ones require combining multiple rules, and juggling multiple attributes (such as shapes, sizes, and colors). The difficulty in solving the Raven’s items is that you have to sort out the relevant attributes from the irrelevant attributes and hold the rules in your mind while testing them. And when some rules don’t work out, you have to know when to stop going down that path and start over. This task—which requires the ability to discover the abstract relations among novel stimuli—is a good measure of nonverbal fluid reasoning.

Note that I said fluid reasoning, not fluid intelligence. In recent years Wendy Johnson and Thomas Bouchard Jr. have shown that the traditional distinction
between “fluid” and “crystallized” intelligence is misguided. Through rigorous statistical testing, they found that cognitive tests of ability tend to covary depending on whether they are verbal or nonverbal,* not by how much they are influenced by culture. As we saw in Chapter 4, this has important implications for assessing giftedness. It is incorrect to assume that tests of nonverbal fluid reasoning are somehow measuring intelligence in its “purest” form, completely divorced from experience and culture. Likewise, just because a test measures vocabulary or general knowledge doesn’t mean it doesn’t place demands on fluid reasoning. Vocabulary tests often require choosing the right meaning for infrequent words. That requires quite a bit of inference and deduction. Both verbal and nonverbal tests of cognitive ability involve fluid reasoning, and fluid reasoning skills in turn are influenced both by genes and culture (which are always interacting with each other; see Chapter 1).

Fluid reasoning is not just any kind of reasoning, however. Tasks that engage a well-organized knowledge base place fewer demands on fluid reasoning. Fluid reasoning involves inferring underlying patterns based on minimal evidence. This requires concentration, planning, breaking the task down into subgoals, and managing these problem-solving goals in working memory.

If you recall from Chapter 9, the human working memory capacity limit is on average about four chunks. This is the number of meaningful bits of information that can be held in consciousness at any point in time. People differ in capacity, however, with the typical range being two to six chunks. And even within this small window, variation in working memory is significantly related to variation in fluid reasoning performance. Consider a commonly administered working memory task called the reading span test:

INSTRUCTIONS: I will present a series of five sentences one at a time. Read each sentence and remember the last word of the sentence. When you see the word “RECALL” you should recall the last words of all the sentences.

*Technically, Johnson and Bouchard distinguish between “verbal” and “perceptual” abilities. Because the term “nonverbal reasoning” will be more intuitive to readers, I decided to replace “perceptual” with “nonverbal.” Also, as Colin DeYoung has noted, given that nonverbal memory and perceptual fluid reasoning tasks were both included in the perceptual factor identified by Johnson and Bouchard, “nonverbal” is sensible as an inclusive label for the factor. I should also note that Johnson and Bouchard also discovered a smaller third factor representing mental rotation skills. See C. G. DeYoung, “Intelligence and Personality,” in The Cambridge Handbook of Intelligence, ed. R. J. Sternberg and S. B. Kaufman 711–737 (New York: Cambridge University Press, 2011).
1. The hunter chased after the man on horseback.
2. After a time, the correspondent looked around.
3. The children ate the cake with a spoon.
4. Helen is expecting tomorrow to be a bad day.
5. The ship left with a noise and went up into space.

RECALL

ANSWERS: Horseback, around, spoon, day, space

Most people can’t go beyond five sentences without getting a serious headache! Daneman and Carpenter found a correlation of .50 between this task and SAT verbal scores, and they found even higher correlations with tests of reading comprehension.33 They found much lower correlations between fluid reasoning and versions of a memory task that only required remembering words.

It seems it’s the dual requirement of remembering the last word of each sentence while processing the sentences that makes it a test of working memory. To illustrate, consider this lovely passage from Proust:

On the occasion of this first call which, after leaving Saint-Loup, I went to pay on Mme de Villeparisis following the advice given by M. De Norpois to my father, I found her in a drawing-room hung with yellow silk, against which the settees and the admirable armchairs upholstered in Beauvais tapestry stood out with the almost purple redness of ripe raspberries . . . Mme de Villeparisis herself, wearing an old-fashioned bonnet of black lace (which she preserved with the same shrewd instinct for local or historical colour as a Breton innkeeper who, however Parisian his clientele may have become, thinks it more astute to keep his maids in coifs and wide sleeves), was seated at a little desk on which, as well as her brushes, her palette and an unfinished flower-piece in water-colour, were arranged—in glasses, in saucers, in cups—moss-roses, zinnias, maidenhair ferns, which on account of the sudden influx of callers she had just left off painting, and which gave the impression of being arrayed on a florist’s counter in some eighteenth-century mezzotint.34

If you made it all the way through without your brain short-circuiting, try answering some questions about what you just read: Why was Mme de
Villeparisis wearing a black bonnet? Why would the innkeeper dress his maids in coifs and wide sleeves? Why did Mme de Villeparisis stop painting?

And this passage is only two sentences long! Comprehending complex passages such as this one is difficult in large part because it requires holding in memory the gist of prior sentences while reading new sentences, and then integrating the meaning of all of the sentences to form a coherent narrative. The more sentences you have to read, and the longer each sentence, the greater the demand on working memory. At the same time, task-irrelevant information, such as the humming of the radiator, or wandering thoughts about the big party Saturday, must be ignored.

