1

MEMORY

The DNA of Consciousness

With Miles Davis's *Kind of Blue* playing in the background, you focus on the computer screen. The rest of the world recedes as you juggle the numbers in your mind while your deadline looms. Suddenly, your landline rings, the front door slams, and your children race into the kitchen. The dog escapes; TV, video games, and stereo blare simultaneously. One of your printers jams. Your cell phone twitters. You search for some last-minute proofs, while opening an incoming e-mail. Do you have a strategy for what to do and in what order? Can you remember what you were thinking in time to act on it? Is your prefrontal cortex working properly?

Multitasking is a unique prefrontal talent that falls under the general rubric of "working memory." Working memory comprises the mind's intersynaptic DNA, its central operating system for thinking-in-time. Or to use another metaphor, working memory provides the musical notation system from which the higher brain's symphonies are composed. (Or to use another metaphor, it is something like cache memory in a computer.)[1] Yes, the PFC is

the engine of choice, flexibility, decision-making, and foresight. But these functions are built on working memory's underlying action: holding information online. Working memory's dynamic processes drive PFC function across all time frames, at all levels of complexity, and in reasoned and emotional thinking. From the shortest-term memory fragment of remembering a telephone number to calculating advanced physics equations, masterminding a large corporation, or creating a large-scale work of art, music, or narrative, all partake of increasingly integrated levels of working memory to accomplish goals.

The brain bases of working memory have been discovered within the life span of a single generation of scientists, yet the origins of the concept are difficult to trace. The neuroscientist Karl Pribram may have been the first to wield the term in 1960. Or the phrase may have been adopted first by information theorists formulating computer programs for an artificial intelligence process that was an entrée into long-term memory, a short-term memory that is, as one neurobiologist crudely phrased it, "that fragile period when if you . . . hit an animal on the head with a hammer in the first twenty hours after it learns something, it won't remember it." The father of the term "working memory" is the British psychologist Alan Baddeley. In 1974, with Graham J. Hitch, he proposed a remarkably useful, if mechanistic, model wherein sentient behavior partakes of an "executive" function that controls behavior and two "slave" units that hold the relevant information "in mind" and available to the executive. Within this model, "working" was the operational word, stressing its dynamic nature, as opposed to the passive quality implied in "short-term" memory. But for decades most hard-core neuroscientists presumed the working-memory phenomenon was just like long-term memory, only shorter.

"Nobody in neuroscience knew about working memory until I started talking about it in the eighties," Patricia Goldman-Rakic declares in the early days of the twenty-first century. Goldman-Rakic, sixty-six, was killed in July 2003. Struck by an SUV on a street in a Connecticut suburb, she was jaywalking, probably thinking about her work, her PFC not attending to traffic concerns, having shuttled that processing to brain regions where automatic-pilot stuff is relegated. Her death was a stunning shock to the brain science community, and she was openly mourned in ways members of

her tribe seldom are. Her campaign to bring the prefrontal cortex front and center where it could be explored with the full intensity of contemporary science has proven to be a triumph of passion, will, foresight, and determination.

Since the 1970s, Goldman-Rakic struggled to get working memory accepted as research worthy of establishment neuroscience. Virtually all wetware neuroscientists then viewed it as the conjurings of the "soft" psychologists. "Real" neurobiologists focused on long-term memory, the stuff stored in the attic of the mind. Long-term memory, whose essential activity, a kind of "stamping in information for archival purposes," as Goldman-Rakic once dismissed it, was then seen as antithetical to working memory. (Amazingly, even into the early 1990s, leaders in memory research omitted the prefrontal cortex from their anatomical diagrams—it simply wasn't wired in.)

