CONTENTS

CHAPTER 1

What Lies in the Gap between Knowledge and Action?

There is a lot of mystery in a cup of coffee—not just in the molecular structure of the drink or in the chemical interactions needed for that morning fix or even in the origin of those D-grade beans, though each of these surely contains mysteries in its own right. A cup of coffee is mysterious because scientists don't really understand how it got there. Someone made the coffee, yes, but we still don't have a satisfying explanation of how that person's brain successfully orchestrated the steps needed to make that coffee. When we set a goal, like making coffee, how does our brain plan and execute the particular actions we take to achieve it? In other words, how do we get things done?

Questions like these fascinate me because they lie close to the heart of what it means to be human. Our species has a uniquely powerful capacity to think, plan, and act in productive, often ingenious, ways. Moreover, our mental apparatus for doing things is somehow general purpose; that is, we are able to get things done even if we have never been in that exact situation with that exact goal before. This ability to get things done drives our common notions of intelligence and personality. We hold in high esteem those who achieve ambitious goals, and we pity those who struggle to achieve any. Yet, the exceptional achievements of the Olympic athlete or the brilliant mathematician are not really what this book is about. This book is about that cup of coffee. It is about what you, the reader, did today. Because what you accomplish on your most unremarkable day, no other species can rival, and no robot yet built can emulate. And how you did it remains an enduring mystery that science has only started to unravel.

To illustrate, let's consider what you might have done this morning. The very first thing you do is decide that your plan to wake up at 5:30 a.m. for a jog was a bit ambitious, so you hit the snooze button on your smartphone. After doing that a couple more times, you finally get up and walk to the kitchen to prepare a cup of coffee, perhaps thinking about that meeting you'll have later as you walk. You end up in the kitchen on autopilot. Before starting your familiar coffee-making routine, you throw a bagel in the toaster. All at once, you remember your intention to email your sister. You are thinking of buying a house, and you have been meaning to ask her if her mortgage broker was any good. You make a mental note to send that email at your next opportunity. Or maybe you are the sort of person who stops whatever you are doing to pull out your phone and send your message. You know if you don't send it now, you won't remember later. When your coffee is done brewing, you turn to locate a clean mug and then also grab a plate for the bagel while you are at the cabinet. You decide to skip the sugar in your coffee today; surely that makes up for the jog you missed. You hear shouting and your children running down the hall toward you arguing loudly. As you watch the door and await the oncoming storm, your fingers idly nudge your coffee mug away from the counter's edge.

Our routine mental life is simply a marvel of goal management. Catch us at any given moment, and our head is bustling with goals, plans, strategies, and tactics. Goals are coming and going throughout our waking life, and often more than one at a time. Our goals can range from the abstract and open-ended, like buying a house, to the immediate and trifling, like looking for the cream. We frequently reevaluate our goals based on changing circumstances, desires, or maybe an estimate of our own limitations (will we remember to send that email later?). Some goals intrude unpredictably or unwanted. For example, no one plans to wash a mustard stain out of their shirt but we can make room. It is fair to say that the course and tenor of everyday human life is largely defined by our goals and the various actions we undertake to achieve them.

The human brain has the remarkable ability to manage the buzz and hum of all these goals, in order to plan and execute efficient courses of action. Indeed, our brains are so good at this that most of us view the routine of everyday life as just that: routine. No one ever produced a summer blockbuster about a typical morning making coffee. We mostly take our ability to get things done for granted and only notice it on those rare occasions when we struggle or fail. Yet, this ability is actually quite singular and marvelous, and also, unfortunately, quite fragile.

The brain requires its own elaborate class of neural mechanisms devoted to generating plans, keeping track of them, and influencing a cascade of brain states that can link our goals with the correct actions. Scientists refer to these mechanisms and the processes they support as *cognitive control* or *executive function*. Though there are some differences in usage and sense between these terms, they generally refer to the same class of mental function. For consistency, I will use the term cognitive control, only because this is the term currently employed by most cognitive neuroscientists, but I am not distinguishing between these labels.

Regardless of what label it goes by, however, cognitive control has been remarkably difficult to define for scientists and lay people alike. As we will see, some of its slipperiness results from our lack of intuition about cognitive control the way we might have about a memory, a percept, or a movement. Rather, cognitive control processes live in the murky spaces between knowledge and action, influencing the translation from the first to the second while not being either one. Yet, cognitive control is a real class of function, separate from knowledge and action, and is supported by its own systems in the brain. Some of the best evidence we have for this comes from the cases of people who have lost cognitive control function owing to brain disease or disorder. In observing these patients we recognize just how devastating the loss of cognitive control can be to the routine course of our lives and even to our image of ourselves as effective agents in our world.

The Functionless Cortex

Cases of brain disease or damage demonstrate the vulnerability of cognitive control, as well as its necessity for success in everyday life. Control function is fragile, as it suffers some degree of loss across many, if not most, neurological and psychiatric disorders, ranging from stroke to Parkinson's disease to Alzheimer's disease to autism. In general, the loss of cognitive control most conventionally arises when these diseases and disorders affect the frontal lobes, and specifically, the *prefrontal cortex*, or PFC (Figure 1.1).

These days it is widely accepted that the prefrontal cortex is crucial for our highest mental functions, including cognitive control. Indeed, this idea has captured the popular imagination. Movies and television police dramas are fond of blaming frontal lobe damage for everything from irrational, violent behavior to personality change. One health website advertises "10 Exercises for your prefrontal cortex"[1] to help you "focus and think" by doing dubious things like "Learn to Juggle."

In our modern context, then, it might be surprising to learn that at one time neuroscientists wondered if the prefrontal cortex might actually be functionless or, at the very least, so auxiliary in function as to be expendable. In the late nineteenth and early twentieth centuries, neuroscientists started precisely damaging or "lesioning" parts of the brain in animal experiments to discover what function they served. If an animal lost a particular function after a lesion, then it could be inferred that the part of the brain that had been lesioned was necessary for that function. In parallel, clinicians who surgically resected parts of the human brain in the course of a treatment or who saw patients with brain damage from stroke, head injury, or other causes, would try to draw similar inferences from the resulting loss of function. Oddly, however, when the prefrontal cortex was lesioned, these early investigators failed to observe the kinds of dramatic changes one would expect from the loss of higher cognitive function. Indeed, it was surprisingly difficult to define a specific function that was lost at all.