For this reason, psychologists make an important distinction between short-term memory and working memory. Some researchers argue that the reason working memory and fluid reasoning are so strongly correlated with each other is because they both require the control of attention.* Today there is widespread evidence and acknowledgment among intelligence researchers that although working memory is not the same as fluid reasoning, it is indeed an important piece of the g puzzle.36

Cognitive psychologists made headway using their behavioral tests, but once new neuroscience techniques came along they were eager to see what pattern of brain activations were associated with cognitive performance. Of course, any one person’s IQ score is the result of multiple cognitive processes. The human brain is massively interconnected, with many specialized regions contributing to performance on every single cognitive task. To perform even the simplest tasks, such as quickly matching up letters with symbols, requires the activation of many brain regions, including regions associated with concentration, vision, sensory motor functions, and semantic retrieval. The key

*It should be acknowledged that working memory tests are more strongly related to fluid reasoning that match their content (e.g., spatial working memory is more strongly related to spatial reasoning than verbal reasoning, and verbal working memory is more strongly related to verbal reasoning than spatial reasoning). See P. Shah and A. Miyake, “The Separability of Working Memory Resources for Spatial Thinking and Language Processing: An Individual Differences Approach,” Journal of Experimental Psychology: General 125 (1996): 4–27; N. J. Mackintosh and E. S. Bennett, “The Fractionation of Working Memory Maps onto Different Components of Intelligence,” Intelligence 31 (2003): 519–531; S. B. Kaufman, “Sex Differences in Mental Rotation and Spatial Visualization Ability: Can They Be Accounted for by Differences in Working Memory Capacity?,” Intelligence 35 (2007): 211–223.
question is this: What brain regions are consistently associated with differences in performance across diverse cognitive tests?

In one of the earliest neuroscience studies of $g$, neuroscientist John Duncan and colleagues found that spatial, verbal, and perceptual-motor tasks that correlated more strongly with $g$ recruited the lateral prefrontal cortex (located just behind the forehead in the outermost region of the brain) much more than tasks that had low correlations with $g$. This study was important, because it was one of the first studies linking the prefrontal cortex—particularly the lateral region—to fluid reasoning.

Of course, the lateral prefrontal cortex doesn’t contribute to fluid reasoning in isolation. This brain region is heavily connected to other cortical and subcortical brain structures in the right and left hemispheres and recruits these areas depending on the content (verbal, spatial, etc.) and task requirements (number of dimensions, number of distractors). Nevertheless, further research on both humans and monkeys suggests that cells in the lateral prefrontal cortex are critical for maintenance of diverse content in working memory.

Consider a study by Jeremy Gray, Christopher Chabris, and Todd Braver. They gave 48 university students the Raven’s Matrices test to solve as well as a measure of working memory called the $n$-back working memory test. The $n$-back task requires people to view a sequence of items presented once every 2.36 seconds and indicate whether the current item is the same as or different from the item presented a certain number of steps back in the sequence. So on a one-back task, all you’d have to do is indicate if an item is the same as that presented on the immediately preceding presentation. Easy. But it quickly gets more difficult. The three-back requires thinking three items back. Still seems easy in theory, but trust me, it’s hard. This requires fast and constant online updating of the sequence in memory.

Gray and colleagues used words and faces in two separate three-back tasks. They also made the task even more difficult by presenting “lures,” which were stimuli that matched a stimulus that was seen recently, but not exactly three back. They included lures that were a two-back, four-back, or five-back match. These interfering lure trials placed a higher demand on attentional control mechanisms. Figure 10.3 shows an example of the different kinds of trials on the three-back task, using faces as stimuli.

They found that subjects who scored highest on the Raven’s test were better at the working memory task and showed more neural activity in the lateral prefrontal cortex and parietal lobe regions (located at the top of
the head toward the back) associated with sensory integration. What’s more, brain activity differences in these regions were most pronounced on lure trials, in which interference was high and attentional control was particularly necessary. This study suggests that fluid reasoning tasks recruit areas of the lateral prefrontal and parietal cortices to maintain focus on a goal to inhibit distractions, and rapidly update the contents of working memory. *

By 2007 a number of studies on the neuroscience of cognitive ability had accumulated. Rex Jung and Richard Haier reviewed thirty-seven neuroimaging studies that were then available, using a variety of methodologies (PET, fMRI, MRI spectroscopy, structural MRI imaging), and concluded that a specific network of brain regions is critical for cognitive test performance across different content. 

According to their Parieto-Frontal Integration Theory of Intelligence (P-FIT), different brain regions play a role at different stages of information processing.

In the first few stages, the brain’s temporal regions (located behind the ears) and occipital regions (located in the back of the head) process basic sen-

sory information, and the parietal cortex integrates this information. Then, the next “hypothesis testing” stage involves higher levels of abstraction and requires efficient information flow between the frontal and parietal regions, with white matter helping to reliably move information across the frontal parietal network. (White matter consists of axons surrounded by a fatty insulation called myelin that helps different gray matter regions of the brain communicate with each other. The prefrontal cortex and parietal cortex are gray matter.) Then once the best solution is determined, the anterior cingulate is recruited to select the appropriate response and inhibit alternative responses.

The P-FIT theory is not without its critics. One criticism is that Jung and Haier based their theory on individual differences at multiple levels of analysis: brain structure, function, and task comparisons (for instance, differences between high g and low g tasks). While these questions are all interesting, they are very different questions, and results from one level don’t necessarily apply to results at the other level. Another criticism is the consistency of the brain regions: only a few brain areas showed activations by more than 50 percent of the studies. Part of the problem may have been that the studies included in the review differed in the assortment of cognitive tests they administered to participants. Remember, Spearman argued that to properly measure g, you must administer a wide variety of cognitive tasks, or else you’re more likely to measure domain-specific skills.

