the producer that poor Truman, unwittingly on TV in front of an audience of millions, is less a performer than a prisoner. The producer calmly replies:

> And can you tell me, caller, that you're not a player on the stage of life—playing out your allotted role? He can leave at any time. If his was more than just a vague ambition, if he were absolutely determined to discover the truth, there's no way we could prevent him. I think what really distresses you, caller, is that ultimately Truman prefers the comfort of his "cell," as you call it.

As we begin to explore the stage we're on, we find that there is quite a bit beyond our umwelt. The search is a slow, gradual one, but it engenders a deep sense of awe at the size of the wider production studio.

We're now ready to move one level deeper into the brain, uncovering another layer of secrets about what we've been blithely referring to as *you*, as though you were a single entity.

David Eagleman, Pantheon, 2011

5

The Brain Is a Team of Rivals

> "Do I contradict myself?
> Very well then I contradict myself,
> (I am large, I contain multitudes.)"
> —Walt Whitman, *Song of Myself*

WILL THE TRUE MEL GIBSON PLEASE STAND UP?

On July 28, 2006, the actor Mel Gibson was pulled over for speeding at nearly twice the posted speed limit on the Pacific Coast Highway in Malibu, California. The police officer, James Mee, administered a breathalyzer test, which revealed Gibson's blood alcohol level to be 0.12 percent, well over the legal limit. On the seat next to Gibson sat an open bottle of tequila. The officer announced to Gibson that he was under arrest and asked him to get into the squad car. What distinguished this arrest from other Hollywood inebriations was Gibson's surprising and out-of-place inflammatory remarks. Gibson growled, "Fucking Jews. . . . Jews are responsible for all the wars in the world." He then asked the officer, "Are you a Jew?" Mee was indeed Jewish. Gibson refused to get into the squad car and had to be handcuffed.

Less than nineteen hours later, the celebrity website TMZ.com obtained a leak of the handwritten arrest report and posted it immediately. On July 29, after a vigorous response from the media, Gibson offered a note of apology:

> After drinking alcohol on Thursday night, I did a number of things that were very wrong and for which I am ashamed. . . . I acted like a person completely out of control when I was

> arrested, and said things that I do not believe to be true and which are despicable. I am deeply ashamed of everything I said and I apologize to anyone who I have offended. . . . I disgraced myself and my family with my behavior and for that I am truly sorry. I have battled the disease of alcoholism for all of my adult life and profoundly regret my horrific relapse. I apologize for any behavior unbecoming of me in my inebriated state and have already taken necessary steps to ensure my return to health.

Abraham Foxman, head of the Anti-Defamation League, expressed outrage that there was no reference in the apology to the anti-Semitic slurs. In response, Gibson extended a longer note of contrition specifically toward the Jewish community:

> There is no excuse, nor should there be any tolerance, for anyone who thinks or expresses any kind of anti-Semitic remark. I want to apologize specifically to everyone in the Jewish community for the vitriolic and harmful words that I said to a law enforcement officer the night I was arrested on a DUI charge. . . . The tenets of what I profess to believe necessitate that I exercise charity and tolerance as a way of life. Every human being is God's child, and if I wish to honor my God I have to honor his children. But please know from my heart that I am not an anti-Semite. I am not a bigot. Hatred of any kind goes against my faith.

Gibson offered to meet one-on-one with leaders of the Jewish community to "discern the appropriate path for healing." He seemed genuinely contrite, and Abraham Foxman accepted his apology on behalf of the Anti-Defamation League.

Are Gibson's true colors that of an anti-Semite? Or are his true colors those he showed afterward, in his eloquent and apparently heartfelt apologies?

In a *Washington Post* article entitled "Mel Gibson: It Wasn't Just the Tequila Talking," Eugene Robinson wrote, "Well, I'm sorry

about his relapse, but I just don't buy the idea that a little tequila, or even a lot of tequila, can somehow turn an unbiased person into a raging anti-Semite—or a racist, or a homophobe, or a bigot of any kind, for that matter. Alcohol removes inhibitions, allowing all kinds of opinions to escape uncensored. But you can't blame alcohol for forming and nurturing those opinions in the first place."

Lending support to that outlook, Mike Yarvitz, the television producer of *Scarborough Country*, drank alcohol on the show until he raised his blood alcohol level to 0.12 percent, Gibson's level that night. Yarvitz reported "not feeling anti-Semitic" after drinking.

Robinson and Yarvitz, like many others, suspected that the alcohol had loosened Gibson's inhibitions and revealed his true self. And the nature of their suspicion has a long history: the Greek poet Alcaeus of Mytilene coined a popular phrase *En oino álétheia* (In wine there is the truth), which was repeated by the Roman Pliny the Elder as *In vino veritas*. The Babylonian Talmud contains a passage in the same spirit: "In came wine, out went a secret." It later advises, "In three things is a man revealed: in his wine goblet, in his purse, and in his wrath." The Roman historian Tacitus claimed that the Germanic peoples always drank alcohol while holding councils to prevent anyone from lying.

But not everyone agreed with the hypothesis that alcohol revealed the true Mel Gibson. The *National Review* writer John Derbyshire argued, "The guy was drunk, for heaven's sake. We all say and do dumb things when we are drunk. If I were to be judged on my drunken escapades and follies, I should be utterly excluded from polite society, and so would you, unless you are some kind of saint." The Jewish conservative activist David Horowitz commented on Fox News, "People deserve compassion when they're in this kind of trouble. I think it would be very ungracious for people to deny it to him." Addiction psychologist G. Alan Marlatt wrote in *USA Today*, "Alcohol is not a truth serum. . . . It may or may not indicate his true feelings."

In fact, Gibson had spent the afternoon before the arrest at the house of a friend, Jewish film producer Dean Devlin. Devlin stated, "I have been with Mel when he has fallen off, and he becomes a completely different person. It is pretty horrifying." He also stated, "If Mel is an anti-Semite, then he spends a lot of time with us [Devlin and his wife, who is also Jewish], which makes no sense."

So which are Gibson's "true" colors? Those in which he snarls anti-Semitic comments? Or those in which he feels remorse and shame and publicly says, "I am reaching out to the Jewish community for its help"?

Many people prefer a view of human nature that includes a true side and a false side—in other words, humans have a single genuine aim and the rest is decoration, evasion, or cover-up. That's intuitive, but it's incomplete. A study of the brain necessitates a more nuanced view of human nature. As we will see in this chapter, we are made of many neural subpopulations; as Whitman put it, we "contain multitudes." Even though Gibson's detractors will continue to insist that he is truly an anti-Semite, and his defenders will insist that he is not, both may be defending an incomplete story to support their own biases. Is there any reason to believe that it's not possible to have both racist and nonracist parts of the brain?

I AM LARGE, I CONTAIN MULTITUDES

Throughout the 1960s, artificial intelligence pioneers worked late nights to try to build simple robotic programs that could manipulate small blocks of wood: find them, fetch them, stack them in patterns. This was one of those apparently simple problems that turn out to be exceptionally difficult. After all, finding a block of wood requires figuring out which camera pixels correspond to the block and which do not. Recognition of the block shape must be accomplished regardless of the angle and distance of the block. Grabbing it requires visual guidance of graspers that must clench at the correct time, from the correct direction, and with the correct force. Stacking requires an analysis of the rest of the blocks and adjustment to those details. And all these programs need to be

coordinated so that they happen at the correct times in the correct sequence. As we have seen in the previous chapters, tasks that appear simple can require great computational complexity.

Confronting this difficult robotics problem a few decades ago, the computer scientist Marvin Minsky and his colleagues introduced a progressive idea: perhaps the robot could solve the problem by distributing the labor among specialized subagents—small computer programs that each bite off a small piece of the problem. One computer program could be in charge of the job *find*. Another could solve the *fetch* problem, and yet another program could take care of *stack block*. These mindless subagents could be connected in a hierarchy, just like a company, and they could report to one another and to their bosses. Because of the hierarchy, *stack block* would not try to start its job until *find* and *fetch* had finished theirs.

This idea of subagents did not solve the problem entirely—but it helped quite a bit. More importantly, it brought into focus a new idea about the working of biological brains. Minsky suggested that human minds may be collections of enormous numbers of machine-like, connected subagents that are themselves mindless.[1] The key idea is that a great number of small, specialized workers can give rise to something like a society, with all its rich properties that no single subagent, alone, possesses. Minsky wrote, "Each mental agent by itself can only do some simple thing that needs no mind or thought at all. Yet when we join these agents in societies—in certain very special ways—this leads to intelligence." In this framework, thousands of little minds are better than one large one.

To appreciate this approach, just consider how factories work: each person on the assembly line is specialized in a single aspect of production. No one knows how to do everything; nor would that equate to efficient production if they did. This is also how government ministries operate: each bureaucrat has one task or a few very specific tasks, and the government succeeds on its ability to distribute the work appropriately. On larger scales, civilizations operate in the same manner: they reach the next level of sophistication when they learn to divide labor, committing some experts to agriculture, some to art, some to warfare, and so on.[2] The division of labor allows specialization and a deeper level of expertise.

The idea of dividing up problems into subroutines ignited the young field of artificial intelligence. Instead of trying to develop a single, all-purpose computer program or robot, computer scientists shifted their goal to equipping the system with smaller "local expert" networks that know how to do one thing, and how to do it well.[3] In such a framework, the larger system needs only to switch which of the experts has control at any given time. The learning challenge now involves not so much how to do each little task but, instead, how to distribute who's doing what when.[4]

As Minsky suggests in his book *The Society of Mind*, perhaps that's all the human brain has to do as well. Echoing William James' concept of instincts, Minsky notes that if brains indeed work this way—as collections of subagents—we would not have any reason to be aware of the specialized processes:

> Thousands and, perhaps, millions of little processes must be involved in how we anticipate, imagine, plan, predict, and prevent—and yet all this proceeds so automatically that we regard it as "ordinary common sense." . . . At first it may seem incredible that our minds could use such intricate machinery and yet be unaware of it.[5]

When scientists began to look into the brains of animals, this society-of-mind idea opened up new ways of looking at things. In the early 1970s, researchers realized that the frog, for example, has at least two separate mechanisms for detecting motion: one system directs the snapping of the frog's tongue to small, darting objects, such as flies, while a second system commands the legs to jump in response to large, looming objects.[6] Presumably, neither of these systems is conscious—instead, they are simple, automated programs burned down into the circuitry.