Patients, particularly those with damage affecting only the prefrontal cortex, demonstrated normal sensory and motor function on a neurological exam. They could speak with their doctor in a knowledgeable and articulate way about their past experiences or the topics with which they were familiar, and they often performed quite well on the types of bedside or clinical neuropsychological tests sensitive to mental decline in other patient groups. In his landmark 1890 psychology text, *The*

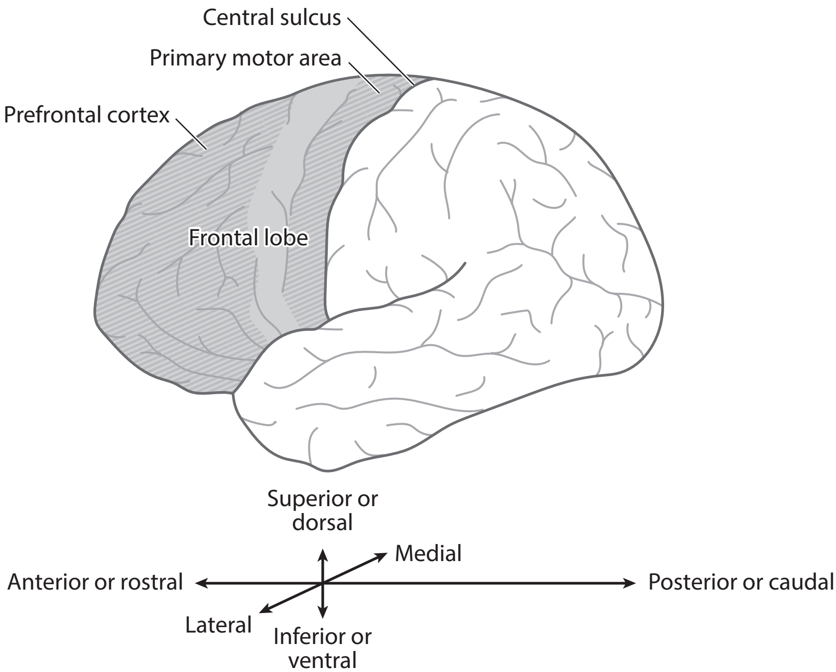

FIGURE 1.1. Drawing of the left lateral surface of the human brain showing the location of the frontal lobe and prefrontal cortex. The compass rose at the bottom provides the direction indicated by relative anatomical terms. For example, the terms *anterior* or *rostral* both mean toward the front of the head, and *dorsal* means toward the top.

Principles of Psychology, Harvard psychologist William James spent a chapter reviewing the known functions of the brain at the time, and he captured the perplexed state of the field concerning the prefrontal cortex.[2] He called the frontal lobes a "puzzle" and stated, "The anterior frontal lobes, for example, so far as is yet known, have no definite functions. [. . .] neither stimulation nor excision of the prefrontal lobes produces any symptoms whatever."

The puzzle of the frontal lobe led to a half-century debate among neuroscientists regarding its function. Many thought it might serve no crucial intellectual function at all or, at most, might serve an auxiliary or supporting function. In 1945, legendary neuroscientist Donald Hebb even described a patient who seemed to improve after surgical removal of frontal tissue.[3] This patient's frontal lobe had been damaged in a sawmill accident. According to Hebb, the patient went from impulsive and unreliable before the surgery to a model citizen after, with maybe only an odd tendency to change jobs every couple of months.[4]

Sadly, it was in part this ambivalence toward prefrontal function that set the stage for the widespread use of frontal lobotomy as a treatment for psychological disorder in the 1950s. After all, if the prefrontal cortex was superfluous, then the subtle cost of its loss was surely outweighed by the potential benefit of relieving major psychiatric symptoms. However, the frontal lobes are not expendable, and tragically, only after some 40,000 estimated lobotomies in the United States alone, and the individual suffering that attended them, was the practice stopped. Outside of the clinic, a different story was evident in the lives of even the earliest cases of prefrontal damage.

In 1928, a 43-year-old woman was referred to the Montreal Neurological Institute because of recurring seizures she had periodically experienced since about the age of 20. She reported to her doctors that she was "mentally not up to her own standard in looking after household arrangements." Among her doctors was Wilder Penfield, the brilliant pioneering neurosurgeon. Penfield had revolutionized treatment for epilepsy with his "Montreal Method" of electrically stimulating an awake patient's brain during surgery, providing a means of locating seizure foci to remove, as well as functional cortex to avoid. However, this occasion and this patient were far from usual. This patient was Penfield's only sister.

Upon removing her cranium during surgery, her doctors discovered a large tumor located in the right frontal lobe. The tumor had infiltrated a wide area, and to excise it, the surgeon had to remove almost the entire right frontal lobe, up to about a centimeter before the motor cortex. Penfield's sister was discharged following recovery from the surgery, and she tried to return to her 1920s lifestyle as a wife and mother of six. Tragically, due to regrowth of the tumor, Penfield's sister died about two years after first being seen at the MNI.

Though conflicted because of his deep personal involvement, Penfield decided to report her case in the medical literature.[5] He felt that the details of her case might be uniquely important to posterity.

One of the early obstacles to understanding the kinds of problems that patients were experiencing following frontal lobe damage was a lack of baseline reference. Patients would be referred to doctors when there was already a problem, perhaps after years of tumor growth, for example, and so observations of the patient could not easily be compared with their behavior prior to the brain disorder. In this case, however, Penfield had known his sister his entire life. He was able to comment in personal and sometimes intimate prose about the changes he saw. The picture he painted was of a woman who retained a pleasant disposition along with much of her core intellectual function, yet who became lost in the everyday activities she had easily performed years earlier.

Even following removal of a substantial amount of frontal tissue, Penfield's sister impressed the doctors attending her with her manner and courtesy. Dr. Colin Russel, who had examined her, noted the following interaction to Penfield in a letter:

> [S]he expressed her appreciation of what she considered my kindness in giving up my time [to see her], so perfectly and with so much courtesy that it was really very impressive. She said that she had felt so afraid of causing you distress by making an exhibition of herself and that I had helped her. When I remarked that the only exhibition I had seen was one of the best exhibitions of courage that it had been my fortune to witness, she expressed her gratitude so nicely that one could not help wondering how much the frontal lobe had to do with the higher association processes.

Professor Russel was likewise impressed by her intellectual function, such as the quality of her personal memories or her ability to discuss with him books she was reading. None of this sounded like a person who had experienced much loss of function following such a large removal of frontal tissue.

Her own take on matters was quite different, however. She described in one of her letters to Penfield that "Dr. Taylor asked me if I felt that mental activity was improving, and I said 'Yes,' but it seems as though each time I feel encouraged that way, *I do a series of very stupid things* [emphasis mine]." The intellectual loss she felt was not experienced as an overt loss of knowledge or an inability to function at all. Rather, in her words, she was frustrated by frequent failures to perform the complex but routine activities that are the fabric of everyday life.

For example, on one occasion, she invited Penfield and four other guests over for dinner. Penfield described her distress upon his arrival.