Although these criticisms are certainly valid, since Haier and Jung’s 2007 paper numerous studies have consistently shown that communication between the lateral prefrontal cortex and posterior regions of the parietal lobe is related to differences in cognitive test performance across a range of tasks. The most striking evidence comes from a growing number of large-scale lesion studies that have all implicated the prefrontal parietal network as crucial for fluid reasoning.

In one recent large-scale lesion study, Aron Barbey and colleagues obtained access to a remarkable sample of 182 Vietnam veterans with highly localized brain damage due to penetrating head injuries. Consistent with the P-FIT theory, they found that damage to specific areas of the lateral prefrontal and parietal cortex impaired the integration and control of a distributed pattern of neural activity throughout the brain. They found that the tests of g and executive function that were most affected by prefrontal and parietal damage involved verbal comprehension, working memory, mental flexibility, and attentional control. The researchers argue that the prefrontal parietal network is at an “ideal site” in the brain to support goal-directed
behavior because of its tight connections and broad access to perceptual and motor representations.

They found that a major player in the prefrontal parietal network was the frontal pole, an area at the very front of the prefrontal cortex, just above our eyes. There is evolutionary significance of this brain region. Compared to other species, we don’t have the largest brains; sperm whales and elephants beat us on that front. We don’t even have a larger prefrontal cortex overall. We do, however, have a larger frontal pole and inferior parietal lobe compared to macaques and apes. Intriguingly, a recent analysis of Albert Einstein’s brain shows that although the overall size and shape of his brain weren’t abnormal, he did have unusual frontal and parietal lobes (among other unusual brain regions), with the frontal pole being particularly unusual within his prefrontal cortex.

What are the functions of the frontal pole? Answering that question may help us understand some of the most uniquely human aspects of human intelligence. But the matter isn’t so simple. According to cognitive neuroscientist Paul Burgess and colleagues, the study of the functions of the frontal pole “presents one of the greatest scientific puzzles to cognitive neuroscience.” The frontal pole has been linked to a diverse range of cognitive processes, including fluid reasoning, relational complexity (the number of interdependent items that must be simultaneously considered in working memory), abstract integration during analogical reasoning, semantically distant mapping during creative analogical reasoning, abstract mental flexibility, evaluation of creative ideas, memory retrieval, moral decision making, “reality monitoring” (the ability to judge whether an event was imagined or actually occurred), and metacognition (the ability to think about one’s own cognition).

Because g, reality monitoring, and metacognition are all at least partially distinct sources of human variation, this suggests that the frontal pole serves functions that go beyond fluid reasoning. So what is the common theme of the frontal pole, if there even is one? Accumulating research suggests that the frontal pole sits at the top of a hierarchy in the prefrontal cortex that monitors the current contents of consciousness to make sure the larger goal is maintained, and helps integrate the prior stages of cognitive processing. Surely more research is needed on this fascinating area of the brain, but it is becoming increasingly clear that the frontal pole is a critical team player in the prefrontal parietal network.

Based on their lesion mapping results across all of their tests of g and executive function, Barbey and colleagues put forward an integrative archi-
tecture for general intelligence and executive function critical for novel, goal-directed problem solving. Consistent with the P-FIT theory proposed by Jung and Haier, this architecture involves strong communication between specific areas of the lateral prefrontal cortex and posterior parietal lobe. The critical brain regions involved in this neural architecture, along with the white matter tract binding these regions together into a coordinated network, are shown in Figure 10.4.* Note that although the regions in this prefrontal parietal network are distributed, they aren’t everywhere. As the researchers note, “Despite its distributed nature, the neural substrates of

*Technically, this particular white matter tract is called the superior longitudinal/arcuate fasciculus. Say that ten times real fast.
$g$ and executive function were remarkably circumscribed, concentrated in the core of white matter and comprising a narrow subset of regions.”

This is worth emphasizing. For one thing, $g$ should not be equated with the entire prefrontal cortex. The prefrontal cortex consists of billions of nerve cells linked together by trillions of connections. All areas of the prefrontal cortex are massively connected with each other (and with many other areas of the brain), but $g$ appears to be primarily associated with broad brain activation in the lateral prefrontal cortex, the outermost part of the brain. Barbey and colleagues didn’t measure the self-regulation of emotions, the learning of reward values, daydreaming, creative improvisation, emotional intelligence, or social reasoning. If they had, they would most likely have found additional recruitment in medial regions of the prefrontal cortex (which reside in the center of the brain) along with connections to subcortical areas of the brain associated with emotional functioning (see Chapter 12).

Also, $g$ isn’t equally related to all executive functions. There are a variety of executive functions in humans, including working memory, mental flexibility, and inhibition. While these various executive functions are significantly correlated with each other, they also show diversity in their functions. A striking demonstration of this can be found at the developmental level of analysis. Naomi Friedman and colleagues found that the development of better behavioral restraint across four testing periods (ages 14, 20, 24, and 36 months) was accompanied by the development of worse mental flexibility. Out of the entire suite of executive functions, $g$ appears to be most strongly associated with working memory-related functions.

Nevertheless, the discovery of the prefrontal parietal network gives us insight into the information processing critical for novel, goal-directed problem solving. Barbey and colleagues argue that a major function of the prefrontal parietal network is the manipulation, integration, and control of distributed patterns of neural activity throughout the brain, including lower-level sensory and motor modules. We’ve already discussed some of these prefrontal parietal network functions (such as goal maintenance, attentional control, updating information in working memory, ignoring distractors, relational complexity), but I believe one of the most important functions of the prefrontal parietal network for understanding $g$ is often overlooked.