The society-of-mind framework was an important step forward.

But despite the initial excitement about it, a collection of experts with divided labor has never proven sufficient to yield the properties of the human brain. It is still the case that our smartest robots are less intelligent than a three-year-old child.

So what went wrong? I suggest that a critical factor has been missing from the division-of-labor models, and we turn to that now.

THE DEMOCRACY OF MIND

The missing factor in Minsky's theory was *competition* among experts who all believe they know the right way to solve the problem. Just like a good drama, the human brain runs on conflict.

In an assembly line or government ministry, each worker is an expert in a small task. In contrast, parties in a democracy hold differing opinions *about the same issues*—and the important part of the process is the battle for steering the ship of state. Brains are like representative democracies.[7] They are built of multiple, overlapping experts who weigh in and compete over different choices. As Walt Whitman correctly surmised, we are large and we harbor multitudes within us. And those multitudes are locked in chronic battle.

There is an ongoing conversation among the different factions in your brain, each competing to control the single output channel of your behavior. As a result, you can accomplish the strange feats of arguing with yourself, cursing at yourself, and cajoling yourself to do something—feats that modern computers simply do not do. When the hostess at a party offers chocolate cake, you find yourself on the horns of a dilemma: some parts of your brain have evolved to crave the rich energy source of sugar, and other parts care about the negative consequences, such as the health of your heart or the bulge of your love handles. Part of you wants the cake and part of you tries to muster the fortitude to forgo it. The final vote of the parliament determines which party controls your action—that is, whether you put your hand out or up. In the end, you either eat the chocolate cake or you do not, but you cannot do both.

Because of these internal multitudes, biological creatures can be conflicted. The term *conflicted* could not be sensibly applied to an entity that has a single program. Your car cannot be conflicted about which way to turn: it has one steering wheel commanded by only one driver, and it follows directions without complaint. Brains, on the other hand, can be of two minds, and often many more. We don't know whether to turn toward the cake or away from it, because there are several little sets of hands on the steering wheel of our behavior.

Consider this simple experiment with a laboratory rat: if you put both food *and* an electrical shock at the end of an alley, the rat finds himself stuck at a certain distance from the end. He begins to approach but withdraws; he begins to withdraw but finds the courage to approach again. He oscillates, conflicted.[8] If you outfit the rat with a little harness to measure the force with which he pulls toward food alone and, separately, you measure the force with which he pulls away from an electric shock alone, you find that the rat gets stuck at the point where the two forces are equal and cancel out. The pull matches the push. The perplexed rat has two pair of paws on his steering wheel, each pulling in opposite directions—and as a result he cannot get anywhere.

Brains—whether rat or human—are machines made of conflicting parts. If building a contraption with internal division seems strange, just consider that we already build social machines of this type: think of a jury of peers in a courtroom trial. Twelve strangers with differing opinions are tasked with the single mission of coming to a consensus. The jurors debate, coax, influence, relent—and eventually the jury coheres to reach a single decision. Having differing opinions is not a drawback to the jury system, it is a central feature.

Inspired by this art of consensus building, Abraham Lincoln chose to place adversaries William Seward and Salmon Chase in his presidential cabinet. He was choosing, in the memorable phrase

of historian Doris Kearns Goodwin, a team of rivals. Rivalrous teams are central in modern political strategy. In February 2009, with Zimbabwe's economy in free fall, President Robert Mugabe agreed to share power with Morgan Tsvangirai, a rival he'd earlier tried to assassinate. In March 2009, Chinese president President Hu Jintao named two indignantly opposing faction leaders, Xi Jinping and Li Keqiang, to help craft China's economic and political future.

I propose that the brain is best understood as a team of rivals, and the rest of this chapter will explore that framework: who the parties are, how they compete, how the union is held together, and what happens when things fall apart. As we proceed, remember that competing factions typically have the same goal—success for the country—but they often have different ways of going about it. As Lincoln put it, rivals should be turned into allies "for the sake of the greater good," and for neural subpopulations the common interest is the thriving and survival of the organism. In the same way that liberals and conservatives both love their country but can have acrimoniously different strategies for steering it, so too does the brain have competing factions that all believe they know the right way to solve problems.

THE DOMINANT TWO-PARTY SYSTEM: REASON AND EMOTION

When trying to understand the strange details of human behavior, psychologists and economists sometimes appeal to a "dual-process" account.[9] In this view, the brain contains two separate systems: one is fast, automatic, and below the surface of conscious awareness, while the other is slow, cognitive, and conscious. The first system can be labeled automatic, implicit, heuristic, intuitive, holistic, reactive, and impulsive, while the second system is cognitive, systematic, explicit, analytic, rule-based, and reflective.[10] These two processes are always battling it out.

Despite the "dual-process" moniker, there is no real reason to assume that there are only two systems—in fact, there may be several systems. For example, in 1920 Sigmund Freud suggested three competing parts in his model of the psyche: the id (instinctive), the ego (realistic and organized), and the superego (critical and moralizing).[11] In the 1950s, the American neuroscientist Paul MacLean suggested that the brain is made of three layers representing successive stages of evolutionary development: the reptilian brain (involved in survival behaviors), the limbic system (involved in emotions), and the neocortex (used in higher-order thinking). The details of both of these theories have largely fallen out of favor among neuroanatomists, but the heart of the idea survives: brains are made of competing subsystems. We will proceed using the generalized dual-process model as a starting point, because it adequately conveys the thrust of the argument.

Although psychologists and economists think of the different systems in abstract terms, modern neuroscience strives for an anatomical grounding. And it happens that the wiring diagram of the brain lends itself to divisions that generally map onto the dual-process model.[12] Some areas of your brain are involved in higher-order operations regarding events in the outside world (these include, for example, the surface of the brain just inside your temples, called the dorsolateral prefrontal cortex). In contrast, other areas are involved with monitoring your internal state, such as your level of hunger, sense of motivation, or whether something is rewarding to you (these areas include, for example, a region just behind your forehead called the medial prefrontal cortex, and several areas deep below the surface of the cortex). The situation is more complicated than this rough division would imply, because brains can simulate future states, reminisce about the past, figure out where to find things not immediately present, and so on. But for the moment, this division into systems that monitor the external and internal will serve as a rough guide, and a little later we will refine the picture.

In the effort to use labels tied neither to black boxes nor to neuroanatomy, I've chosen two that will be familiar to everyone: the

rational and *emotional* systems. These terms are underspecified and imperfect, but they will nonetheless carry the central point about rivalries in the brain.[13] The rational system is the one that cares about analysis of things in the outside world, while the emotional system monitors internal state and worries whether things will be good or bad. In other words, as a rough guide, rational cognition involves external events, while emotion involves your internal state. You can do a math problem without consulting your internal state, but you can't order a dessert off a menu or prioritize what you feel like doing next.[14] The emotional networks are absolutely required to rank your possible next actions in the world: if you were an emotionless robot who rolled into a room, you might be able to make analyses about the objects around you, but you would be frozen with indecision about what to do next. Choices about the priority of actions are determined by our internal states: whether you head straight to the refrigerator, bathroom, or bedroom upon returning home depends not on the external stimuli in your home (those have not changed), but instead on your body's internal states.

A TIME FOR MATH, A TIME TO KILL

The battle between the rational and emotional systems is brought to light by what philosophers call the trolley dilemma. Consider this scenario: A trolley is barreling down the train tracks, out of control. Five workers are making repairs way down the track, and you, a bystander, quickly realize that they will all be killed by the trolley. But you also notice that there is a switch nearby that you can throw, and that will divert the trolley down a different track, where only a single worker will be killed. What do you do? (Assume there are no trick solutions or hidden information.)

If you are like most people, you will have no hesitation about throwing the switch: it's far better to have one person killed than five, right? Good choice.

Now here's an interesting twist to the dilemma: imagine that the same trolley is barreling down the tracks, and the same five workers are in harm's way—but this time you are a bystander on a footbridge that goes over the tracks. You notice that there is an obese man standing on the footbridge, and you realize that if you were to push him off the bridge, his bulk would be sufficient to stop the train and save the five workers. Do you push him off?

If you're like most people, you bristle at this suggestion of murdering an innocent person. But wait a minute. What differentiates this from your previous choice? Aren't you trading one life for five lives? Doesn't the math work out the same way?

What exactly is the difference in these two cases? Philosophers working in the tradition of Immanuel Kant have proposed that the difference lies in how people are being used. In the first scenario, you are simply reducing a bad situation (the deaths of five people) to a less bad situation (the death of one). In the case of the man on the bridge, he is being exploited as a means to an end. This is a popular explanation in the philosophy literature. Interestingly, there may be a more brain-based approach to understand the reversal in people's choices.

In the alternative interpretation, suggested by the neuroscientists Joshua Greene and Jonathan Cohen, the difference in the two scenarios pivots on the emotional component of actually touching someone—that is, interacting with him at a close distance.[15] If the problem is constructed so that the man on the footbridge can be dropped, with the flip of switch, through a trapdoor, many people will vote to let him drop. Something about interacting with the person up close stops most people from pushing the man to his death. Why? Because that sort of personal interaction activates the emotional networks. It changes the problem from an abstract, impersonal math problem into a personal, emotional decision.

When people consider the trolley problem, here's what brain imaging reveals: In the footbridge scenario, areas involved in motor planning and emotion become active. In contrast, in the track-switch scenario, only lateral areas involved in rational thinking

become active. People register emotionally when they have to push someone; when they only have to tip a lever, their brain behaves like *Star Trek*'s Mr. Spock.