> She looked forward to it with pleasure and had the whole day for preparation. This was a thing she could have done with ease ten years before. When the appointed hour arrived she was in the kitchen, the food was all there, one or two things were on the stove, but the salad was not ready, the meat had not been started and she was distressed and confused by her long continued effort alone. It seemed evident that she would never be able to get everything ready at once.

She had managed to get all the food out and maybe even had some items bubbling away. But the full meal, as a concerted act, would never have been completed had she been left to do so independently. As anyone knows who has prepared a large meal for several people, you don't just cook each dish on its own and then set it aside to work on the next. Rather, you coordinate multiple preparations at once to produce a coherent whole. You pause headway on one dish to move to another. You manage bottlenecks in the kitchen. You make progress checks. And the various components of the meal must all be finished at roughly the same time, with none too hot and none too cold. This is a challenge even for the experienced amateur cook. It is insurmountable for the patient without cognitive control, as it was for Penfield's sister.

Injury to our cognitive control system results in real, qualitative loss. It becomes challenging to get things done. For Penfield's sister, making a meal for a dinner party was as impenetrable as forming a new memory is for an amnesic patient or speaking is for an aphasic patient. Yet, she described this experience not as a performance issue but as an intellectual loss, as feeling "stupid."

It is important to clarify that the mental decline she expressed is not necessarily related to a loss of core knowledge itself, in the colloquial sense of the word, as in our stored collection of facts, experiences, beliefs, and our conscious understanding of the world about us. Rather, problems with cognitive control can be present in patients whose knowledge base appears entirely normal or even better than normal.

This paradox was dramatically demonstrated in the case of another patient, called EVR, who was reported by Paul Eslinger and the neurologist and neuroscientist Antonio Damasio in 1986.[6] EVR was a 44-year-old accountant who had undergone surgery to remove a frontal tumor around a decade prior to Damasio's report and with the surgery had lost a large portion of his medial and ventral frontal lobes. On multiple tests he was given in the clinic, EVR consistently displayed a superior intellect, even a decade after his surgery. He performed at the highest levels (97–99th percentile) on the Wechsler Adult Intelligence Scale, or WAIS, a widely used intelligence test. In conversation, he drew on a rich knowledge base that he exhibited freely, for instance, offering depth and insight about Reagan's neofederalist political philosophy (this was the '80s, after all). He even performed at normal levels on conventional neuropsychological tests intended to detect frontal lobe dysfunction.

While EVR was acing all the doctors' tests, however, his life outside the clinic was a mess. Prior to his brain damage, EVR had been a respectable married father of two, was active in his local church, and had risen to the level of comptroller at a construction firm. But as the tumor grew, compressing his frontal lobes, and following the surgery to remove it, he had undergone a transformation.

EVR had returned to work after his surgery but eventually lost this job and several subsequent ones because he was repeatedly late and disorganized. He lost his life savings in a risky business scheme with a questionable partner. His wife of seventeen years divorced him and took their two children with her. He remarried, divorced, and remarried again. At the time of Damasio's report, EVR was living with his parents and concocting wild money-making schemes, which he never saw through.

In his day-to-day life, EVR was similarly ineffective. For example, he might take two hours trying to get ready in the morning. And he could spend whole days just concerned with washing his hair. A plan to go to out in the evening would never come to fruition. He would initiate, but ultimately abort, multiple visits to candidate restaurants, or read and repeatedly reread their menus. EVR could express a larger goal, even a rough plan for the future, like eating out at a restaurant, but he could not plan the actions to realize that goal. It was the organization and management of the shorter- to intermediate-term goals that eluded him. EVR found himself stuck—with an image of an endpoint but bewildered by the options, overwhelmed by the paths that he might take to get there, or lacking the initiative to start culling them.

We are perhaps starting to understand why neurologists and neuroscientists had a hard time pinpointing what function had been lost with frontal lobe damage. Like EVR or Penfield's sister, frontal patients were fluent in conversation, knowledgeable, and competent in their moment-to-moment interactions. Further, many of the problems these patients exhibited in their lives might be seen as annoying personality traits in lots of perfectly healthy people. In his review of frontal lobe case studies, behavioral neurologist Frank Benson compiled several descriptive terms commonly used by neurologists in describing frontal cases.[7] His table of "Frontal Lobishness," reproduced as Box 1.1, includes such traits as puerility, boastfulness, or lewd conversation, which might also be characteristic of that annoying person who corners you at a party. The difference is that these traits manifest consistently and aberrantly with frontal lobe disorder and often reflect a change in that person's personality.

As with personality traits, it was hard to distinguish which adverse life events or failures of enterprise were precipitated by brain damage versus which might just be bad luck or follow from some eccentricity or intemperance of personality. Making foolish business decisions, lacking punctuality, changing jobs frequently, or being married and divorced multiple times are certainly not, in and of themselves, diagnostic signs of frontal lobe disorder. However, causes of these life events and their exaggerated frequency following damage to the frontal lobe suggested that one or more functions were being compromised in these patients that made them unusually susceptible to this pattern.

Modern clinical studies confirm in large groups of patients what these case studies illustrated in their anecdotes. On average, frontal patients in these studies exhibit worse job and scholastic performance

Box 1.1. Benson's Manifestations of Prefrontal Damage (Frontal Lobishness)

Tastelessness
Poorly restrained behavior
Decreased social concern
Jocularity
Facetiousness
Witzelsucht [inappropriate joking]
Moria [foolish euphoria, not place in Middle Earth]
Boastfulness
Grandiosity
Decreased initiative
Decreased attentiveness
Forgetfulness
Poor memory
Indifference
Apathy
Shallow effect [*sic*]
Lack of spontaneity
Abulia [slowing of mind, motor, and speech]
Asthenia [lack of energy]
Akinesia [loss of motor spontaneity]
Deterioration of work quality
Depression
Morose discontent
Restlessness
Delusions:
Grandiosity (strength, wealth, intelligence)
Nihilism
Paranoia
Hypochondriasis

Irritability
Disinhibition
Coarseness
Hyperkinetic
Hypokinetic
Flare with anger
Puerile (silly) attitude
Disinhibition of social graces
Inappropriate sexual advances
Sexual exhibitionism
Lewd conversation
Erotic behavior
Euphoria
Poor planning ability
Diminished concern for the future
Capriciousness
Loss of abstract attitude
Loss of esthetic sense
Impulsiveness
Distractibility
Stimulus bound
Concreteness
Perseveration

Adapted from Table 11.1 in D. F. Benson, *The Neurology of Thinking*. (New York: Oxford University Press.)

and are unable to manage their households or themselves. They often get into financial or legal trouble, and many end up hospitalized or in care of others because of their inability to complete the basic tasks of daily life.[8]

Nonetheless, these problems in patients' everyday lives are often missed by scientists and clinicians and the measures they have available to assess them. Indeed, there is still no widely accepted gold standard measure of cognitive control.[9] Some studies have found that the most widely used measures we have explain as little as 18%–20% of the variation among patients in the problems they experience in their everyday lives.[10]

Why are these tests such poor predictors of real world performance? There are several reasons, but one clear problem is that most laboratory tests are too simple. They lack the open-ended complexity we all confront in planning and executing actions in the real world. A 1991 study by neuropsychologists Tim Shallice and Paul Burgess put this disconnect into sharp focus.[11] Three patients with damage to the frontal lobes from traumatic head injury were included in the study, along with a group of healthy people as controls. All three patients performed well on standardized tests of intelligence and cognitive control. However, all three patients also exhibited problems with cognitive control in their everyday lives. For example, one of the patients had at one point excused himself from a therapy session to get a cup of coffee, disappeared, and was later found on the golf course.