In his recent book *The Ravenous Brain*, neuroscientist Daniel Bor argues that consciousness is a form of information processing. In particular, Bor argues
that consciousness processes information that is useful and relevant to a particular goal and that captures some pattern in the world. In formulating his cogent argument, Bor links the prefrontal parietal brain network, and its associated working memory and attentional control functions, with consciousness. While attention and consciousness are certainly not the same, Bor argues that attention, as a way of selecting content and attending to stimuli, is a necessary aspect of consciousness.

But Bor doesn’t stop there. He cites numerous studies showing that chunking heavily activates the prefrontal parietal brain network, sometimes more so than working memory. Consider one elegant study conducted by Bor and Adrian Owen. They asked participants to memorize novel verbal and numerical double-digit sequences. Critically, the information was either randomly arranged or structured. For instance, 57 68 79 90 is a structured sequence because each number goes up by 11, but 31 24 89 65 is an unstructured sequence because you can’t readily apply a chunking strategy to organize it more easily in memory. They also had a mnemonic condition in which participants could apply phone extension numbers they memorized earlier. What did they find?

Unsurprisingly, participants performed better when the information was structured compared to unstructured, as it allowed them to reorganize and chunk the information into more efficient forms. Also, consistent with prior research, the prefrontal parietal network (consisting mainly of communication between the dorsolateral prefrontal cortex and the posterior parietal lobe) lit up more brightly during the structured and mnemonic conditions than during the completely unstructured condition where chunking wasn’t possible. Most interestingly, the prefrontal parietal network was consistently most active during the structured trials compared to the unstructured and mnemonic sequences.

This is interesting because the unstructured sequences were more difficult to memorize and placed a higher demand on working memory than the structured trials did. According to Bor, these findings suggest that “the prefrontal parietal network will activate for many complex tasks, but it will be

*Additionally, structured mathematical sequences more robustly activated the prefrontal parietal network compared to a control condition in which participants performed equivalent mental calculations without any chunking component. The same goes for the mnemonic sequences when compared with the same level of memory recall, but with still no chunking aspect. Therefore, it appears that the prefrontal parietal activation wasn’t due just to mental arithmetic or memory recall but was specific to the act of chunking.
most excited when subjects are actively searching for and finding entirely new patterns.” Bor raises the intriguing possibility that chunking is one of the most central functions of the prefrontal parietal network and may have evolved to provide innovative solutions to complex or novel problems.

Even though Bor doesn’t explicitly make the connection, I believe his theory may help us better understand $g$. Perhaps one of the fundamental characteristics of individuals who rank high on $g$ is their ability—when given an explicit goal—to quickly focus on the relevant aspects of the task, ignore the irrelevant aspects, and find efficient means of organizing the information in working memory. This would free up important resources for fluid reasoning. Spearman’s hypothesis was that the best measures of $g$ are those that require grasping relationships, inferring rules, spotting similarities and differences, and “educing” (Latin for “drawing out”) the relevant relations within a complex pattern. Fluid reasoning tasks do seem to be measuring these skills. But fluid reasoning also places a heavy burden on working memory, because you have to conduct all the reasoning in your head with no external aids and often with a timer counting down! Those who have better cognitive strategies for lessening the cognitive load will be at a distinct cognitive advantage in this testing environment. If true, I think these findings allow for a tighter integration between the scientific study of $g$ (which emphasizes cognitive efficiency) and the expertise performance framework (which emphasizes efficient chunking strategies).

I witnessed the importance of cognitive strategies firsthand during a study I conducted for my doctoral dissertation. I measured $g$ by administering three highly correlated tests of cognitive ability: Raven’s, verbal analogical reasoning, and mental rotation. I found that working memory, processing speed, and the deliberate learning of complex associations were each significantly and independently related to $g$. During the interviews at the end of the experiment, some participants told me they tried coming up with elaborate mnemonic strategies to memorize the list of words on the tests of associative learning. At the time, I just filed this away as interesting.

More recently I came across a study by Kiruthiga Nandagopal, Roy Roring, K. Anders Ericsson, and Jeanette Taylor that had twins think aloud during measures of associative learning, working memory, and processing speed—the same skills I measured in my study. They found that performance on all three cognitive tests was heavily influenced by strategies. Most compellingly, differences in strategy use on the associative learning task (which was most amenable to the use of strategies) explained a significant amount of the
genetic influences on performance. Their study is the first to demonstrate that the heritability of performance on cognitive tasks is due, in part, to the use of specific cognitive strategies.

The use of strategies to facilitate performance is consistent with studies linking $g$ with neural efficiency. Cognitive effort consumes a lot of glucose in the brain. As tasks become more automated, we require fewer cognitive resources. A number of studies suggest that people who do well on IQ tests are faster at recruiting the right brain resources for the right task. Consider an earlier study by Haier and colleagues using PET. They had 8 participants practice the video game Tetris for fifty days. At the time of the study, Tetris had just been introduced in the United States and the participants weren’t familiar with it. The researchers found that the highest scorers on the Raven’s matrices test showed the largest decreases in brain activity with practice, particularly in the prefrontal and cingulate cortical areas.

Newer studies using fMRI generally support the notion that people who do well on cognitive tests use fewer brain resources to solve novel and complex problems. But while these findings are generally expressed in terms of “intelligent people” having more “efficient brains,” I think a more precise framing is that people who are “good at fluid reasoning” are “faster at freeing up expensive brain resources through the use of efficient cognitive strategies.” I concede it’s more of a mouthful, but I think it’s more precise.