* * *

The battle between emotional and rational networks in the brain is nicely illustrated by an old episode of *The Twilight Zone*. I am paraphrasing from memory, but the plot goes something like this: A stranger in an overcoat shows up at a man's door and proposes a deal. "Here is a box with a single button on it. All you have to do is press the button and I will pay you a thousand dollars."

"What happens when I press the button?" the man asks.

The stranger tells him, "When you press the button, someone far away, someone you don't even know, will die."

The man suffers over the moral dilemma through the night. The button box rests on his kitchen table. He stares at it. He paces around it. Sweat clings to his brow.

Finally, after an assessment of his desperate financial situation, he lunges to the box and punches the button. Nothing happens. It is quiet and anticlimactic.

Then there is a knock at the door. The stranger in the overcoat is there, and he hands the man the money and takes the box. "Wait," the man shouts after him. "What happens now?"

The stranger says, "Now I take the box and give it to the next person. Someone far away, someone you don't even know."

The story highlights the ease of impersonally pressing a button: if the man had been asked to attack someone with his hands, he presumably would have declined the bargain.

In earlier times in our evolution, there was no real way to interact with others at a distance any farther than that allowed by hands, feet, or possibly a stick. That distance of interaction was salient and consequential, and this is what our emotional reaction reflects. In modern times, the situation differs: generals and even soldiers commonly find themselves far removed from the people they kill.

In Shakespeare's *Henry VI, Part 2*, the rebel Jack Cade challenges Lord Say, mocking the fact that he has never known the firsthand danger of the battlefield: "When struck'st thou one blow in the field?" Lord Say responds, "Great men have reaching hands: oft have I struck those that I never saw, and struck them dead." In modern times, we can launch forty Tomahawk surface-to-surface missiles from the deck of navy ships in the Persian Gulf and Red Sea with the touch of a button. The result of pushing that button may be watched by the missile operators live on CNN, minutes later, when Baghdad's buildings disappear in plumes. The proximity is lost, and so is the emotional influence. This impersonal nature of waging war makes it disconcertingly easy. In the 1960s, one political thinker suggested that the button to launch a nuclear war should be implanted in the chest of the President's closest friend. That way, should the President want to make the decision to annihilate millions of people on the other side of the globe, he would first have to physically harm his friend, ripping open his chest to get to the button. That would at least engage his emotional system in the decision making, so as to guard against letting the choice be impersonal.

Because both of the neural systems battle to control the single output channel of behavior, emotions can tip the balance of decision making. This ancient battle has turned into a directive of sorts for many people: *If it feels bad, it is probably wrong.*[16] There are many counter examples to this (for example, one may find oneself put off by another's sexual preference but still deem nothing morally wrong with that choice), but emotion nonetheless serves as a generally useful steering mechanism for decision making.

The emotional systems are evolutionarily old, and therefore shared with many other species, while the development of the rational system is more recent. But as we have seen, the novelty of the rational system does not necessarily indicate that it is, by itself, superior. Societies would *not* be better off if everyone were like Mr. Spock, all rationality and no emotion. Instead, a balance—a teaming up of the internal rivals—is optimal for brains. This is

because the disgust we feel at pushing the man off the footbridge is critical to social interaction; the impassivity one feels at pressing a button to launch a Tomahawk missile is detrimental to civilization. Some balance of the emotional and rational systems is needed, and that balance may already be optimized by natural selection in human brains. To put it another way, a democracy split across the aisle may be just what you want—a takeover in either direction would almost certainly prove less optimal. The ancient Greeks had an analogy for life that captured this wisdom: you are a charioteer, and your chariot is pulled by two thunderous horses, the white horse of reason and the black horse of passion. The white horse is always trying to tug you off one side of the road, and the black horse tries to pull you off the other side. Your job is to hold on to them tightly, keeping them in check so you can continue down the middle of the road.

The emotional and rational networks battle not only over immediate moral decisions, but in another familiar situation as well: how we behave in time.

WHY THE DEVIL CAN SELL YOU FAME NOW FOR YOUR SOUL LATER

Some years ago, the psychologists Daniel Kahneman and Amos Tversky posed a deceptively simple question: If I were to offer you $100 right now or $110 a week from now, which would you choose? Most subjects chose to take $100 right then. It just didn't seem worthwhile to wait an entire week for another $10.

Then the researchers changed the question slightly: If I were to offer you $100 fifty-two weeks from now, or $110 fifty-three weeks from now, which would you choose? Here people tended to switch their preference, choosing to wait the fifty-three weeks. Note that the two scenarios are identical in that waiting one extra week earns you an extra $10. So why is there a preference reversal between the two?[17]

It's because people "discount" the future, an economic term meaning that rewards closer to now are valued more highly than rewards in the distant future. Delaying gratification is difficult. And there is something very special about *right now*—which always holds the highest value. Kahneman and Tversky's preference reversal comes about because the discounting has a particular shape: it drops off very quickly into the near future, and then flattens out a bit, as though more distant times are all about the same. That shape happens to look like the shape you would get if you combined two simpler processes: one that cares about short-term reward and one that holds concerns more distantly into the future.

That gave an idea to neuroscientists Sam McClure, Jonathan Cohen, and their colleagues. They reconsidered the preference-reversal problem in light of the framework of multiple competing systems in the brain. They asked volunteers to make these something-now-or-more-later economic decisions while in a brain scanner. The scientists searched for a system that cared about immediate gratification, and another that involved longer-term rationality. If the two operate independently, and fight against each other, that just might explain the data. And indeed, they found that some emotionally involved brain structures were highly activated by the choice of immediate or near-term rewards. These areas were associated with impulsive behavior, including drug addiction. In contrast, when participants opted for longer-term rewards with higher return, lateral areas of the cortex involved in higher cognition and deliberation were more active.[18] And the higher the activity in these lateral areas, the more the participant was willing to defer gratification.

Sometime between 2005 and 2006, the United States housing bubble burst. The problem was that 80 percent of recently issued mortgages were adjustable-rate. The subprime borrowers who had signed up for these loans suddenly found themselves stuck with higher payment rates and no way to refinance. Delinquencies soared. Between late 2007 and 2008, almost one million U.S. homes were foreclosed on. Mortgage-backed securities rapidly

lost most of their value. Credit around the world tightened. The economy melted.

What did this have to do with competing systems in the brain? Subprime mortgage offers were perfectly optimized to take advantage of the I-want-it-now system: buy this beautiful house now with very low payments, impress your friends and parents, live more comfortably than you thought you could. At some point the interest rate on your adjustable-rate mortgage will go up, but that's a long way away, hidden in the mists of the future. By plugging directly into these instant-gratification circuits, the lenders were able to almost tank the American economy. As the economist Robert Shiller noted in the wake of the subprime mortgage crisis, speculative bubbles are caused by "contagious optimism, seemingly impervious to facts, that often takes hold when prices are rising. Bubbles are primarily social phenomena; until we understand and address the psychology that fuels them, they're going to keep forming."[19]

When you begin to look for examples of I-want-it-now deals, you'll see them everywhere. I recently met a man who accepted $500 while he was a college student in exchange for signing his body away to a university medical school after he dies. The students who accepted the deal all received ankle tattoos that tell the hospital, decades from now, where their bodies should be delivered. It's an easy sell for the school: $500 now feels good, while death is inconceivably distant. There is nothing wrong with donating one's body, but this serves to illustrate the archetypical dual-process conflict, the proverbial deal with the Devil: your wishes granted now for your soul in the distant future.

These sorts of neural battles often lie behind marital infidelity. Spouses make promises in a moment of heartfelt love, but later can find themselves in a situation in which present temptations tip their decision making the other way. In November 1995, Bill Clinton's brain decided that risking the future leadership of the free world was counterbalanced by the pleasure he had the opportunity to experience with the winsome Monica in the present moment.

So when we talk about a virtuous person, we do not necessarily mean someone who is not tempted but, instead, someone who is able to *resist* that temptation. We mean someone who does not let that battle tip to the side of instant gratification. We value such people because it is easy to yield to impulses, and inordinately difficult to ignore them. Sigmund Freud noted that arguments stemming from the intellect or from morality are weak when pitted against human passions and desires,[20] which is why campaigns to "just say no" or practice abstinence will never work. It has also been proposed that this imbalance of reason and emotion may explain the tenacity of religion in societies: world religions are optimized to tap into the emotional networks, and great arguments of reason amount to little against such magnetic pull. Indeed, the Soviet attempts to squelch religion were only partially successful, and no sooner had the government collapsed than the religious ceremonies sprang richly back to life.

The observation that people are made of conflicting short- and long-term desires is not a new one. Ancient Jewish writings proposed that the body is composed of two interacting parts: a body (*guf*), which always wants things now, and a soul (*nefesh*), which maintains a longer-term view. Similarly, Germans use a fanciful expression for a person trying to delay gratification: he must overcome his *innerer schweinehund*—which translates, sometimes to the puzzlement of English speakers, as "inner pigdog."

Your behavior—what you do in the world—is simply the end result of the battles. But the story gets better, because the different parties in the brain can learn about their interactions with one another. As a result, the situation quickly surpasses simple arm wrestling between short- and long-term desires and enters the realm of a surprisingly sophisticated process of negotiation.

THE PRESENT AND FUTURE ULYSSES

In 1909, Merkel Landis, treasurer of the Carlisle Trust Company in Pennsylvania, went on a long walk and was struck with a new

financial idea. He would start something called a Christmas club. Customers would deposit money with the bank throughout the year, and there would be a fee if they withdrew their money early. Then, at the end of the year, people could access their money just in time for holiday shopping. If the idea worked, the bank would have plenty of capital to reinvest and profit from all year. But would it work? Would people willingly give up their capital all year for little or no interest?

Landis tried it, and the concept immediately caught fire. That year, almost four hundred patrons of the bank socked away an average of $28 each—quite a bit of money in the early 1900s. Landis and the other bankers couldn't believe their luck. Patrons *wanted* them to hold on to their money.