The patients and controls were asked to complete a set of errands by themselves around London. They were given some money to budget and various tasks to perform, like buying a loaf of bread or finding what yesterday's weather had been. The study asked not only whether the patients could do these errands but also how efficiently and correctly they would go about conducting them. For instance, you or I might choose a shop where we can buy two items on the same visit rather than going to two different stores. The investigators hypothesized that the patients would have trouble with this kind of efficient planning. To test their hypothesis, they developed a sophisticated analysis scheme by which they could assign a numeric score not just to how many tasks were completed but also to how efficiently they were carried out and how many rules were broken along the way.

In the end, however, the researchers probably didn't even need their fancy analysis scheme to see the outcome. Two of the three patients barely completed half of the eight tasks they were given. Controls barely missed one. The single patient who managed to complete all the errands did so inefficiently and broke several rules along the way. For example, this patient successfully found a previous day's newspaper to determine the weather, but he was then chased down by the store clerk for failing to pay for it.

Perhaps the patients just didn't know what they were supposed to do and that's why they failed to complete all those tasks? Nope. Shallice and Burgess included a score for failures of interpretation, and none of the patients differed from controls in these types of errors. They knew what their goals were. Rather, their failures were due to inefficiency, rule breaking, or failing to reach those goals.

To summarize, then, damage to the brain's cognitive control system can result in a deficit in efficiently and cleanly getting things done. This deficit can be as dense and as devastating as any other loss of function seen in other patient groups. The tricky part for the clinician and scientist is that cognitive control is built to help us achieve our goals in the complexity of the world outside the laboratory. This is its niche, and so it is in this setting that control is needed and thus where deficits are consistently observed.

The Gap between Knowledge and Action

Why do these frontal lobe patients fail at the basic tasks of everyday life? What has been lost? Answering these questions is difficult. There is likely no such thing as a single, uniform "dysexecutive syndrome," or loss of cognitive control as a single whole. Rather, as we shall discuss over the course of this book, cognitive control is a complex system with lots of moving parts, a fault in any one of which can affect our ability to get things done in numerous ways. Thus, two patients might show the same failures on a task for entirely different reasons. However, whatever the particular reason, these patients have in common an inability to connect knowledge and action. Of course, this statement implies there is something to lose that is neither knowledge nor action itself but, rather, is a special class of thing that lies between these two endpoints. Is there any evidence that such a thing exists?

First, it is clear that knowing is not sufficient to acting appropriately. We often experience this disconnect when we try to communicate our ideas to others. We know what we want to say or what we'd like to write, but we just can't find the right words to get the point across. Language is probably not exceptional in this regard. We experience a similar disconnect between more general goals and intentions and the actions that realize them. For example, I have a notion to make that recipe for parsnip soup I saw online, and a fairly firm idea of what it will look and taste like if I do, but I have some serious mental work left to do to actually make one.

Thus, even with the best-formed image of the world, the clearest awareness of a rule for action, and the most urgent desire to achieve an outcome, the brain still requires a way to implement that knowledge. There must be a means to translate an abstract goal into the intricate, moment-to-moment sequencing of behavior.

It follows, then, that intact knowledge and intention are never enough to ensure intact action. Indeed, frontal patients will often be able to state the rules for a task explicitly, and yet, under certain circumstances, they will be unable to follow those very rules.

The neuropsychologist Brenda Milner first noted this paradox in a 1964 study of 71 frontal patients, citing what she termed a "curious dissociation" between her patients' ability to verbalize the rules for the tests she was administering and their inability to follow those rules.[12] Her patients were performing the Wisconsin Card Sorting Test. In this test, the patient sorts cards with some number of colored shapes on them into piles based on the shape, color, or number of shapes printed on the card. So, for example, a patient sorting according to color would make separate piles for red, blue, or yellow shapes, without regard to the identity of the shapes or their number.

The patient is not told the correct sorting rule, and all three of these features are printed on every card. So, the patient has to identify the appropriate sorting rule based on feedback from the tester. For example, if the rule is to sort by shape, and the patient starts sorting by color, the experimenter tells them "incorrect." The patient then shifts to sort by shape, to which the tester says "correct." If they learn, the patient will then continue to sort by shape. Typically, frontal patients are able to figure out the first sorting rule just as quickly as healthy people. However, once the patient has figured out the rule and sorted for a while, the tester changes the rule without telling the patient. So, quite unexpectedly, the patient starts hearing "incorrect" from the tester when they sort based on the rule that had been working until that point.

Healthy people adjust to this change fairly quickly and shift their sorting to find the new rule within a few cards or so. By contrast, frontal patients continue to sort based on the old rule, even while being told "incorrect" over and over. The neuropsychologist's jargon for this inability to stop a previously valid but now obsolete behavior is *perseveration*. What Milner noticed, however, is that some patients who perseverate would also comment with growing frustration on their accumulating errors, all the while continuing to follow the old rule again and again. The patients knew the rule for the task, stating spontaneously, "it has to be the color, the form, or the number." They knew that they were taking the wrong actions. They even knew the tester was going to tell them "incorrect." Yet, they were simply unable to use this knowledge to guide how they behaved: to stop sorting based on the old rule.

In his rich set of short case examples, Frank Benson provided an anecdote that illustrates this phenomenon in more everyday behavior of a patient with frontal damage:

> While being evaluated for the presence of diabetes insipidus, the patient was instructed, "Don't drink any water; don't go near the water fountain." Within a few minutes he was observed having a drink at the water fountain. When asked by the examiner what he had just been told, he immediately replied: "Don't drink any water; don't go near the water fountain."[13]

Of course, one doesn't need to be a frontal patient to see this disconnect. How many parents have experienced the following interaction with their toddler?

"Maria, please don't touch that outlet."

"Okay, Mama." Touches outlet.

"Maria! What did I just say?"

"Don't touch the outlet."