Importantly, different people can solve the same problem using different strategies. For instance, some people may rely more on their verbal strengths to solve verbal and spatial problems, while others may use a spatial strategy across the board. There is evidence that this is the case. On a test of verbal processing speed, Erik Reichle and colleagues found that the use of verbal strategies produced more brain activation in language-related cortical areas (such as Broca’s area), whereas the use of visual-spatial strategies produced more activation in visual-spatial areas (such as the parietal cortex). Also, those with better verbal working memory skills showed less activation in the language-related regions when using the verbal strategy, and those with better mental rotation skills showed less activation in visual-spatial brain regions when they used the visual-spatial strategy. The researchers concluded

---

*Provided that the task is difficult enough—but not so difficult that even high IQ test scorers (on average) have to recruit all of their brain resources or people with lower IQs (on average) quit out of frustration; I. J. Deary, L. Penke, and W. Johnson, “The Neuroscience of Human Intelligence Differences,” *Nature Reviews Neuroscience* 11 (2010): 201–211.*
that cognitive strategies are useful because they help minimize cognitive workload.

If people can recruit different brain regions to solve the same novel or complex problem, it might not make sense to ask “Where is g in the brain?” A recent study by Rogier Kievit and colleagues illustrates this point. They administered tests of verbal comprehension, perceptual organization, working memory, and processing speed. They also gave participants MRI scans and estimated their level of white matter, gray matter density, and brain volume in eight brain regions of interest that have shown correlations with g in prior studies. They found that people could arrive at a similar g score with a very different neural composition. These findings are important because they suggest that studies that average over many different brains mask the fact that every single person may arrive at the same g score via different neural pathways.

Additionally, Kievit and colleagues found that individual differences in brain structure jointly determined individual differences in g. In other words, g, like consciousness, was an emergent property of the brain, not a single feature. This finding is consistent with a number of other recent studies that suggest that the positive manifold is largely the emergent result of overlap between multiple specialized brain regions. As Cristina Rabaglia (who is currently conducting research on this topic with Gary Marcus and Sean Lane) told me, “to the extent that individual cognitive tasks each require a particular, unique set of neural resources drawn from a common pool that is reused across tasks, individual differences in performance on those tasks will be correlated—even if particular neural resources have prescribed functions.”

Although these findings don’t negate the existence of the positive manifold (those who do well on one IQ test item are statistically more likely to do well on the other items), they do suggest that to measure g, it’s best to measure a diverse range of cognitive skills, not look at any individual’s brain and attempt to predict their g ranking (or IQ score, for that matter).

These findings are also consistent with recent work by Han van der Maas and colleagues. According to their dynamical model of intelligence, g is an emergent property of a number of independent mental processes that beneficially interact and mutually reinforce each other throughout the course of development. This research is exciting because it shifts our level of understanding from individual differences to the development of intellectual skills across an individual’s life span.
In childhood and adolescence, brain maturation is particularly vulnerable to plasticity and change. The relation between brain size and IQ is lower in children than in adults, and IQ is related to brain development in complex ways. One study followed 300 children up to early adulthood. At age 7, the higher IQ children (IQ > 120) tended to have less cortical thickness. Soon after, however, the high-IQ children showed a rapid increase in cortical thickness, overtaking the other children and peaking at age 11 to 12 before slowly declining to about the same level as the others. The researchers concluded, “‘Brainy’ children are not cleverer solely by virtue of having more or less gray matter at any one age. Rather, intelligence is related to dynamic properties of cortical maturation” (678).

To be sure, genes substantially influence the development of brain structure. But the relation between genes and cortical maturation need not be direct. Higher IQ children, due to their prefrontal parietal brain wiring, may act differently in the world and elicit different reactions from others, which in turn sculpts their brains. A number of studies across various domains show the importance of experience on both gray and white matter structures in the brain.

People learning to juggle over the course of three months showed increased gray matter volume in brain regions associated with visuomotor coordination, reaching, and grasping. In fact, brain plasticity in gray matter was observed after only seven days of juggling training! After three months of juggling, changes were also seen in the organization of underlying white matter pathways. Likewise, research among expert taxi drivers has found enlargement in the volume of the posterior hippocampus, an area of the brain crucial to spatial navigation, and mindfulness meditation training has been found to alter cortical representations of interoceptive attention (attention turned inward).

Multiple studies have also found significant brain plasticity among people engaging in music. Expert musicians display greater gray volume matter and cortical thickness in auditory, somatosensory, and motor cortices, with the effects increasing as a function of years of musical practice. There are also important effects of early ability. In one study, skilled string players showed neurological differences in the area of the cortex associated with the representation of their left hand, and these differences were significantly correlated with the age at which they first started playing the instrument.
There even appear to be very specific age effects: Sara Bengtsson, Fredrik Ullén, and colleagues identified several brain regions where white matter was directly related to the amount of piano practice during particular periods in childhood and adolescence. Along similar lines, Krista Hyde and colleagues identified numerous changes in brain structure among 5- and 6-year-olds who engaged in keyboard training thirty minutes every week for a year.

Recent research has also demonstrated the plasticity of brain regions associated with cognitive ability. Hikaru Takeuchi and colleagues found that working memory training resulted in measurable changes in the structural connectivity critical for working memory performance, including areas of the parietal lobe and the anterior part of the body of the corpus callosum. In another recent study, Allyson Mackey, Kirstie Whitaker, and Silvia Bunge found plasticity in the white matter structure of the frontal and parietal lobes after just three months of reasoning training among a sample of participants enrolled in a course to prepare for the Law School Admission Tests (LSAT).