The popularity of Christmas banking clubs grew quickly, and banks soon found themselves battling each other for the holiday nest egg business. Newspapers exhorted parents to enroll their children in Christmas clubs "to develop self-reliance and the saving habit."[21] By the 1920s, several banks, including the Dime Saving Bank of Toledo, Ohio, and the Atlantic Country Trust Co. in Atlantic City, New Jersey, began manufacturing attractive brass Christmas club tokens to entice new customers.[22] (The Atlantic City tokens read, "Join our Christmas Club and Have Money When You Need It Most.")

But why did Christmas clubs catch on? If the depositors controlled their own money throughout the year, they could earn better interest or invest in emerging opportunities. Any economist would advise them to hold on to their own capital. So why would people willingly ask a bank to take away their money, especially in the face of restrictions and early withdrawal fees? The answer is obvious: people wanted someone to stop them from spending their money. They knew that if they held on to their own money, they were likely to blow it.[23]

For this same reason, people commonly use the Internal Revenue Service as an ersatz Christmas club: by claiming fewer deductions on their paychecks, they allow the IRS to keep more of their money during the year. Then, come next April, they receive the joy of a check in the mailbox. It feels like free money—but of course it's only your own. And the government got to earn interest on it instead of you. Nonetheless, people choose this route when they intuit that the extra money will burn a hole in their pocket during the year. They'd rather grant someone else the responsibility to protect them from impulsive decisions.

Why don't people take control of their own behavior and enjoy the opportunities of commanding their own capital? To understand the popularity of the Christmas club and IRS phenomena, we need to step back three millennia to the king of Ithaca and a hero of the Trojan War, Ulysses.

After the war, Ulysses was on a protracted sea voyage back to his home island of Ithaca when he realized he had a rare opportunity in front of him. His ship would be passing the island of Sirenum scopuli, where the beautiful Sirens sang melodies so alluring they beggared the human mind. The problem was that sailors who heard this music steered toward the tricky maidens, and their ships were dashed into the unforgiving rocks, drowning all aboard.

So Ulysses hatched a plan. He knew that when he heard the music, he would be as unable to resist as any other mortal man, so he came up with an idea to deal with his *future self*. Not the present, rational Ulysses, but the future, crazed Ulysses. He ordered his men to lash him to the mast of the ship and tie him there securely. This way he would be unable to move when the music wafted over the bow of the ship. Then he had them fill their ears with beeswax so they could not be seduced by the voices of the Sirens—or hear his crazed commands. He made it clear to them that they should not respond to his entreaties and should not release him until the ship was well past the Sirens. He surmised that he would be screaming, yelling, cursing, trying to force the men to steer toward the mellifluous women—he knew that this future Ulysses would be in no position to make good decisions. Therefore, the Ulysses of sound mind structured things in such a way as to

prevent himself from doing something foolish when they passed the upcoming island. It was a deal struck between the present Ulysses and the future one.

This myth highlights the way in which minds can develop a meta-knowledge about how the short- and long-term parties interact. The amazing consequence is that minds can negotiate with different time points of themselves.[24]

So imagine the hostess pressing the chocolate cake upon you. Some parts of your brain want that glucose, while others parts care about your diet; some parts look at the short-term gain, other parts at long-term strategy. The battle tips toward your emotions and you decide to dig in. But not without a contract: you'll eat it only if you promise to go to the gym tomorrow. Who's negotiating with whom? Aren't both parties in the negotiation *you?*

Freely made decisions that bind you in the future are what philosophers call a Ulysses contract.[25] As a concrete example, one of the first steps in breaking an alcohol addiction is to ensure, during sober reflection, that there is no alcohol in the house. The temptation will simply be too great after a stressful workday or on a festive Saturday or a lonely Sunday.

People make Ulysses contracts all the time, and this explains the immediate and lasting success of Merkel Landis's Christmas club. When people handed over their capital in April, they were acting with a wary eye toward their October selves, who they knew would be tempted to blow the money on something selfish instead of deferring to their generous, gift-giving December selves.

Many arrangements have evolved to allow people to proactively bind the options of their future selves. Consider the existence of websites that help you lose weight by negotiating a business deal with your future self. Here's how it works: you pay a deposit of $100 with the promise that you will lose ten pounds. If you succeed by the promised time, you get all the money back. If you don't lose the weight by that time, the company keeps the money. These arrangements work on the honor system and could easily be cheated, but nonetheless these companies are profiting. Why? Because people understand that as they come closer to the date when they can win back their money, their emotional systems will care more and more about it. They are pitting short- and long-term systems against each other.*

Ulysses contracts often arise in the context of medical decision making. When a person in good health signs an advance medical directive to pull the plug in the event of a coma, he is binding himself in a contract with a possible future self—even though it is arguable that the two selves (in health and in sickness) are quite different.

An interesting twist on the Ulysses contract comes about when someone else steps in to make a decision for you—and binds your present self in deference to your future self. These situations arise commonly in hospitals, when a patient, having just experienced a traumatic life change, such as losing a limb or a spouse, declares that she wants to die. She may demand, for example, that her doctors stop her dialysis or give her an overdose of morphine. Such cases typically go before ethics boards, and the boards usually decide the same thing: don't let the patient die, because the future patient will eventually find a way to regain her emotional footing and reclaim happiness. The ethics board here acts simply as an advocate for the rational, long-term system, recognizing that the present context allows the intellect little voice against the emotions.[26] The board essentially decides that the neural congress is unfairly tilted at the moment, and that an intervention is needed to prevent a one-party takeover. Thank goodness that we can sometimes rely on the dispassion of someone else, just as Ulysses relied on his sailors to ignore his pleas. The rule of thumb is this: when

*Although this system works, it strikes me that there is a way to better match this business model to the neurobiology. The problem is that weight loss demands a sustained effort, while the approaching deadline for the loss of money is always distantly in the future until the day of reckoning is suddenly upon you. In a neurally optimized model, you would lose a little money each day until you have shed the ten pounds. Each day, the amount you'd lose would increase by fifteen percent. So every day brings the immediate emotional sting of monetary loss, and the sting constantly grows worse. When you've lost the ten pounds, then you stop losing money. This encourages a sustained diet ethic over the entire time window.

you cannot rely on your own rational systems, borrow someone else's.[27] In this case, patients borrow the rational systems of the board members. The board can more easily take responsibility for protecting the future patient, as its members do not hear the emotional Siren songs in which the patient is ensnared.

OF MANY MINDS

For the purpose of illustrating the team-of-rivals framework, I have made the oversimplification of subdividing the neuroanatomy into the rational and emotional systems. But I do not want to give the impression that these are the only competing factions. Instead, they are only the beginning of the team-of-rivals story. Everywhere we look we find overlapping systems that compete.

One of the most fascinating examples of competing systems can be seen with the two hemispheres of the brain, left and right. The hemispheres look roughly alike and are connected by a dense highway of fibers called the corpus callosum. No one would have guessed that the left and right hemispheres formed two halves of a team of rivals until the 1950s, when an unusual set of surgeries were undertaken. Neurobiologists Roger Sperry and Ronald Meyers, in some experimental surgeries, cut the corpus callosum of cats and monkeys. What happened? Not much. The animals acted normal, as though the massive band of fibers connecting the two halves was not really necessary.

As a result of this success, split-brain surgery was first performed on human epilepsy patients in 1961. For them, an operation that prevented the spread of seizures from one hemisphere to the other was the last hope. And the surgeries worked beautifully. A person who had suffered terribly with debilitating seizures could now live a normal life. Even with the two halves of his brain separated, the patient did not seem to act differently. He could remember events normally and learn new facts without trouble. He could love and laugh and dance and have fun.

But something strange was going on. If clever strategies were used to deliver information only to one hemisphere and not the other, then one hemisphere could learn something while the other would not. It was as though the person had two independent brains.[28] And the patients could do different tasks at the same time, something that normal brains cannot do. For example, with a pencil in each hand, split brain patients could simultaneously draw incompatible figures, such as a circle and a triangle.

There was more. The main motor wiring of the brain crosses sides, such that the right hemisphere controls the left hand and the left hemisphere controls the right hand. And that fact allows a remarkable demonstration. Imagine that the word *apple* is flashed to the left hemisphere, while the word *pencil* is simultaneously flashed to the right hemisphere. When a split-brain patient is asked to grab the item he just saw, the right hand will pick up the apple while the left hand will simultaneously pick up the pencil. The two halves are now living their own lives, disconnected.

Researchers came to realize, over time, that the two hemispheres have somewhat different personalities and skills—this includes their abilities to think abstractly, create stories, draw inferences, determine the source of a memory, and make good choices in a gambling game. Roger Sperry, one of the neurobiologists who pioneered the split-brain studies (and garnered a Nobel Prize for it), came to understand the brain as "two separate realms of conscious awareness; two sensing, perceiving, thinking and remembering systems." The two halves constitute a team of rivals: agents with the same goals but slightly different ways of going about it.

In 1976, the American psychologist Julian Jaynes proposed that until late in the second millennium B.C.E., humans had no introspective consciousness, and that instead their minds were essentially divided into two, with their left hemispheres following the commands from their right hemispheres.[29] These commands, in the form of auditory hallucinations, were interpreted as voices from the gods. About three thousand years ago, Jaynes suggests, this division of labor between the left and right hemispheres began to

break down. As the hemispheres began to communicate more smoothly, cognitive processes such as introspection were able to develop. The origin of consciousness, he argues, resulted from the ability of the two hemispheres to sit down at the table together and work out their differences. No one yet knows whether Jaynes's theory has legs, but the proposal is too interesting to ignore.

The two hemispheres look almost identical anatomically. It's as though you come equipped with the same basic model of brain hemisphere in the two sides of your skull, both absorbing data from the world in slightly different ways. It's essentially one blueprint stamped out twice. And nothing could be better suited for a team of rivals. The fact that the two halves are doubles of the same basic plan is evidenced by a type of surgery called a hemispherectomy, in which one entire half of the brain is removed (this is done to treat intractable epilepsy caused by Rasmussen's encephalitis). Amazingly, as long as the surgery is performed on a child before he is about eight years old, the child is fine. Let me repeat that: the child, with only half his brain remaining, is fine. He can eat, read, speak, do math, make friends, play chess, love his parents, and everything else that a child with two hemispheres can do. Note that it is not possible to remove *any* half of the brain: you cannot remove the front half or the back half and expect survival. But the right and left halves reveal themselves as something like copies of each other. Take one away and you still have another, with roughly redundant function. Just like a pair of political parties. If the Republicans or Democrats disappeared, the other would still be able to run the country. The approach would be slightly different, but things would still work.