Sound familiar? We commonly interpret these interactions as disobedience or an assertion of independence on the part of a child. Though this might sometimes be the case, this interpretation implicitly assumes that if the child knows the rule, to the point of being able to state it, then she should be able to follow that rule if she wants to. The knowledge is there, so just follow the rule, thinks the exasperated parent. However, a child's brain is still developing, and on some occasions, they may not lack the knowledge of a rule or even the will to follow it. Rather, they might lack the mature cognitive control processes needed to implement it.

We routinely overlook this gap when interpreting our own or others' apparent failures to get something done. "If only Uncle Joe knew how unhealthy soft drinks are, he would stop drinking three a day. Let's email him another article about diabetes." It might be that Uncle Joe somehow missed the public education campaign regarding the health risks of sugar and obesity, but it may also be that Uncle Joe has not been able to structure his life to act on this knowledge in a sustainable way. In sum, knowing may indeed be half the battle, but the other half is evidently not to be trifled with.

On the other end, cognitive control is also sometimes confused with action itself. But, as with knowledge, intact cognitive control is not required to simply execute intact actions, even complex ones. Consider the following bedside test used by neurologists to identify a potential frontal lobe disorder.[14] A pair of glasses is placed in front of the patient. The neurologist then makes a simple motion to her face, as though putting on a pair of glasses. Upon seeing this, the frontal patient will take the glasses from the table and put them on. This is termed *imitation behavior*, and it demonstrates that the action of picking up and putting on a pair of glasses is intact. It is held somewhere in the brain as a precompiled routine, and it can be elicited given the right trigger in the environment. Once triggered, the action can arise without regard to the broader situation. The patient can't stop the action from being triggered by the neurologist's gesture just because those glasses might not belong to them, or they don't need them, or, indeed, because they are already wearing their own pair.

Psychologists sometimes refer to these types of actions as "automatic." They are the actions we can do without thinking about them. They are fast, effortless, and obligatory once triggered. Akin to a habit, they are often performed regardless of our goals, plans, or the broader context of our behavior. Thus, given the right trigger, we can occasionally perform automatic actions without meaning to do so. If you'd like to see an example of this, just search for "Senator Hatch removes glasses that aren't there" on your web browser.

Automatic actions are not limited to simple behaviors like putting on a pair of glasses. In a series of cases, neurologist F. Lhermitte documented several striking examples of what he called "utilization behavior" and "environmental dependency syndrome."[15] Lhermitte's patients performed complex and lengthy action routines that were triggered by objects or cues in the environment, even when these were entirely inappropriate given the broader circumstances.

For example, one patient was brought to Dr. Lhermitte's house. On entering a bedroom and seeing the sheets pulled back on the bed, the patient proceeded to disrobe down to his underwear and get in the bed. In this case, the patient had no trouble executing the complex sequence of actions needed to get ready for bed. However, doing so was also obviously inappropriate merely because he was in a bedroom and saw a bed. There are some strong social norms and conditions that must be satisfied before we can just hop into someone else's bed, much less the bed of our neurologist. To be clear, however, this patient was not deluded about where he was or the nature of the broader circumstances or the correct rules for behavior. If you asked him, he could likely tell you he was in his neurologist's home and that this was not his own bed. Similarly, he would likely confirm for you that it is socially inappropriate to jump into just any bed one might encounter. So what happened here?

We can probably assume that this sequence of actions was a nightly ritual for this man, one that countless bedtimes had hardwired into his brain. Because of his brain damage, however, this patient was unable to use the broader situation to overrule his strong association between beds and getting into them. While automatic actions do not require cognitive control to be initiated, they do require cognitive control to be regulated. Cognitive control allows us to take our knowledge of the broader situation into consideration when deciding things like when we get ready for bed.

When I discuss examples like this in lectures, I am often asked, how can this happen? It's one thing to remove your invisible glasses in a Senate hearing, but it is quite another to go through a lengthy process of undressing and getting into a bed even knowing that doing so is not correct. The patient is conscious after all. But then again, we all act like a frontal patient occasionally. Have you ever accidentally put the milk in the cabinet and the can of soup in the refrigerator? Have you ever missed the turn to your friend's house because going straight was part of your more regular route to work? Even William James writing in 1890 made this point in a delightfully dated way in his discussion of the power of habits to drive our behavior:

> Who is there that has never wound up his watch on taking off his waistcoat in the daytime, or taken his latch-key out on arriving at the door-step of a friend? Very absent-minded persons in going to their bedroom to dress for dinner have been known to take off one garment after another and finally to get into bed, merely because that was the habitual issue of the first few movements when performed at a later hour.[16]

Most of us are not winding our watches or dressing for dinner anymore, but we know exactly what James means. Indeed, the absentminded person in James's example behaves much like our patient, compelled by the habit of this setting to carry out a complex sequence of actions. But the difference is that a healthy person can, if properly focused, wrest control over their automatic actions. The frontal patient is incapable of doing so.

There is a significant gap between knowledge and action the brain must bridge to achieve our goals. Intact knowledge does not guarantee aligned action. Rather, cognitive control processes are required to put it all together—to plan, select, sequence, and monitor actions with knowledge in mind. Further, cognitive control is its own class of mental function, distinct from knowledge and action that is required to bridge this gap. So, what does a cognitive control process look like? We will answer this question in the next section.

Bridging the Gap with Control Processes

So far, we have seen there is a substantive gap between knowledge and action. The fact that just knowing about something is not enough to act on this knowledge is neither obvious nor theoretically convenient, and, indeed, it also eluded psychologists for years. Rather, the contrary assumption—that a direct line links an input to the human brain and its appropriate output as behavior—had been fundamental to many psychologists' thinking at least as early as William James. And this way of thinking continued through the major schools of psychology that dominated the field in the early part of the twentieth century.

It was not until we started building thinking machines ourselves, namely, the modern computer, that cognitive psychologists recognized that control structures were necessary for such a machine to do much of anything. The computer gave psychologists an example of an executive agent that could control itself, and it provided a first example of what a control process might look like: the control flow of a computer program.

Your computer, tablet, smartphone, car, coffee maker, washing machine, and everything with a computer chip in it, which is most things these days, all operate on the basis of programs. Programs are essentially lists of instructions that tell the computer what to do. These lists can be quite massive. For example, a simple iPhone app might have around 10,000 lines of code, the Mac OSX Tiger operating system has more than 80 million, and all of Google's Internet services are estimated at close to 2 billion. These programs are so long because computers are very literal. They won't get the gist of what you want them to do. For a computer to do anything, commands must be exact and explicit. Thus, programming computers made clear to the psychologist just how hard it is to get from what you know to what you actually do.