Besides brain changes, there are also behavioral changes. Multiple studies suggest that working memory training programs produce reliable short-term improvements in working memory skills, and there’s accumulating evidence that fluid reasoning training can also result in short-term improvements in fluid reasoning skills. Some of the strongest results have come from Cogmed computer-based training and interactive games.

Transfer effects appear to be limited, however. Training children in working memory improves performance on other working memory tasks, but the evidence is mixed on whether it improves fluid reasoning. Vice versa, fluid reasoning training has been found to improve untrained fluid reasoning performance but doesn’t appear to improve working memory or processing speed. Also, unsurprisingly, considering that $g$ reflects variation in a variety of cognitive skills, there’s little evidence that working memory training changes people’s ranking on $g$.

Clearly, the picture is muddled at the moment, and we will have to wait for more research to clarify it. In the meantime I have a few suggestions that

*Interestingly, some working memory tasks, such as the $n$-back task, do seem to transfer to fluid reasoning more reliably than other measures of working memory, although training on the $n$-back task doesn’t necessarily improve performance on other working memory tasks even if it does improve fluid reasoning!"
might help bring things into focus. I think it’s worth keeping in mind that working memory may serve a domain-independent function for fluid reasoning across multiple forms of content (verbal, nonverbal, mental rotation, and so on), but it surely isn’t emotion-independent. As we’ve seen throughout this book, stress, anxiety, and stereotype threat can significantly impact working memory and can cause a person’s prefrontal parietal brain network to shut down (see Chapter 7). This is probably one of the reasons the most widespread gains in brain training come from programs that simultaneously address multiple aspects of a person, such as traditional martial arts training and enriched school curricula. I believe researchers underestimate the extent to which multiple aspects of development—cognitive, physical, social, and emotional—all feed off each other.

A related consideration is individual differences in response to training. One recent study found that variation in a gene that codes for dopamine transportation was related to improvements in working memory and fluid reasoning in preschool children following training. While the results certainly require replication (the sample size was small, and the effects did not remain significant after correcting for multiple comparisons), the results suggest that dopamine may play an important role for cognitive performance and brain plasticity (see Chapters 6 and 12 for more on the role of dopamine in motivation and cognition).

Another recent study found effects of personality on cognitive training. Conscientious participants were better than other participants on working memory training, but their improvement in working memory didn’t transfer to a measure of fluid reasoning. In contrast, participants scoring higher in neuroticism (a proxy for anxiety) showed lower training scores on a difficult version of a working memory task, but showed more gains on an easier version. It seems that in the easier version, higher levels of emotion may have been an advantage, allowing them to maintain their concentration and vigilance, whereas on the more difficult version of the task they became overwhelmed. Therefore, personal characteristics should be taken into account when considering the effectiveness of cognitive training.

Another thing to keep in mind is that those who need the training the most are the ones who are most likely to benefit. In recent discussions with neuroscientist Silvia Bunge, she noted that most reasoning tests only require the ability to maintain and manipulate a few bits of information in working memory. So training working memory among those who are already at the minimum necessary to solve the task won’t improve reasoning performance.
Individuals with low working memory who have difficulty keeping a few things in mind and integrating, comparing, or sequencing them are more likely to show meaningful improvement in training, because for them working memory serves as a bottleneck. This is why it’s important we teach all people the importance of strategies to relieve the mental burden.

According to Philip Johnson-Laird and his colleagues, people solve syllogisms (a form of deductive reasoning) by constructing a mental model—an internal representation of the premises. The juggling of mental models in your mind places a very heavy burden on working memory. That’s why it’s important to choose your mental models wisely. For instance, Johnson-Laird and Mark Steedman asked people to describe their mental models for the following syllogism:

1. All of the artists are beekeepers.
2. Some of the beekeepers are clever.
3. Are all artists clever?

One participant remarked, “I thought of all the little . . . artists in the room and imagined they all had beekeeper’s hats on.” You may have relied on an altogether different mental model. How you represented the problem, however, strongly determined how well you were able to reason with the information. If you started with an inaccurate representation, you were more likely to overburden your working memory and end up with an inaccurate answer, regardless of the level of syllogistic reasoning you are actually capable of attaining.

People can be taught to improve their reasoning by learning how to draw diagrams to represent a problem. Even blind people can create spatial mental models. When Kenneth Gilhooly and his colleagues presented syllogisms orally, it placed a higher demand on working memory, as participants had to store the premises in their head. But when the syllogisms were presented with all the premises remaining on the projector screen, people performed better because they could unload the premises from their working memory and free up limited resources to construct efficient mental models.

Over the past decade John Sweller and colleagues have designed instructional techniques that relieve working memory burdens on students and increase learning and interest. Drawing on both the expertise and working memory literatures, they match the complexity of learning situations to the
learner, attempting to reduce unnecessary working memory loads that may interfere with reasoning and learning, and optimize cognitive processes most relevant to learning.

Finally, it’s important to consider length of training. In a recent New York Times Op-Ed called “IQ Points for Sale, Cheap,” David Z. Hambrick notes his skepticism that a few hours of working memory training can create long-lasting and meaningful improvements in IQ. He makes an important point that such increases aren’t likely without substantial commitment of resources. Hambrick points to a few kindergarten after-school programs that involve substantial enrichment and support but don’t demonstrate a large increase in IQ scores, and whatever increases are found don’t sustain in the long-term.