CEASELESS REINVENTION

I've begun with examples of rational systems versus emotional systems, and the two-factions-in-one-brain phenomenon unmasked by split-brain surgeries. But the rivalries in the brain are far more numerous, and far more subtle, than the broad-stroke ones I have introduced so far. The brain is full of smaller subsystems that have overlapping domains and take care of coinciding tasks.

Consider memory. Nature seems to have invented mechanisms for storing memory more than once. For instance, under normal circumstances, your memories of daily events are consolidated (that is, "cemented in") by an area of the brain called the hippocampus. But during frightening situations—such as a car accident or a robbery—another area, the amygdala, also lays down memories along an independent, secondary memory track.[30] Amygdala memories have a different quality to them: they are difficult to erase and they can pop back up in "flashbulb" fashion—as commonly described by rape victims and war veterans. In other words, there is more than one way to lay down memory. We're not talking about a memory of different events, but multiple memories of the *same* event—as though two journalists with different personalities were jotting down notes about a single unfolding story.

So we see that different factions in the brain can get involved in the same task. In the end, it is likely that there are even more than two factions involved, all writing down information and later competing to tell the story.[31] The conviction that memory is one thing is an illusion.

Here's another example of overlapping domains. Scientists have long debated how the brain detects motion. There are many theoretical ways to build motion detectors out of neurons, and the scientific literature has proposed wildly different models that involve connections between neurons, or the extended processes of neurons (called dendrites), or large populations of neurons.[32] The details aren't important here; what's important is that these different theories have kindled decades of debates among academics. Because the proposed models are too small to measure directly, researchers design clever experiments to support or contradict various theories. The interesting outcome has been that most of the experiments are inconclusive,

supporting one model over another in some laboratory conditions but not in others. This has led to a growing recognition (reluctantly, for some) that there are *many* ways the visual system detects motion. Different strategies are implemented in different places in the brain. As with memory, the lesson here is that the brain has evolved multiple, redundant ways of solving problems.[33] The neural factions often agree about what is out there in the world, but not always. And this provides the perfect substrate for a neural democracy.

The point I want to emphasize is that biology rarely rests with a single solution. Instead, it tends to ceaselessly reinvent solutions. But why endlessly innovate—why not find a good solution and move on? Unlike the artificial intelligence laboratory, the laboratory of nature has no master programmer who checks off a subroutine once it is invented. Once the *stack block* program is coded and polished, human programmers move on to the next important step. I propose that this moving on is a major reason artificial intelligence has become stuck. Biology, in contrast to artificial intelligence, takes a different approach: when a biological circuit for *detect motion* has been stumbled upon, there is no master programmer to report this to, and so random mutation continues to ceaselessly invent new variations in circuitry, solving *detect motion* in unexpected and creative new ways.

This viewpoint suggests a new approach to thinking about the brain. Most of the neuroscience literature seeks *the* solution to whatever brain function is being studied. But that approach may be misguided. If a space alien landed on Earth and discovered an animal that could climb a tree (say, a monkey), it would be rash for the alien to conclude that the monkey is the only animal with these skills. If the alien keeps looking, it will quickly discover that ants, squirrels, and jaguars also climb trees. And this is how it goes with clever mechanisms in biology: when we keep looking, we find more. Biology never checks off a problem and calls it quits. It reinvents solutions continually. The end product of that approach is a highly overlapping system of solutions—the necessary condition for a team-of-rivals architecture.[34]

THE ROBUSTNESS OF A MULTIPLE-PARTY SYSTEM

The members of a team can often disagree, but they do not have to. In fact, much of the time rivals enjoy a natural concordance. And that simple fact allows a team of rivals to be robust in the face of losing parts of the system. Let's return to the thought experiment of a disappearing political party. Imagine that all the key decision makers of a particular party were to die in an airplane crash, and let's consider this roughly analogous to brain damage. In many cases the loss of one party would expose the polarized, opposing opinions of a rival group—as in the case when the frontal lobes are damaged, allowing for bad behavior such as shoplifting or urinating in public. But there are other cases, perhaps much more common, in which the disappearance of a political party goes unnoticed, because all the other parties hold roughly the same opinion on some matter (for example, the importance of funding residential trash collection). This is the hallmark of a robust biological system: political parties can perish in a tragic accident and the society will still run, sometimes with little more than a hiccup to the system. It may be that for every strange clinical case in which brain damage leads to a bizarre change in behavior or perception, there are hundreds of cases in which parts of the brain are damaged with no detectable clinical sign.

An advantage of overlapping domains can be seen in the newly discovered phenomenon of *cognitive reserve*. Many people are found to have the neural ravages of Alzheimer's disease upon autopsy—but they never showed the symptoms while they were alive. How can this be? It turns out that these people continued to challenge their brains into old age by staying active in their careers, doing crossword puzzles, or carrying out any other activities that kept their neural populations well exercised. As a result of staying mentally vigorous, they built what neuropsychologists call cognitive reserve. It's not that cognitively fit people don't get

Alzheimer's; it's that their brains have protection against the symptoms. Even while parts of their brains degrade, they have other ways of solving problems. They are not stuck in the rut of having a single solution; instead, thanks to a lifetime of seeking out and building up redundant strategies, they have alternative solutions. When parts of the neural population degraded away, they were not even missed.

Cognitive reserve—and robustness in general—is achieved by blanketing a problem with overlapping solutions. As an analogy, consider a handyman. If he has several tools in his toolbox, then losing his hammer does not end his career. He can use his crowbar or the flat side of his pipe wrench. The handyman with only a couple of tools is in worse trouble.

The secret of redundancy allows us to understand what was previously a bizarre clinical mystery. Imagine that a patient sustains damage to a large chunk of her primary visual cortex, and an entire half of her visual field is now blind. You, the experimenter, pick up a cardboard shape, hold it up to her blind side, and ask her, "What do you see here?"

She says, "I have no idea—I'm blind in that half of my visual field."

"I know," you say. "But take a guess. Do you see a circle, square, or triangle?"

She says, "I really can't tell you. I don't see anything at all. I'm blind there."

You say, "I know, I know. But *guess*."

Finally, with exasperation, she guesses that the shape is a triangle. And she's *correct*, well above what random chance would predict.[35] Even though she's blind, she can tease out a hunch—and this indicates that *something* in her brain is seeing. It's just not the conscious part that depends on the integrity of her visual cortex. This phenomenon is called blindsight, and it teaches us that when conscious vision is lost, there are still subcortical factory workers behind the scenes running their normal programs. So removal of parts of the brain (in this case, the cortex) reveals underlying structures that do the same thing, just not as well. And from a neuroanatomical point of view, this is not surprising: after all, reptiles can see even though they have no cortex at all. They don't see as well as we do, but they see.[36]

* * *

Let's pause for a moment to consider how the team-of-rivals framework offers a different way of thinking about the brain than is traditionally taught. Many people tend to assume that the brain will be divisible into neatly labeled regions that encode, say, faces, houses, colors, bodies, tool use, religious fervor, and so on. This was the hope of the early-nineteenth-century science of phrenology, in which bumps on the skull were assumed to represent something about the size of the underlying areas. The idea was that each spot in the brain could be assigned a label on the map.

But biology rarely, if ever, pans out that way. The team-of-rivals framework presents a model of a brain that possesses multiple ways of representing the same stimulus. This view rings the death knell for the early hopes that each part of the brain serves an easily labeled function.

Note that the phrenological impulse has crept back into the picture because of our newfound power to visualize the brain with neuroimaging. Both scientists and laypeople can find themselves seduced into the easy trap of wanting to assign each function of the brain to a specific location. Perhaps because of pressure for simple sound bites, a steady stream of reports in the media (and even in the scientific literature) has created the false impression that the brain area for such-and-such has just been discovered. Such reports feed popular expectation and hope for easy labeling, but the true situation is much more interesting: the continuous networks of neural circuitry accomplish their functions using multiple, independently discovered strategies. The brain lends itself well to the complexity of the world, but poorly to clear-cut cartography.

KEEPING THE UNION TOGETHER: CIVIL WARS IN THE BRAIN DEMOCRACY

In the campy cult movie *Evil Dead 2*, the protagonist's right hand takes on a mind of its own and tries to kill him. The scene degenerates into a rendition of what you might find on a sixth-grade playground: the hero uses his left hand to hold back his right hand, which is trying to attack his face. Eventually he cuts off the hand with a chain saw and traps the still-moving hand under an upside-down garbage can. He stacks books on top of the can to pin it down, and the careful observer can see that the topmost book is Hemingway's *A Farewell to Arms*.

As preposterous as this plotline may seem, there is, in fact, a disorder called *alien hand syndrome*. While it's not as dramatic as the *Evil Dead* version, the idea is roughly the same. In alien hand syndrome, which can result from the split-brain surgeries we discussed a few pages ago, the two hands express conflicting desires. A patient's "alien" hand might pick up a cookie to put it in his mouth, while the normally behaving hand will grab it at the wrist to stop it. A struggle ensues. Or one hand will pick up a newspaper, and the other will slap it back down. Or one hand will zip up a jacket, and the other will unzip it. Some patients with alien hand syndrome have found that yelling "Stop!" will cause the other hemisphere (and the alien hand) to back down. But besides that little modicum of control, the hand is running on its own inaccessible programs, and that is why it's branded as alien—because the conscious part of the patient seems to have no predictive power over it; it does not feel as though it's part of the patient's personality at all. A patient in this situation often says, "I swear I'm not doing this." Which revisits one of the main points of this book: who is the *I*? His own brain is doing it, not anyone else's. It's simply that he doesn't have conscious access to those programs.