To illustrate, imagine I want the computer to add together any two numbers I give it. I might give it the following instructions:

Input first number
Assign first number to X
Input second number
Assign second number to Y
Look up X + Y in stored addition table
Assign the result to variable ANSWER
Output the value of ANSWER
End program

Now, I'm expressing all this in what programmers would call "pseudocode," meaning a string of program-like commands that are not in the true syntax of a computer language, like C++ or Python. Real coding would be different and even more explicit. Nonetheless, my pseudocode is enough to illustrate how control flow allows a computer to get tasks done. This little program can be used for adding any two numbers I wish to assign to X and Y and for which I have an entry stored in my computer's lookup table.

It would be annoying to write and rewrite these instructions every time I wanted to add two numbers. A useful trick for a common task like this is to store it as an individual subroutine to be called as needed by other programs. Now that I have my little program for adding two numbers, I might store it as a subroutine called "add_2" and call it anytime I want to add two numbers.

Importantly, however, to do most tasks, I can't call subroutines in just any order. Rather, they must be called at particular times and often under particular conditions. For example, what if I wanted my program to add the two numbers only if I am doing a task called "cash register"? I could use a control process: an "if" statement. For example, I might write the code:

Input task from user
Assign task to variable called TASK
If the value of TASK is "cash register"
    Then run add_2
Otherwise
    Output "I can't do that task!"
End program

If the user tells the computer the task is to be a cash register, now my little program will execute its add_2 subroutine. For any other task, it will just complain. This is an example of a branching control structure. Similar to encountering a fork in the road, the program will follow one branch under one situation and another otherwise.

Directing control flow can allow the computer to perform substantially more complex tasks. For instance, our cash register program is pretty crummy so far, as it can total only two items at checkout. If you want any more items, you'd need to get back in line. So, instead, we should want our cash register to take the prices of a series of items of any length and given in any order and add them together. To do this, I might add another control structure, a "loop," that allows me to iteratively call my add_2 subroutine, adding each new item price to a growing total until there are no further items to add.

The computer has a memory with stored knowledge, in this case the addition tables. It has various devices, like keyboards and touchpads to take input, and screens and speakers for output. It also has lots of operations and subroutines, like add_2, which might be analogous to our automatic action routines. But to make the computer do something, like be a cash register, control has to be passed from one subroutine to the other at the right time and under the right conditions. Control determines whether an operation is relevant, how long to execute it, and whether to stop. Strictly speaking, these control commands are not doing the task themselves. The branching and looping control structures are not adding the numbers or outputting the results. Rather, the control processes guide the flow of processing in the service of an overall computational goal. This control of flow is necessary to get the smaller parts working together in the right way to do a task that is greater than the sum of those parts.

Cognitive psychologists seized on the computer program and its control structures as a guiding analogy for human cognitive control. Prominent among these early ideas was the Test-Operate-Test-Exit structure—or TOTE—proposed by George Miller, Eugene Galanter, and Karl Pribram.[17] TOTE sought to describe what a control structure for human behavior might look like and provides a helpful illustration of several features of any cognitive control system. The basic TOTE structure is illustrated in Figure 1.2A.

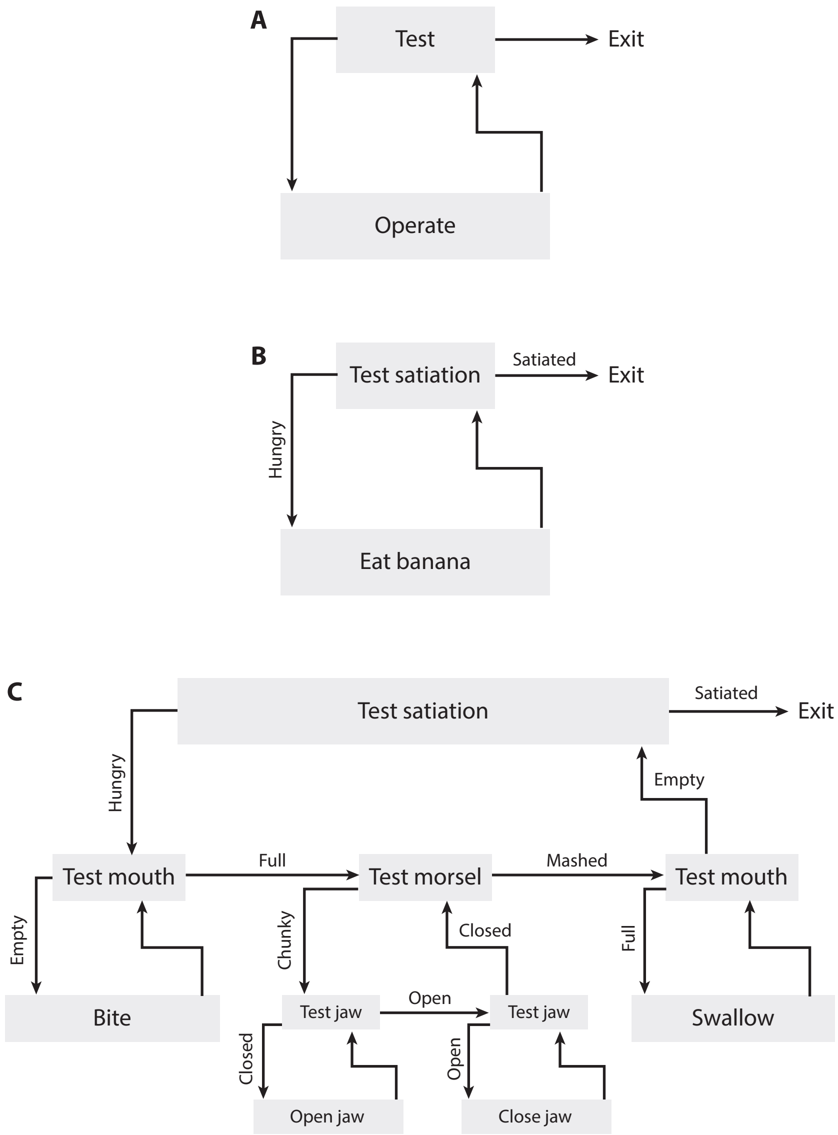

FIGURE 1.2. Schematic of the Miller, Galanter, and Pribram (1960) Test-Operate-Test-Exit (TOTE) structure. (A) The basic TOTE with arrows showing flow of control from Test to Operate to Test to Exit. (B) A simple TOTE example for eating a banana. (C) Hierarchical elaboration of the TOTE control flow for eating a banana. The operator "Eat Banana" is now represented by flow among three sub-TOTEs. The Chew operator of the middle sub-TOTE is further elaborated as two sub-sub-TOTEs that control the jaw.

A TOTE is essentially a loop applied to human behavior. There is an initial "Test" phase, during which an input to the system is compared with a particular condition. If that condition is not met, this incongruity between the desired condition and the input drives an "Operation" to be performed. Following that operation, there is another "Test." If the incongruity still exists, then the operation is performed again, and so forth, until the test condition is met, at which point the loop "Exits," passing control to the next TOTE.