Also irksome was that working memory was too complicated to study with their rather rigid systems. Many memory researchers, she claimed, were actually only studying stimulus-response conditioning—mere forms of Pavlovian training. "But human behavior," she declares to Douglas Stein during one long conversation in the 1990s, "cannot be explained by even an infinitely large set of conditioned responses." Or they confused learning by rote with long-term memory. Those who did study memory typically focused on the temporal lobe, never the prefrontal cortex. One player in particular—she snorted at the name of a well-known memory researcher—"called the medial temporal lobe memory system 'The Show,' the major leagues of memory research!" But long-term memory, she says, "could never ever explain this distinctive quality of intelligence: the ability to use the knowledge you've stored throughout the cortex to modulate your response to the moment." The big boys, she thought, either missed or denied the PFC's crucially human component.

"I was a bit surprised by the resistance at the beginning," she admits. By the 1980s, however, Goldman-Rakic sensed she was headed toward some remarkably unique neural system. "I saw that working memory, this elemental physiological function, is the equivalent atomic basis of all cognitive architecture. I felt we had the very essence of cognition!" Working memory is all about adaptability, she thought. Human behavior quintessentially involves new responses, changing constantly, based on information available at

that moment. This ability to update information from moment to moment is what evolved with the frontal lobe and associative cortices in primates, and is further elaborated in humans: the capacity enabling us to base our choice of actions on experience and knowledge. Knowledge and experience are representations encoded in neurons. One must hold these *internal* representations online to guide behavior in the absence of *external* clues.

Goldman-Rakic wanted to know how individual neurons—the o's and x's of the brain's computing machine—communicated in the PFC. She set up experiments in what is called single-unit analysis, a painless electrophysiological procedure that allows a researcher to record activity simultaneously from numbers of individual neurons with ultrathin electrodes embedded in PFC areas. One records from neurons one by one to understand how populations of them converse with one another.

She and colleagues trained monkeys in various delayed-response tasks, sophisticated wired-up versions of Jacobsen's experiments of the early 1930s. Recording from ten, twelve, or twenty-four separate neurons in monkeys' prefrontal cortices, the team isolated groups of cells that fired during these specific delayed-response tasks. This neuronal activity indicated that the monkeys were storing information about spatial location after the cue disappeared and before they acted. The pattern of firing when the monkey used only a mental representation to guide its response showed Goldman-Rakic just how the neurons did the computation.

She was well aware that her test apparatus was a substitute for cues humans use to access information in working memory. "To you I could say, 'Remember the name of the restaurant you were just in'; 'the last five words I just said to you'; 'the last face you saw in the next room before you walked in here,'" she explains. "But I can't say that to an animal. Our presentation in the lab is a way of providing and controlling information, presenting it briefly, and seeing if the animal can hold on to it in a kind of scratchpad memory. Working memory is what you have in mind at the present moment."

Whatever task Goldman-Rakic set her monkeys to do, she saw that the delay response always worked the same. The common denominator was the inner image of the cue encoded in PFC neurons that remained active after the cue vanished. What once was present in the outside world now only existed in the inside world.

The mind's ability to create an internal representation of the withdrawn image, sound, or thought is central to this process. That representation in the mind's eye could originate either from long-term memory or from something you saw flash by the car window ten seconds ago. Some adaptations for auditory information may differ from visual or touch information, but holding the representation online in the absence of the cue is what's replicated across different prefrontal areas. That holding online of the representation is the basic DNA of mind.

Goldman-Rakic investigated the way PFC pyramidal neurons talked to each other: "Two parties are talking to this neuron, giving it information, and it may be building up its information. A single pyramidal cell can hold on to that information for ten, maybe twenty seconds. But an ensemble of neurons interconnected in a column can keep restimulating themselves, maintaining the conversation, and so may have some emergent quality keeping the information active longer. For a very fast system the time limit is much less. To hold on to the subject of a sentence while you go for the verb must be milliseconds," she speculates. "My constructing a sentence, or comprehending yours, requires a rapid integration."