If the skills that determine someone’s IQ test score operate by the same developmental principles as any other abilities, then why would we expect them to be radically improved through enrichment programs that may last no more than sixty hours, let alone working memory training that lasts only a few hours? As Michael Howe notes in his book IQ in Question, “If we start by assuming that the skills that contribute to a person’s IQ score have no special or unlearned status, and are acquired by processes that are similar to the ones involved in the acquisition of other kinds of mental expertise, then it makes sense to ask how much time is typically needed in order to gain those mental abilities that are acknowledged to be acquired through learning.”

Howe suggests that the amount of training would be at least as large as the time necessary to acquire high levels of expertise in music, chess, and sports. After all, IQ tests sample an assortment of mental forms expertise. Therefore, to build up the high levels of mental expertise required to do well on IQ tests would require thousands of hours, not just a couple.

This doesn’t mean that there are no biological contributions to the acquisition of cognitive expertise. There are biological contributions to the acquisition of any forms of expertise (see Chapter 11). Some children from a very early age, due to their prefrontal parietal brain wiring, may gain a greater reward from engaging in fluid reasoning and so become more motivated to engage in intellectually demanding activities. All those hours of continual intellectual engagement add up (see Chapter 1). In recent years psychologists—including Robert Sternberg and David Lohman—have begun to view IQ test scores at any single point in time as a measure of developing expertise or ability.

From this perspective, we also shouldn’t be surprised that IQ gains don’t last very long after an intervention is over. We certainly wouldn’t expect
people who stop practicing music or Spanish or chess to maintain the same levels of expertise across their life span. Repeated practice and challenge are crucial.100

But increased IQ test scores may not even be the most practically meaningful outcome of interventions for all children. As we saw in Chapter 8, there are successful interventions that improve multiple aspects of self-regulation—skills that may be more important for people in reaching their own personal goals.

Nevertheless, this shift in conceptualization of $g$ from “mental energy” to various forms of cognitive expertise serves a real practical purpose. Instead of focusing so much on how people differ from one another on tests that place heavy demands on working memory, it shifts the focus to what people can actually achieve if we really give them the chance. A truly potent demonstration of this isn’t even found in the current generation.

The twentieth century witnessed a dramatic increase in IQ scores, as much as 3 points per decade. This phenomenon is dubbed “The Flynn Effect.”101 Not all IQ subtests showed the same level of increase, however. As we saw earlier, all tests of cognitive ability involve at least some fluid reasoning and prior knowledge, but some tests are better measures of fluid reasoning than others. The most dramatic increases in the twenty-first century were found on those that made the strongest demands on fluid reasoning. Tests of vocabulary, short-term memory, and general knowledge showed much smaller increases. But what caused these increases?

In his book What Is Intelligence? Beyond the Flynn Effect, James Flynn offers a plausible account of the rise in fluid reasoning.102 According to Flynn, our ancestors thought very differently about the world. The Industrial Revolution brought to prominence a particular type of thought—scientific abstract thought. People back then didn’t receive the same scientific instruction we do today. Once the Industrial Revolution brought with it a different set of demands, abstract thinking flooded the classroom curriculum, and the average person became much more comfortable at hypothetical thinking and making abstract generalizations.

Today, with more opportunities for educating the scientific mind, more people hold a ticket that gives them the chance to properly develop their abstract reasoning ability. According to Flynn, many people may have been smart back in the day, and may have been capable of answering the IQ items
requiring abstract generalizations, but may have found the items so foreign and absurd that they may have answered the test questions with a certain habit of mind that wouldn’t have earned them a high score.

Let’s take an example of an actual IQ testing session from the early part of the twentieth century:

**IQ TEST EXAMINER:** All bears are white where there is always snow; in Novaya Zemlya there is always snow; what color are the bears there?

**SOVIET UNION PEASANT:** I have seen only black bears and I do not talk of what I have not seen.

**IQ TEST EXAMINER:** But what do my words imply?

**SOVIET UNION PEASANT:** If a person has not been there he can not say anything on the basis of words. If a man was 60 or 80 and had seen a white bear there and told me about it, he could be believed.

This peasant, hailing from a remote area of the Soviet Union, was interviewed by the psychologist Alexander Luria. It turns out that this peasant’s way of thinking was quite common back then. This type of thinking involves reasoning based on personal experience and reference to the concrete, functional use of objects. As Flynn notes, “You are just not supposed to be preoccupied with how we use something or how much good it does you to possess it.”

A recent paper supports the notion that abstract reasoning comes more naturally to people who grow up in scientifically advanced cultures. Mark Fox and Ainsley Mitchum conducted an analysis of how test takers of different cohorts solve problems on the Raven’s test. As we saw earlier, performance on the Raven’s test involves high levels of relational abstraction. The important aspect of the task isn’t the figures themselves, but how they relate to each other. Multiple attributes of the figures have to be integrated in working memory to spot the pattern.

Fox and Mitchum classified all of the Raven’s items in terms of their amount of relational abstraction and found that individuals tested in 1961 were less likely to map objects at higher levels of relational abstraction than individuals tested in 2006. It’s not that those tested in 1961 were incapable of solving the problems; they were just not as accustomed to this way of thinking. When given the Raven’s test, they became so fixated on the literal appearance
of the figures, it didn’t occur to them to apply more abstract strategies. The researchers argue that today’s test-takers have a qualitatively different strategy to responding to test items: they know to search for the analogical relations among information when the initial interpretations do not automatically lend themselves to such a strategy.

The Flynn Effect raises a number of important issues. For one, it supports the findings from the behavioral and biological levels of analysis that \( g \) consists of multiple, interacting, cognitive mechanisms. The Flynn Effect clearly shows that the various cognitive abilities that form \( g \) can come apart across generations.