To see how this control system might operate, consider a possible TOTE structure for eating a banana, illustrated in Figure 1.2B. First, we need a test; we'll call it Test Satiation. If we are hungry, then the incongruence of our hungry state with our satiated test condition drives the operation "Eat Banana." In this loop, we will continually eat a banana until we are satiated. At that point, when we Test Satiation again, we will exit the TOTE. We have a simple control structure that determines when and for how long we eat a banana.

Of course, the TOTE structure is an oversimplification. This is likely not the true description of the plan for eating a banana that is implemented in our brain, and no current psychological theory relies on TOTEs as Miller, Galanter, and Pribram described them. But even in its simplicity, the TOTE framework illustrates some key points about the basic structure of a control system.

First, the TOTE structure makes clear why knowing that you want to do something is not enough. Doing so entails elaborating a plan for action that has to be built, for example, around organizing and updating test conditions and directing control flow to the right operations, like eating a banana, at the right times. These conditions are not necessarily expressed in the initial goal, and they might change from one situation to the next. Thus, there must be some kind of control flow to structure behavior.

Second, the TOTE incorporates the concept of both conditional testing and feedback as a means of dynamically governing control flow. In other words, operators are selected based on a condition of the world. To the degree that this condition changes as a consequence of the actions of that operator, then this condition is a way of controlling the operator. As engineers know, incorporating feedback as a way of regulating a system is a powerful means of control.

Third, TOTEs make explicit the importance of incorporating a stop rule that determines when to cease doing something so control can be passed on from a TOTE. As we shall see, locating the conditions for a stop and then executing one, whether by inhibiting action or passing control to a new action, are important features of the brain's control system.

Fourth, TOTEs can be embedded into one another, yielding a hierarchical organization for control. To illustrate, in our example TOTE, the operational stage "Eat Banana" itself can be decomposed into a hierarchical set of sub-TOTEs for biting, chewing, and swallowing the banana that can control flow among each other to consume the banana.

Figure 1.2C depicts an example of a banana-eating TOTE elaborated in this way. Of course, we can further specify even these sub-TOTE operator steps as sub-sub-TOTEs. For example, the Chew operator in Figure 1.2C can be specified as a flow between two subordinate jaw-control TOTEs, the first of which opens a jaw that tests closed, and a second that closes a jaw that tests open.

This capacity for hierarchical structure means that we can continually specify the control flow at finer and finer levels of detail by making the operator phase of each TOTE another subordinate TOTE, which itself can have additional subordinate TOTEs embedded in its operation.

The hierarchical structure of TOTEs matches the hierarchical structure of action itself. Any given task can be described at multiple levels of abstraction. For example, my morning coffee-making routine has this rough sequence: fill the grinder with beans, turn on the grinder, fill the carafe with water, pour water into the reservoir, put the grounds in the filter, turn on the drip machine, wait. Each of these can be further broken down. For example, putting the grounds in the filter involves getting a coffee filter from the cabinet, unfolding it, filling it from the grinder, and so forth. Each of these sub-sub-tasks can itself be decomposed into sub-sub-sub-tasks until I ultimately end up at a very specific sequence of movements. Thus, the ability to control actions hierarchically is a necessary feature of any control system.

Finally, hierarchical control structures like TOTEs illustrate that we require higher degrees of monitoring the more finely we specify our plan. This is evident even in our simple banana example. Just count up the number of Tests versus Operators in the elaboration of our banana-eating TOTE between Figures 1.2b and 1.2c. As our TOTE tree gets deeper, we add tests at a faster rate than we add operations. This is because whenever we embed, we add a new test but replace an operation one-to-one. Thus, the deeper we plan an action, the greater the demands we place on managing, tracking, and executing all those tests. As we shall see, this is analogous to our cognitive control system, too. Much about cognitive control concerns setting up the right test conditions to govern what operations to engage and when.

The Problem of Cognitive Control

Miller, Galanter, and Pribram's TOTE architecture provided a first hypothetical answer to the question of what lies between knowledge and action: a control structure, a plan capable of translating a fundamentally nonhierarchical and fuzzy concept of a task or goal into a form that can guide the rigid, precisely timed, and biophysically constrained hierarchy of the motor system.

As a theory of human cognitive control, however, production systems like TOTE oversimplify the problem in significant ways. For instance, there is no place for motivation or mental effort in these test and operator loops. For our TOTE, bananas are eaten. Why one is eating that banana, or how desperately one wants to do so, or, indeed, at what cost is not a part of the structure.

Further, though the computer metaphor has been enormously useful for recognizing the value in control processes and what they might look like at a functional level, this metaphor also has limits. The mind is not a computer program, and the brain is unlike a digital computer. It is not established that the human brain or mind features the kind of context-free computation that enables the versatility of modern computer programming languages to do virtually any task given the right instructions. Thus, before we set out trying to understand cognitive control in this book, it is important to elaborate the problem of cognitive control a bit more and the unique challenges it presents.

Among the reasons that early schools of psychology rejected the notion of a cognitive control faculty was that doing so seemed a theoretical shortcut, the so-called homunculus problem. The worry was that positing control processes that planned and carried out actions amounted to accounting for how we do things in terms of a little person in the head, a *homunculus*, who knows how to act given the information available and who always takes the right actions to do so. We are left needing to explain what is happening in that little person's head, and so we are forced to posit another, even littler, person in the first little person's head, and so on ad infinitum.

Ideas like TOTE and the computer metaphor helped break this infinite regress. In his landmark first textbook on cognitive psychology in 1967, Ulrich Neisser cited the basic homunculus problem and noted that until the advent of the programmable computer, the only clear example of an executive computational agent we had was the human being.[18] But a computer demonstrated that it was possible for a machine to control itself, given the right productions, without need for a homunculus. So, models like TOTE started to chip away at the homunculus problem by showing that one could build an executive agent without resorting to magical little people in the head. However, these models did not fully banish the homunculus.

Take our case of the TOTE agent eating a banana. This model does not require a little person pulling various levers and buttons in our head to coordinate jaw and banana successfully. The control policies it specifies in its TOTE loops and subloops make clear how test-operate cycles could program the behavior without any guiding agent pulling the strings. But how does this TOTE get set up in the first place? How did this system know to test the banana and the jaw in just the way it did? Indeed, if one flipped the Test Jaw Open, Test Jaw Closed operational phases, one could end up with either lockjaw or a serious drooling problem.

Of course, the answer to how that TOTE got set up that particular way is that I wrote it that way. Whoops. We have posited a homunculus, again. But now it is a little programmer, who can figure out what to do and furiously builds the right productions and control policies into our heads as we encounter new situations. That means we're now back where we started.