A pyramidal cell in the PFC, she noted, works differently from one in other parts of the cortex. "A cell in the visual area, say, would just stop firing"—she snapped her fingers—"when the visual image disappears." Unlike the pyramidal cell elsewhere in the cortex, those in the PFC are not stimulus-bound. "That's the secret of cognition," she repeats. "How the cell holds that information? What is the nature of its input and output that give it that unique ability? These are the $64,000 questions."

Why are people's working memory abilities so variable? What are the prefrontal correlates of extraordinary talent? "Every person has delay cells in his prefrontal cortex," she says, "but some people may have wonderful cells that fire for fifteen seconds, during which time that person could integrate volumes of information. Take an arithmetic problem: you've got to hold all this in mind while you keep computing. Some eight-year-old will do all these mathematical gymnastics with incredible ease and accuracy. So if we were to put electrodes into a particular target, like the dorsolateral PFC, Brodmann area forty-six, we might find millions of cells that are exceptionally clean, sharply tuned, that have a capacity to hold on to

information for, maybe, fifteen to twenty seconds. Whereas in another person those same cells will only do it for three seconds. "How working memory functions takes us in a direct line of reasoning to the underpinnings of intelligence. If there is a bell curve for intelligence, there is probably one for working memory capacity. There is a high correlation between performance on working-memory tasks and standard reasoning tests, and why not? Reasoning ability, simply put, is the ability to use representation to guide behavior."

Ironically, in light of the "mindless" accidental nature of her death, Goldman-Rakic once mused about how common, even comic, is the range of low-grade working-memory disorders we all experience. "Your frontal lobes fail you a lot," she reminisces merrily. "I often find myself making automatic responses that don't work. I started parking my car in a different lot, and had to take a different exit route from my office to get to it. Going to the left to get out of this building, which actually is a sort of radial arm maze, is the customary response I had made. Day after day for years I turned left at a choice point. Now I'm confronted by a new choice. Should I go left or right? Well, on three occasions I went left. I absolutely perseverated in my error! This is where the frontal lobe is so important: it overrides the automatic response. I didn't realize until far down that long corridor that I had to walk all the way back. My car was in another spot, but still I, the human with the PhD, was doing exactly what the frontally damaged animal does in the delayed-response task. He can't hold on to what he just saw. Lacking a representational system, he responds with the system that's not damaged."

There must be some neuronal dialogue, ensembles of cells conversing, each playing a role, like the keys in a chord on the piano. Jazz and improvisational piano playing obviously employ dynamic working memory. Playing extemporaneously, a pianist summons ideas from the well of the musical mind, but the self-generated ideas guiding these responses are not fixed notes, an inflexible score, but inner notations that change from moment to moment. The improvising musician never plays a piece the same way twice. Compare jazz to delivering a speech: "When I read the text of my speech," says Goldman-Rakic, "I've thrown a switch in the brain, turned off the PFC, and am using the sensory-guided mode of per-

formance. Certain words are connected to certain vocal responses. But if I put down my text and speak extemporaneously, I'm constructing the speech as I go, and cannot even repeat the last sentence I made, because I have to erase it to move along. Working memory is like a mill; long-term memories are the grist for that mill."

The Perception-Action Cycle

Long before Goldman-Rakic began talking up working memory on the East Coast, Joaquin Fuster at UCLA had already published a paper demonstrating a unique firing pattern in the monkey dorsolateral PFC during a delayed-response task. Today a courtly, white-haired man, the Spanish-born Fuster has the quick, efficient moves of an athlete as he carries out research, teaches, and writes books. "I immediately thought the cells we saw were the mediators of working memory," he declares. Maybe the world was not yet ready for the discoveries Fuster and Garrett Alexander made in 1971. After all, Fuster published his findings years before Baddeley brought the first working-memory model to the table. In next three decades, Fuster elaborated his findings, presciently suggesting that this blazing activity in the PFC was part of a complex circuit involving many areas of the brain simultaneously, and several varieties of long- and short-term memory.[2]

Fuster had a unique insight into the special genius of PFC cells. In 1982, by training monkeys to switch between "what" (object) and "where" (location) memory tasks, he isolated neurons that selectively fire during each. That these cells are intermixed throughout the lateral PFC suggests that the PFC infrastructure integrates memory of object identity and location at the cellular level. He says, "In the prefrontal cortex, representations are highly idiosyncratic, very much related to one's experience—and therefore highly variable from one individual to another."