The Flynn Effect also highlights the important influence of cultural priorities on the development of specific forms of cognitive expertise. As an analogy, consider sports. There’s evidence that many of today’s world records are at least 50 percent superior to those of a century ago.\(^{105}\) Accomplishments for such events as the marathon and swimming have increased so much that the gold medal winners of the early days of the Olympics would barely meet today’s requirements for entrance into high school swimming teams.\(^{106}\) Officials almost banned the double somersault in dives during the Fourth Olympic Games in 1908 because they believed they were too dangerous and impossible to achieve. Today, triple somersault dives are part of the standard repertoire of competitive divers! But if you go to any high school gym class within a generation and measure performance across various tests of physical fitness, a general physical fitness factor will most likely emerge.\(^{107}\)

Likewise, within a generation, people who tend to do well on one test of cognitive ability do tend to do well on other tests causing a cognitive \( g \) factor to emerge. But across generations, tests placing heavy demands on fluid reasoning have increased the most, because culture has placed greater emphasis on this type of thinking, compared to short-term memory, vocabulary, and general knowledge—all tests that have not increased nearly as much throughout the generations. According to Flynn, we have to face a different set of problems today that were unheard of to our ancestors. Today, we take such thought for granted. But that doesn’t necessarily make us smarter than prior generations.

Flynn’s explanation of the Flynn Effect is just one of many “social multipliers” that have been proposed to explain the rise in fluid reasoning. According to Dickens and Flynn, a social multiplier is any societal factor that
can multiply through the generations, causing large effects.\textsuperscript{108} In addition to increased scientific spectacles, other factors undoubtedly played a role, including increased nutrition, increased literacy, increased test familiarity, video games, complexity of TV shows and movie plotlines, modernization, decreased prevalence of infectious disease and parasites, and more.\textsuperscript{109}

Surely a combination of factors contributed to the rise in fluid reasoning. Regardless of the specific causes, the Flynn Effect serves as a reminder that when we give people more opportunities to prosper, more people do prosper.

\begin{quotation}
So have we tamed $g$? Even though many issues remain unresolved and require further research, I think we can at least wrap our head around the phenomenon. In Thomson’s time, they didn’t yet have a language for talking about cognitive processes. That’s why Thomson had to rely on his abstract, theoretical notion of “bonds.” But modern cognitive science suggests that the truth is somewhere between Spearman’s and Thomson’s extreme positions. There appears to be a manageable number of cognitive processes available for use in any given cognitive task, with each cognitive test engaging a different—but overlapping—subset of human cognition. These processes are related to each other, and to $g$, but there is no reason to believe they are the same mental “energy.”

An important implication of these findings is that although $g$ is a real statistical phenomenon that emerges when we examine performance on a diverse range of cognitive ability tests across a large number of diverse individuals, $g$ does not exist within any single individual. IQ test scores are probably best thought of as the summary of a restricted but important range of cognitive forms of expertise, involving fluid reasoning, abstract integration, working memory, short-term memory, cognitive strategies, mental rotation, verbal comprehension, vocabulary, and processing speed. Of course, which particular subset of these skills is being measured depends on the IQ test battery (see Chapters 3 and 4).

Importantly, not all test items equally measure $g$. Multiple threads of research suggest that when a large and diverse-enough battery of cognitive tests is administered, the very best measures of $g$ will tend to involve a high amount of on-the-spot manipulation, integration, and control of cognitive processes in working memory (although all indicators of $g$ require these
skills to some degree). When tasks place such a heavy demand on working memory, the prefrontal parietal regions of the brain play an important role in helping us find efficient cognitive strategies to chunk the material and reduce the mental burden. People can draw on different strategies, however, to achieve the same $g$ score, and there is accumulating evidence that the skills that are most strongly related to $g$—working memory and fluid reasoning—can be trained. Although evidence for transfer is mixed, I’ve highlighted the need for long-term cognitive training interventions to simultaneously target multiple cognitive mechanisms as well as social and emotional functioning.

Though $g$ describes the common variance between individuals on tests of cognitive ability, $g$ doesn’t capture all the richness of human intelligence. In fact, there is accumulating evidence that at the upper end of IQ scores the positive manifold starts to break down, and scattered cognitive profiles become more prominent. In 1925 Spearman published a paper in which he compared the performance of 78 “normal” and 22 “defective” children on a battery of twelve cognitive tests. He found that the positive correlations among the tests were much higher among the “defective” group. He labeled this discovery the Law of Diminishing Returns. Modern-day researchers have confirmed Spearman’s initial findings, but refer to it as the differentiation hypothesis, since cognitive abilities tend to become more differentiated as $g$ increases.

On this point, we may have finally found some common ground. Let’s end our long survey of $g$ by considering two viewpoints from two intelligence researchers who come from vastly different time periods and appear to be completely at odds in their conceptualization of human intelligence: Charles Spearman and Howard Gardner.

First, here’s Spearman in his 1925 paper:

> At the extreme ends of the distribution will lie a very small number of performances for which the person is, on one side a genius, and on the other an idiot. Every normal man, woman, and child is, then, a genius at something as well as an idiot at something. It remains to discover what—at any rate in respect of the genius. (439)

Now here’s Gardner, as quoted in a 1995 interview with Daniel Goleman:

> The time has come to broaden our notion of the spectrum of talents. The single most important contribution education can make to a child’s
development is to help him towards a field where his talents best suit him... We should spend less time ranking children and more time helping them to identify their natural competencies and gifts, and cultivate those. There are hundreds and hundreds of ways to succeed, and many, many different abilities that will help us get there.¹¹²

Maybe they aren’t all that different after all.