What does alien hand syndrome tell us? It unmasks the fact that we harbor mechanical, "alien" subroutines to which we have no access and of which we have no acquaintance. Almost all of our actions—from producing speech to picking up a mug of coffee—are run by alien subroutines, also known as zombie systems. (I use these terms interchangeably: *zombie* emphasizes the lack of conscious access, while *alien* emphasizes the foreignness of the programs.)[37] Some alien subroutines are instinctual, while some are learned; all of the highly automated algorithms that we saw in Chapter 3 (serving the tennis ball, sexing the chicks) become inaccessible zombie programs when they are burned down into the circuitry. When a professional baseball player connects his bat with a pitch that is traveling too fast for his conscious mind to track, he is leveraging a well-honed alien subroutine.

Alien hand syndrome also tells us that under normal circumstances, all the automated programs are tightly controlled such that only one behavioral output can happen at a time. The alien hand highlights the normally seamless way in which the brain keeps a lid on its internal conflicts. It requires only a little structural damage to uncover what is happening beneath. In other words, keeping the union of subsystems together is not something the brain does without effort—instead, it is an active process. It is only when factions begin to secede from the union that the alienness of the parts becomes obvious.

A good illustration of conflicting routines is found in the Stroop test, a task that could hardly have simpler instructions: name the color of the *ink* in which a word is printed. Let's say I present the word JUSTICE written in blue letters. You say, "Blue." Now I show you PRINTER written in yellow. "Yellow." Couldn't be easier. But the trick comes when I present a word that is itself the name of a color. I present the word BLUE in the color green. Now the reaction is not so easy. You might blurt out, "Blue!", or you might stop yourself and sputter out, "Green!" Either way, you have a much slower reaction time—and this belies the conflict going on under the hood. This *Stroop interference* unmasks the clash between the strong, involuntary and automatic impulse to read the word and the unusual, deliberate, and effortful task demand to state the color of the print.[38]

Remember the implicit association task from Chapter 3, the one that seeks to tease out unconscious racism? It pivots on the slower-than-normal reaction time when you're asked to link something you dislike with a positive word (such as *happiness*). Just as with the Stroop task, there's an underlying conflict between deeply embedded systems.

E PLURIBUS UNUM

Not only do we run alien subroutines; we also justify them. We have ways of retrospectively telling stories about our actions as though the actions were always our idea. As an example at the beginning of the book, I mentioned that thoughts come to us and we take credit for them ("I just had a great idea!"), even though our brains have been chewing on a given problem for a long time and eventually served up the final product. We are constantly fabricating and telling stories about the alien processes running under the hood.

To bring this sort of fabrication to light, we need only look at another experiment with split-brain patients. As we saw earlier, the right and left halves are similar to each other but not identical. In humans, the left hemisphere (which contains most of the capacity to speak language) can speak about what it is feeling, whereas the mute right hemisphere can communicate its thoughts only by commanding the left hand to point, reach, or write. And this fact opens the door to an experiment regarding the retrospective fabrication of stories. In 1978, researchers Michael Gazzaniga and Joseph LeDoux flashed a picture of a chicken claw to the left hemisphere of a split-brain patient and a picture of a snowy winter scene to his right hemisphere. The patient was then asked to point at cards that represented what he had just seen. His right hand pointed to a card with a chicken, and his left hand pointed to a card with a snow shovel. The experimenters asked him why he was pointing to the shovel. Recall that his left hemisphere (the one with the capacity for language), had information only about a chicken, and nothing else. But the left hemisphere, without missing a beat, fabricated a story: "Oh, that's simple. The chicken claw goes with the chicken, and you need a shovel to clean out the chicken shed." When one part of the brain makes a choice, other parts can quickly invent a story to explain why. If you show the command "Walk" to the right hemisphere (the one without language), the patient will get up and start walking. If you stop him and ask why he's leaving, his left hemisphere, cooking up an answer, will say something like "I was going to get a drink of water."

The chicken/shovel experiment led Gazzaniga and LeDoux to conclude that the left hemisphere acts as an "interpreter," watching the actions and behaviors of the body and assigning a coherent narrative to these events. And the left hemisphere works this way even in normal, intact brains. Hidden programs drive actions, and the left hemisphere makes justifications. This idea of retrospective storytelling suggests that we come to know our own attitudes and emotions, at least partially, by inferring them from observations of our own behavior.[39] As Gazzaniga put it, "These findings all suggest that the interpretive mechanism of the left hemisphere is always hard at work, seeking the meaning of events. It is constantly looking for order and reason, even when there is none—which leads it continually to make mistakes."[40]

This fabrication is not limited to split-brain patients. Your brain, as well, interprets your body's actions and builds a story around them. Psychologists have found that if you hold a pencil between your teeth while you read something, you'll think the material is funnier; that's because the interpretation is influenced by the smile on your face. If you sit up straight instead of slouching, you'll feel happier. The brain assumes that if the mouth and spine are doing that, it must be because of cheerfulness.

* * *

On December 31, 1974, Supreme Court Justice William O. Douglas was debilitated by a stroke that paralyzed his left side and confined him to a wheelchair. But Justice Douglas demanded to be checked out of the hospital on the grounds that he was fine. He declared that reports of his paralysis were "a myth." When reporters expressed skepticism, he publicly invited them to join him for a hike, a move interpreted as absurd. He even claimed to be kicking football field goals with his paralyzed side. As a result of this apparently delusional behavior, Douglas was dismissed from his bench on the Supreme Court.

What Douglas experienced is called *anosognosia*. This term describes a total lack of awareness about an impairment, and a typical example is a patient who completely denies their very obvious paralysis. It's not that Justice Douglas was *lying*—his brain actually believed that he could move just fine. These fabrications illustrate the lengths to which the brain will go to put together a coherent narrative. When asked to place both hands on an imaginary steering wheel, a partially paralyzed and anosognosic patient will put one hand up, but not the other. When asked if both hands are on the wheel, he will say yes. When the patient is asked to clap his hands, he may move only a single hand. If asked, "Did you clap?", he'll say yes. If you point out that you didn't hear any sound and ask him to do it again, he might not do it at all; when asked why, he'll say he "doesn't feel like it." Similarly, as mentioned in Chapter 2, one can lose vision and claim to still be able to see just fine, even while being unable to navigate a room without crashing into the furniture. Excuses are made about poor balance, rearranged chairs, and so on—all the while denying the blindness. The point about anosognosia is that the patients are not lying, and are motivated neither by mischievousness nor by embarrassment; instead, their brains are fabricating explanations that provide a coherent narrative about what is going on with their damaged bodies.

But shouldn't the contradicting evidence alert these people to a problem? After all, the patient wants to move his hand, but it is not moving. He wants to clap, but he hears no sound. It turns out that alerting the system to contradictions relies critically on particular brain regions—and one in particular, called the anterior cingulate cortex. Because of these conflict-monitoring regions, incompatible ideas will result in one side or another winning out: a story will be constructed that either makes them compatible or ignores one side of the debate. In special circumstances of brain damage, this arbitration system can be damaged—and then conflict can cause no trouble to the conscious mind. This situation is illustrated by a woman I'll call Mrs. G., who had suffered quite a bit of damage to her brain tissue from a recent stroke. At the time I met her, she was recovering in the hospital with her husband by her bedside, and seemed generally in good health and spirits. My colleague Dr. Karthik Sarma had noticed the night before that when he asked her to close her eyes, she would close only one and not the other. So he and I went to examine this more carefully.

When I asked her to close her eyes, she said "Okay," and closed one eye, as in a permanent wink.

"Are your eyes closed?" I asked.

"Yes," she said.

"Both eyes?"

"Yes."

I held up three fingers. "How many fingers am I holding up, Mrs. G.?"

"Three," she said.

"And your eyes are closed?"

"Yes."

In a nonchallenging way I said, "Then how did you know how many fingers I was holding up?"

An interesting silence followed. If brain activity were audible, this is when we would have heard different regions of her brain battling it out. Political parties that wanted to believe her eyes were closed were locked in a filibuster with parties that wanted the logic to work out: *Don't you see that we can't have our eyes closed* and *be able to see out there?* Often these battles are quickly won by the party with the most reasonable position, but this does

not always happen with anosognosia. The patient will say nothing and will conclude nothing—not because she is embarrassed, but because she is simply locked up on the issue. Both parties fatigue to the point of attrition, and the original issue being fought over is finally dumped. The patient will conclude nothing about the situation. It is amazing and disconcerting to witness.

I was struck with an idea. I wheeled Mrs. G. to a position just in front of the room's only mirror and asked if she could see her own face. She said yes. I then asked her to close both her eyes. Again she closed one eye and not the other.

"Are *both* your eyes closed?"

"Yes."

"Can you see yourself?"

"Yes."

Gently I said, "Does it seem possible to see yourself in the mirror if both your eyes are closed?"

Pause. *No conclusion.*

"Does it look to you like one eye is closed or that both are closed?"

Pause. *No conclusion.*

She was not distressed by the questions; nor did they change her opinion. What would have been a checkmate in a normal brain proved to be a quickly forgotten game in hers.

Cases like Mrs. G.'s allow us to appreciate the amount of work that needs to happen behind the scenes for our zombie systems to work together smoothly and come to an agreement. Keeping the union together and making a good narrative does not happen for free—the brain works around the clock to stitch together a pattern of logic to our daily lives: what just happened and what was my role in it? Fabrication of stories is one of the key businesses in which our brains engage. Brains do this with the single-minded goal of getting the multifaceted actions of the democracy to make sense. As the coin puts it, *E pluribus unum*: out of many, one.

* * *

Once you have learned how to ride a bicycle, the brain does not need to cook up a narrative about what your muscles are doing; instead, it doesn't bother the conscious CEO at all. Because everything is predictable, no story is told; you are free to think of other issues as you pedal along. The brain's storytelling powers kick into gear only when things are conflicting or difficult to understand, as for the split-brain patients or anosognosics like Justice Douglas.