From the 1970s, Fuster increasingly realized that some prefrontal cells are multipurpose, their activity neither job-specific nor restricted to one sensory modality such as vision. These PFC neurons might instead fire to perform a variety of tasks, each calling for a mix and match of the senses—motion and object seen; sound and color; object and sound. Monkeys trained to associate a visual cue

with an arm motion had neurons that fired only in the combined image-arm-movement context. Animals trained to associate a tone with a color showed prefrontal neurons that fired only to integrate both auditory and visual dimensions. Such cells, which treated the two streams as one twinned representation held in mind, are exclusively cross-modal. Audiovisually tuned for a specific set of sound and color, they perpetuate information about the "whatness" of this "soundcolor." No wonder we demand soundtracks with our video games, visuals with our MTV.

Fuster further saw that PFC neurons could form, dissolve, and reform their associations depending on the context. They were not committed to a discrete motor plan or sensory modality, or even the same polymodal associations. One could imagine these multiplex pyramidal cells as versatile freelance consultants—doing work for whoever called them up. Promiscuous even. "This was an entirely new concept when we first announced it. Activated ad hoc, yes!" Fuster exclaims. During that Stone Age stage of PFC research, Fuster alone understood this compelling characteristic of PFC neurons—that they can hang together to do a task, then disband or form other affiliations. The notion that prefrontal neurons are uniquely polymorphic powerfully influenced the next generation of PFC explorers.

The genius of the PFC neuron is even more impressive in its role as bridge over time. The PFC owns time. Working memory is essential for the "execution of successive acts in a structure of behavior over time," stated Fuster, seeing how PFC neurons fire during the space between the stimulus and the response, during the temporal gap when you memorize a telephone number and you punch it into the phone. To organize your actions, you need a neural mechanism to integrate them across time. If now this, then later that. If earlier than that, then now this: cross-temporal contingencies. This for Fuster was the unique prefrontal factor: he called it "temporal integration," marrying past and future across the gulf of now. By the early 1990s, Fuster had concluded that one set of PFC neurons are predictive, prospective cells that look to the future, while others are retrospective, looking to the past.

From this arises a third temporal dimension: the "memory of the future," as the Swedish neurobiologist David Ingvar named it 1985, the "I remembered that I plan to visit her tomorrow . . ." template.

A recollection of the plan to be executed can be as essential to one's sequence of actions as the straightforward preparation to act. Memory of the future is an extension of working memory at its existential bedrock. That is, no matter what the world's chaos, my own internal distractions or physical perturbations, I continue to remember what I intend to do tomorrow, next week, next year. This steadfastness of mind is embedded, for example, in the Latin verb conjugations of the future past participle, generally translated as, say, "I will have done [this thing] before sunset." Steadfastness of mind is something that characterizes the leadership of such heroes as Winston Churchill and Nelson Mandela.

Another prefrontal function is the role of attention over time. Obviously, there is an intimate relationship between working memory and attention. But what is it? Before scientists began picking it apart, "attention" was perceived tautologically: attention was attention. But to Fuster, attention comes in three flavors, quarks of attention. One is focus: the batter focuses on the ball as it leaves the pitcher's hand; in the airport you locate on the departure screen your flight number, gate, and time, and attend to that as you get a newspaper and coffee and go through security. Attention as keeping a representation of a sensory percept zeroed in over short time periods.