To banish the homunculus entirely, then, our theory requires a mechanism not only of control but also of learning. A theory of cognitive control requires a plausible way in which the policies and plans we use are gained through experience, retrieved in the right situation, and elaborated for use by the cognitive control system in a specific situation. Relatedly, that theory of learning should say something about generalization, as in the problem of learning in one specific setting but applying what we learn to multiple other, different settings.

A second fundamental challenge a theory of cognitive control must address is the scaling problem. Again, a central lesson from the study of frontal patients is that cognitive control is concerned with operating in the real world, and often patients' problems become most apparent only in the open-ended complexity of everyday life, rather than in the simple tasks that we test in the laboratory. Yet, scientists working out theories of cognitive control, including me, have spent the majority of our efforts studying these very tasks. For example, more than 3000 papers have been published on the classic Stroop effect alone. The *Stroop effect* refers to the observation that it is difficult to identify the font color of a word when that word names a different color (for example: white). Performing this task undoubtedly requires cognitive control to respond to the font color in the face of interference from our automatic tendency to read the word. Yet, this task is poor at predicting a frontal patient's cognitive control problems in life.[19] Patients show Stroop interference just like everyone else, but whether they show more or less interference on the Stroop task does not consistently predict whether they are likely to get their errands done or successfully hold down a job.

Now, the scientific community is not entirely misguided in our focus on these simple tasks. Rather, we are pursuing the classic scientific strategy of reductionism. Tasks like Stroop are simplified examples—experimental models of cognitive control—that are easily defined and controlled in the laboratory. And because of their simplicity, they also more readily lend themselves to theory. The problem is that we often don't follow through on this reductionist program. We build elegant models of these simple cases, but we rarely attempt to scale these models up to see if they can explain something about the person's behavior in the real world. Indeed, many theories of the simple tasks just can't be scaled. Once the problem becomes complex, the theory that was able to explain the simple problem finds the new problem intractable. Why don't these theories scale easily to the real world? There are many reasons, but we will discuss three big ones: the *curse of dimensionality*, the *degrees of freedom problem*, and the *temporal abstraction problem*.

The curse of dimensionality refers to the fact that the world around us has many features or "dimensions" that might be important for our actions. Take a look around your room and consider all the features that are there. There are likely many objects and lots of different colors, textures, forms, shading, sounds, smells, and so forth. If you were an alien with no particular advance knowledge of the way things work on our planet, which of these features would be most relevant to your behavior right now? What part of the door is important to get the thing open? What if an off-white wall means something important about the way you should walk? The combinations would be limitless.

Any model of learning cognitive control must at some point address this issue, because it is clear that we humans don't need to explicitly learn about every feature in our world and its relevance as a test condition for every possible action. We don't think, "I learned today that the direction of the nap on my rug has no relevance for how I part my hair." Testing every possible case would take many human lifetimes, let alone the relatively few years of human development. The brain must have a way of culling the options and distilling which information in the world is crucial for our behavior.

Contributing to the scaling problem at the other end of the perception-to-action continuum is the so-called degrees of freedom problem. Basically, this is a fancy way of saying that there are many ways to get the same thing done. So, how do we choose the particular one that we end up executing?

This problem was first recognized in basic motor control, referring to the fact that there are actually many hypothetical ways your arm could configure itself to solve a simple movement problem, like sliding a mug across the counter. However, the degrees of freedom problem becomes more complex, less constrained, and less well defined as we reach more abstract levels in the hierarchy of action and focus on more abstract task goals. For example, there are many, many ways to make a cup of coffee. Some choices probably don't matter much to the ultimate goal, while others matter greatly. Turning on the drip machine before adding grounds will probably be a messy decision. But should you grind the beans before filling the carafe? It probably doesn't matter to the ultimate outcome, but the brain still has to choose just one from many ways to do any given thing. Thus, a theory of cognitive control will need to explain how certain plans are specified and actions are taken in a particular situation, even when there are many ways one could act.

Finally, a scalable theory of cognitive control must be able to explain how we generalize what we are doing over time, an ability called temporal abstraction. Many of the simple tasks we conduct in the lab involve a series of short episodes, each of which includes a decision and behavior in response to a particular stimulus. In the experimenter's argot, these episodes are called *trials*. Trials are typically unrelated to one another, in that how you respond on one trial does not affect what happens on the next. As a consequence, we can randomly arrange these trials in any order.

The ability to rearrange trials is a valuable feature for designing scientific experiments, but it is pretty unlike tasks in everyday life. Outside the laboratory, episodes are not randomly tossed at us without cause as we go through our day. We don't lurch from one decision to the next, as if some die roll of fate decides whether we will now eat lunch or take a shower. Rather, our lives and the tasks we do unfold meaningfully in time, over a course of minutes or even hours, rather than seconds. Thus, to control action effectively, the brain's cognitive control system must both leverage this structure and also maintain a sense of continuity, even when the world itself does not signal what task to do. Further, tasks are often open ended, without a clear plan or path to a well-set endpoint. We embark on many tasks with a general idea of what we want to do and then deal with the specifics on the fly. Thus, the control processes that manage temporal abstraction contribute to the very continuity and flow of our lives, and they are at the basis of our ability to be effective, goal-driven people.

The problem of cognitive control is one of bridging the gap from knowledge to action in a complex world. With this problem of cognitive control more clearly before us, in the remainder of the book we will consider the mechanisms by which the brain might solve it. Cognitive and brain sciences in the last several years have provided us with a number of important clues to these mechanisms. We will discuss these discoveries with an eye to both the power and the limits of current evidence and theory.

In the first part of the book, we will lay the theoretical foundation for understanding cognitive control function. We will first consider the evolutionary origins of cognitive control in the mind and brain, with a focus on the emergence in our ancestors of a capacity for detailed, hypothetical future thinking and compositional action planning. Then, we will delve under the hood and take a close look at the nuts and bolts of cognitive control. We will introduce the cognitive and neural mechanisms at the basis of cognitive control function, and we will then consider how they help us to resist impulses, avoid errors, and choose the correct courses of action. We will also see how the brain has elaborated these basic mechanisms to handle complex tasks that are structured hierarchically and change over time and place.

Equipped with this theoretical background, in the latter half of the book we will consider the many facets of cognitive control in our everyday functioning. We will see not only *that* we are bad at multitasking but *why* we are bad at it. We will consider the problem of inhibition, or stopping ourselves from doing unwanted actions or thinking unwanted thoughts. We will explore the close relationship of motivation and cognitive control and will see how control systems not only help us achieve the ends we want but also balance them against the means we don't like. We will see how control makes our memories work for us. And, finally,

(continued...)

INDEX

Page numbers in *italics* refer to illustrations.