In the mid-1990s my colleague Read Montague and I ran an experiment to better understand how humans make simple choices. We asked participants to choose between two cards on a computer screen, one labeled A and the other labeled B. The participants had no way of knowing which was the better choice, so they picked arbitrarily at first. Their card choice gave them a reward somewhere between a penny and a dollar. Then the cards were reset and they were asked to choose again. Picking the same card produced a different reward this time. There seemed to be a pattern to it, but it was very difficult to detect. What the participants didn't know was that the reward in each round was based on a formula that incorporated the history of their previous forty choices—far too difficult for the brain to detect and analyze.

The interesting part came when I interviewed the players afterward. I asked them what they'd done in the gambling game and why they'd done it. I was surprised to hear all types of baroque explanations, such as "The computer liked it when I switched back and forth" and "The computer was trying punish me, so I switched my game plan." In reality, the players' descriptions of their own strategies did not match what they had actually done, which turned out to be highly predictable.[41] Nor did their descriptions match the computer's behavior, which was purely formulaic. Instead, their conscious minds, unable to assign the task to a well-oiled zombie system, desperately sought a narrative. The participants weren't *lying*; they were giving the best explanation they could—just like the split-brain patients or the anosognosics.

Minds seek patterns. In a term introduced by science writer

Michael Shermer, they are driven toward "patternicity"—the attempt to find structure in meaningless data.[42] Evolution favors pattern seeking, because it allows the possibility of reducing mysteries to fast and efficient programs in the neural circuitry.

To demonstrate patternicity, researchers in Canada showed subjects a light that flashed on and off randomly and asked them to choose which of two buttons to press, and when, in order to make the blinking more regular. The subjects tried out different patterns of button pressing, and eventually the light began to blink regularly. They had succeeded! Now the researchers asked them how they'd done it. The subjects overlaid a narrative interpretation about what they'd done, but the fact is that their button pressing was wholly unrelated to the behavior of the light: the blinking would have drifted toward regularity irrespective of what they were doing.

For another example of storytelling in the face of confusing data, consider dreams, which appear to be an interpretative overlay to nighttime storms of electrical activity in the brain. A popular model in the neuroscience literature suggests that dream plots are stitched together from essentially random activity: discharges of neural populations in the midbrain. These signals tickle into existence the simulation of a scene in a shopping mall, or a glimpse of recognition of a loved one, or a feeling of falling, or a sense of epiphany. All these moments are dynamically woven into a story, and this is why after a night of random activity you wake up, roll over to your partner, and feel as though you have a bizarre plot to relate. Ever since I was a child, I have been consistently amazed at how characters in my dreams possess such specific and peculiar details, how they come up with such rapid answers to my questions, how they produce such surprising dialogue and such inventive suggestions—all manner of things I would not have invented "myself." Many times I've heard a new joke in a dream, and this impressed me greatly. Not because the joke was so funny in the sober light of day (it wasn't) but because the joke was not one I could believe that *I* would have thought of. But, at least presumably, it was my brain and no one else's cooking up these interesting plotlines.[43] Like the split-brain patients or Justice Douglas, dreams illustrate our skills at spinning a single narrative from a collection of random threads. Your brain is remarkably good at maintaining the glue of the union, even in the face of thoroughly inconsistent data.

WHY DO WE HAVE CONSCIOUSNESS AT ALL?

Most neuroscientists study animal models of behavior: how a sea slug withdraws from a touch, how a mouse responds to rewards, how an owl localizes sounds in the dark. As these circuits are scientifically brought to light, they all reveal themselves to be nothing but zombie systems: blueprints of circuitry that respond to particular inputs with appropriate outputs. If our brains were composed *only* of these patterns of circuits, why would it feel like anything to be alive and conscious? Why wouldn't it feel like nothing—like a zombie?

A decade ago, neuroscientists Francis Crick and Christof Koch asked, "Why does not our brain consist simply of a series of specialized zombie systems?"[44] In other words, why are we conscious of anything at all? Why aren't we simply a vast collection of these automated, burned-down routines that solve problems?

Crick and Koch's answer, like mine in the previous chapters, is that consciousness exists to control—and to distribute control over—the automated alien systems. A system of automated subroutines that reaches a certain level of complexity (and human brains certainly qualify) requires a high-level mechanism to allow the parts to communicate, dispense resources, and allocate control. As we saw earlier with the tennis player trying to learn how to serve, consciousness is the CEO of the company: he sets the higher-level directions and assigns new tasks. We have learned in this chapter that he doesn't need to understand the software that each department in the organization uses; nor does he need to see their

detailed logbooks and sales receipts. He merely needs to know whom to call on when.

As long as the zombie subroutines are running smoothly, the CEO can sleep. It is only when something goes wrong (say, all the departments suddenly find that their business models have catastrophically failed) that the CEO is rung up. Think about *when* your conscious awareness comes online: in those situations where events in the world *violate your expectations*. When everything is going according to the needs and skills of your zombie systems, you are not consciously aware of most of what's in front of you; when suddenly they cannot handle the task, you become consciously aware of the problem. The CEO scrambles around, looking for fast solutions, dialing up everyone to find who can address the problem best.

The scientist Jeff Hawkins offers a nice example of this: after he entered his home one day, he realized that he had experienced no conscious awareness of reaching for, grasping, and turning the doorknob. It was a completely robotic, unconscious action on his part—and this was because everything about the experience (the doorknob's feel and location, the door's size and weight, and so on) was already burned down into unconscious circuitry in his brain. It was expected, and therefore required no conscious participation. But he realized that if someone were to sneak over to his house, drill the doorknob out, and replace it three inches to the right, he would notice immediately. Instead of his zombie systems getting him directly into his house with no alerts or concerns, suddenly there would be a violation of expectations—and consciousness would come online. The CEO would rouse, turn on the alarms, and try to figure out what might have happened and what should be done next.

If you think you're consciously aware of most of what surrounds you, think again. The first time you make the drive to your new workplace, you attend to everything along the way. The drive seems to take a long time. By the time you've made the drive many times, you can get yourself there without much in the way of conscious deliberation. You are now free to think about other things; you feel as though you've left home and arrived at work in the blink of an eye. Your zombie systems are experts at taking care of business as usual. It is only when you see a squirrel in the road, or a missing stop sign, or an overturned vehicle on the shoulder that you become consciously aware of your surroundings.

All of this is consistent with a finding we learned two chapters ago: when people play a new video game for the first time, their brains are alive with activity. They are burning energy like crazy. As they get better at the game, less and less brain activity is involved. They have become more energy efficient. If you measure someone's brain and see very little activity during a task, it does not necessarily indicate that they're not trying—it more likely signifies that they have worked hard in the past to burn the programs into the circuitry. Consciousness is called in during the first phase of learning and is excluded from the game playing after it is deep in the system. Playing a simple video game becomes as unconscious a process as driving a car, producing speech, or performing the complex finger movements required for tying a shoelace. These become hidden subroutines, written in an undeciphered programming language of proteins and neurochemicals, and there they lurk—for decades sometimes—until they are next called upon.

From an evolutionary point of view, the purpose of consciousness seems to be this: an animal composed of a giant collection of zombie systems would be energy efficient but *cognitively inflexible*. It would have economical programs for doing particular, simple tasks, but it wouldn't have rapid ways of switching between programs or setting goals to become expert in novel and unexpected tasks. In the animal kingdom, most animals do certain things very well (say, prying seeds from the inside of a pine cone), while only a few species (such as humans) have the flexibility to dynamically develop new software.

Although the ability to be flexible sounds better, it does not

come for free—the trade-off is a burden of lengthy childrearing. To be flexible like an adult human requires years of helplessness as an infant. Human mothers typically bear only one child at a time and have to provide a period of care that is unheard-of (and impracticable) in the rest of the animal kingdom. In contrast, animals that run only a few very simple subroutines (such as "Eat foodlike things and shrink away from looming objects") adopt a different rearing strategy, usually something like "Lay lots of eggs and hope for the best." Without the ability to write new programs, their only available mantra is: If you can't outthink your opponents, outnumber them.

So are other animals conscious? Science currently has no meaningful way to make a measurement to answer that question—but I offer two intuitions. First, consciousness is probably not an all-or-nothing quality, but comes in degrees. Second, I suggest that an animal's *degree of consciousness* will parallel its intellectual flexibility. The more subroutines an animal possesses, the more it will require a CEO to lead the organization. The CEO keeps the subroutines unified; it is the warden of the zombies. To put this another way, a small corporation does not require a CEO who earns three million dollars a year, but a large corporation does. The only difference is the number of workers the CEO has to keep track of, allocate among, and set goals for.*

If you put a red egg in the nest of a herring gull, it goes berserk. The color red triggers aggression in the bird, while the shape of the egg triggers brooding behavior—as a result, it tries to simultaneously attack the egg and incubate it.[45] It's running two programs at once, with an unproductive end result. The red egg sets off sovereign and conflicting programs, wired into the gull's brain like competing fiefdoms. The rivalry is there, but the bird

*There may be other advantages to having a large collection of alien systems with flexible allocation. For example, it may reduce our predictability to predators. If you had only one subroutine and ran it every time, a predator would know exactly how to pick you off (think of the crocodiles grazing on the wildebeest that swim across African rivers the same way, at the same time, every year). More complex collections of alien systems enjoy not only flexibility but a better shot at unpredictability.

has no capacity to arbitrate in the service of smooth cooperation. Similarly, if a female stickleback trespasses onto a male's territory, the male will display attack behavior and courtship behavior simultaneously, which is no way to win over a lady. The poor male stickleback appears to be simply a bundled collection of zombie programs triggered by simple lock-and-key inputs (*Trespass! Female!*), and the subroutines have not found any method of arbitration between them. This seems to me to suggest that the herring gull and the stickleback are not particularly conscious.

I propose that a useful index of consciousness is the capacity to successfully mediate conflicting zombie systems. The more an animal looks like a jumble of hardwired input–output subroutines, the less it gives evidence of consciousness; the more it can coordinate, delay gratification, and learn new programs, the more conscious it may be. If this view is correct, in the future a battery of tests might be able to yield a rough measure of a species' degree of consciousness. Think back to the befuddled rat we met near the beginning of the chapter, who, trapped between the drive to go for the food and the impulse to run from the shock, became stuck in between and oscillated back and forth. We all know what it's like to have moments of indecision, but our human arbitration between the programs allows us to escape these conundrums and make a decision. We quickly find ways of cajoling or castigating ourselves toward one outcome or the other. Our CEO is sophisticated enough to get us out of the simple lockups that can thoroughly hamstring the poor rat. This may be the way in which our conscious minds—which play only a small part in our total neural function—really shine.

THE MULTITUDES

Let's circle back to how this allows us to think about our brains in a new way—that is, how the team-of-rivals framework allows us to address mysteries that would be inexplicable if we took

the point of view of traditional computer programs or artificial intelligence.

Consider the concept of a secret. The main thing known about secrets is that keeping them is unhealthy for the brain.[46] Psychologist James Pennebaker and his colleagues studied what happened when rape and incest victims, acting out of shame or guilt, chose to hold secrets inside. After years of study, Pennebaker concluded that "the act of *not* discussing or confiding the event with another may be more damaging than having experienced the event per se."[47] He and his team discovered that when subjects confessed or wrote about their deeply held secrets, their health improved, their number of doctor visits went down, and there were measurable decreases in their stress hormone levels.[48]

The results are clear enough, but some years ago I began to ask myself how to understand these findings from the point of view of brain science. And that led to a question that I realized was unaddressed in the scientific literature: what *is* a secret, neurobiologically? Imagine constructing an artificial neural network of millions of interconnected neurons—what would a secret look like here? Could a toaster, with its interconnected parts, harbor a secret? We have useful scientific frameworks for understanding Parkinson's disease, color perception, and temperature sensation—but none for understanding what it means for the brain to have and to hold a secret.

Within the team-of-rivals framework, a secret is easily understood: it is the result of struggle between competing parties in the brain. One part of the brain wants to reveal something, and another part does not want to. When there are competing votes in the brain—one for telling, and one for withholding—that defines a secret. If no party cares to tell, that's merely a boring fact; if both parties want to tell, that's just a good story. Without the framework of rivalry, we would have no way to understand a secret.*

*Some people are constitutionally incapable of keeping a secret, and this balance may tell us something about the battles going on inside them and which way they tip. Good spies and secret agents are those people whose battle always tips toward long-term decision making rather than the thrill of telling.

The reason a secret is experienced consciously is because it results from a rivalry. It is not business as usual, and therefore the CEO is called upon to deal with it.

The main reason not to reveal a secret is aversion to the long-term consequences. A friend might think ill of you, or a lover might be hurt, or a community might ostracize you. This concern about the outcome is evidenced by the fact that people are more likely to tell their secrets to total strangers; with someone you don't know, the neural conflict can be dissipated with none of the costs. This is why strangers can be so forthcoming on airplanes, telling all the details of their marital troubles, and why confessional booths have remained a staple in one of the world's largest religions. It may similarly explain the appeal of prayer, especially in those religions that have very personal gods, deities who lend their ears with undivided attention and infinite love.

The newest twist on this ancient need to tell secrets to a stranger can be found in the form of websites like postsecret.com, where people go to anonymously disclose their confessions. Here are some examples: "When my only daughter was stillborn, I not only thought about kidnapping a baby, I planned it out in my head. I even found myself watching new mothers with their babies trying to pick the perfect one"; "I am almost certain that your son has autism but I have no idea how to tell you"; "Sometimes I wonder why my dad molested my sister but not me. Was I not good enough?"

As you have doubtless noticed, venting a secret is usually done for its own sake, not as an invitation for advice. If the listener spots an obvious solution to some problem revealed by the secret and makes the mistake of suggesting it, this will frustrate the teller—all she *really* wanted was to tell. The act of telling a secret can itself be the solution. An open question is why the receiver of the secrets has to be human—or human-like, in the case of deities. Telling a wall, a lizard, or a goat your secrets is much less satisfying.

WHERE IS C3PO?

When I was a child, I assumed that we would have robots by now—robots that would bring us food and clean our clothes and converse with us. But something went wrong with the field of artificial intelligence, and as a result the only robot in my home is a moderately dim-witted self-directing vacuum cleaner.

Why did artificial intelligence become stuck? The answer is clear: intelligence has proven itself a tremendously hard problem. Nature has had an opportunity to try out trillions of experiments over billions of years. Humans have been scratching at the problem only for decades. For most of that time, our approach has been to cook up intelligence from scratch—but just recently the field has taken a turn. To make meaningful progress in building thinking robots, it is now clear that we need to decipher the tricks nature has figured out.

I suggest that the team-of-rivals framework will play an important role in dislodging the jammed field of artificial intelligence. Previous approaches have made the useful step of dividing labor—but the resulting programs are impotent without differing opinions. If we hope to invent robots that think, our challenge is not simply to devise a subagent to cleverly solve each problem but instead to ceaselessly reinvent subagents, each with overlapping solutions, and then to pit them against one another. Overlapping factions offer protection against degradation (think of cognitive reserve) as well as clever problem solving by unexpected approaches.

Human programmers approach a problem by assuming there's a *best* way to solve it, or that there's a way it *should* be solved by the robot. But the main lesson we can extract from biology is that it's better to cultivate a team of populations that attack the problem in different, overlapping manners. The team-of-rivals framework suggests that the best approach is to abandon the question "What's the most clever way to solve that problem?" in favor of "Are there multiple, overlapping ways to solve that problem?"

Probably the best way to cultivate a team is with an evolutionary approach, randomly generating little programs and allowing them to reproduce with small mutations. This strategy allows us to continuously discover solutions rather than trying to think up a single perfect solution from scratch. As the biologist Leslie Orgel's second law states: "Evolution is smarter than you are." If I had a law of biology, it would be: "Evolve solutions; when you find a good one, *don't stop*."

Technology has so far not taken advantage of the idea of a democratic architecture—that is, the team-of-rivals framework. Although your computer is built of thousands of specialized parts, they never collaborate or argue. I suggest that conflict-based, democratic organization—summarized as the team-of-rivals architecture—will usher in a fruitful new age of biologically inspired machinery.[49]

* * *

The main lesson of this chapter is that you are made up of an entire parliament of pieces and parts and subsystems. Beyond a collection of local expert systems, we are collections of overlapping, ceaselessly reinvented mechanisms, a group of competing factions. The conscious mind fabricates stories to explain the sometimes inexplicable dynamics of the subsystems inside the brain. It can be disquieting to consider the extent to which all of our actions are driven by hardwired systems, doing what they do best, while we overlay stories about our choices.

Note that the population of the mental society does not always vote exactly the same way each time. This recognition is often missing from discussions of consciousness, which typically assume that what it is like to be you is the same from day to day, moment to moment. Sometimes you're able to read well; other times you drift. Sometimes you can find all the right words; other times your tongue is tangled. Some days you're a stick in the mud; other days you throw caution to the wind. So who's the real you? As

the French essayist Michel de Montaigne put it, "There is as much difference between us and ourselves as there is between us and others."

A nation is at any moment most readily defined by its political parties in power. But it is also defined by the political opinions it harbors in its streets and living rooms. A comprehensive understanding of a nation must include those parties that are not in power but that could rise in the right circumstances. In this same way, you are composed of your multitudes, even though at any given time your conscious headline may involve only a subset of all the political parties.

Returning to Mel Gibson and his drunken tirade, we can ask whether there is such a thing as "true" colors. We have seen that behavior is the outcome of the battle among internal systems. To be clear, I'm not defending Gibson's despicable behavior, but I am saying that a team-of-rivals brain can naturally harbor both racist and nonracist feelings. Alcohol is not a truth serum. Instead, it tends to tip the battle toward the short-term, unreflective faction—which has no more or less claim than any other faction to be the "true" one. Now, we may *care* about the unreflective faction in someone, because it defines the degree to which they're *capable* of antisocial or dangerous behavior. It is certainly rational to worry about this aspect of a person, and it makes sense to say, "Gibson is capable of anti-Semitism." In the end, we can reasonably speak of someone's "most dangerous" colors, but "true" colors may be a subtly dangerous misnomer.

With this in mind, we can now return to an accidental oversight in Gibson's apology: "There is no excuse, nor should there be any tolerance, for anyone who thinks or expresses any kind of anti-Semitic remark." Do you see the error here? Anyone who *thinks* it? I would love it if no one ever thought an anti-Semitic remark, but for better or worse we have little hope of controlling the pathologies of xenophobia that sometimes infect the alien systems. Most of what we call thinking happens well under the surface of cognitive control. This analysis is not meant to exculpate Mel

Gibson for his rotten behavior, but it *is* meant to spotlight a question raised by everything we've learned so far: if the conscious you has less control over the mental machinery than we previously intuited, what does all this mean for responsibility? It is to this question that we turn now.

6

Why Blameworthiness Is the Wrong Question

THE QUESTIONS RAISED BY THE MAN ON THE TOWER

On the steamy first day of August 1966, Charles Whitman took an elevator to the top floor of the University of Texas Tower in Austin.[1] The twenty-five-year-old climbed three flights of stairs to the observation deck, lugging with him a trunk full of guns and ammunition. At the top he killed a receptionist with the butt of his rifle. He then shot at two families of tourists coming up the stairwell before beginning to fire indiscriminately from the deck at people below. The first woman he shot was pregnant. As others ran to help her, he shot them as well. He shot pedestrians in the street and the ambulance drivers that came to rescue them.

The night before Whitman had sat at his typewriter and composed a suicide note:

> I do not really understand myself these days. I am supposed to be an average reasonable and intelligent young man. However, lately (I cannot recall when it started) I have been a victim of many unusual and irrational thoughts.

As news of the shooting spread, all Austin police officers were ordered to the campus. After several hours, three officers and a quickly deputized citizen worked their way up the stairs and