A second form is effortful attention. This is dedication, the drive that compels a person to persevere, keep striving, maintain discipline, and keep his eyes on the prize. It can be inextricably bound up with motivation, will, and desire. Attention with a capital A; attention over the long haul. The third attention is exclusionary, inhibitory. It repels the continuous sensory barrage to which the brain is exposed, and runs interference against distracting thoughts, and inappropriate behaviors and remarks. This attention overrides the habitual old groove that is such an effort to break out of—Goldman-Rakic's automatically walking to the old parking lot. Inhibitory control is absent in babies too undeveloped to curtail reflexive arm motions even when they want to; they lack motor inhibition. When brain damage to the orbitofrontal PFC causes the loss of this attention, primitive drives and emotions can gain the upper hand over reason and social conventions. As we will see, inhibitory control in cognition may be a prime indicator of IQ.

The prefrontal cortex's role in organizing action and the

"attentions" over time led Fuster to see that different categories of information in multiple brain areas play in working memory's big show. "Cells in many other areas of the cortex showed characteristics we'd also seen in PFC memory cells. We thought the entire cortex worked in concert, if you wish, for this form of active memory," he says. Thus Fuster was devising a theoretical model in which all memory, sensory impulses, and planning are interwoven in a giant, cyclical feedback engine he called the "perception-action cycle." Working, or "active," memory, as he calls it, is just one element in the big picture.[3] The PFC is the summit in the hierarchy of structures that form this perception-action cycle, integrating multiple inputs and outputs from many brain levels, translating them into actions that in turn produce changes in the environment, which are then perceived and analyzed in the posterior cortices and once again fed back into the PFC. Fuster's is an architecture of circularity: feedback and feedforward at every level in the ceaseless stream of reciprocal neural processing from spine to brow.

Surprisingly for a bench scientist, Fuster slides easily into the philosophical implications of his system. First of all, he says, it exorcizes the homunculus, the gnomelike puppeteer in the center of the brain. Instead of a miniature boss-operative, there are many semiautonomous agents processing lots of information. "So it goes in a cycle in which there is no true origin, and therefore no need for a center for initiation of actions," he explains. "Because initiation of actions is a factor of the competition of small stimuli acting at the same time, many of which we are not conscious. So what you have is a statistical decision, a summation of impulses that we are not aware of, and to the extent that we are not aware of these stimuli, we feel free!"

So is free will, then, an artifact? Is self-determination merely the end result of summed computations, a calculus of neural events, or consensus voting among tiny unconscious impulses? "No, not an artifact," he replies, "but free will is a by-product of something which is to some degree deterministic." Since the work of the mind is unfolding in a statistical manner, Fuster thinks, stimuli soliciting an action are fiercely competing at any given moment. So the PFC acts as arbiter, awarding the stronger, winning impulse with conscious attention and intent to act. We may be aware of our intent to act, but not the vying neural competitors at work behind that

awareness. We will see this idea of PFC "bias" expressed in the constructs of other investigators as they deploy computer models of the PFC to understand its special genius.

Gradients of memory, then, constitute a relational code emerging from the combinatory nets of neurons in the "grid" that fire and wire together to build a unique and dominant representation that biases and influences all statistical events over time. In this sense, Fuster says, one might consider one's "self" to be embodied in this one-of-a-kind web of neuronal relationships that fire together more frequently than other possible firing webs, and thus become the dominant web of neuronal relationships. The PFC's role is to manage the integration of such competing actions—outward movement, speech, and inner thought—over time. This is one hell of a model and it has inspired work on various levels of scrutiny. One scientist to pursue several of Fuster's observations is Earl Miller.

The Rules of the Game

In 2002, a reporter for German public radio separately interviewed Earl Miller and a professor at Harvard Law School for the same show. The radio producers then surreptitiously spliced both men's taped statements to create a phantom debate. So later, Miller was somewhat taken aback to hear in the midst of the fake face-to-face confrontation the law professor complain that scientists like Miller "think every mental state is attributable to a brain state!" as if this were a dangerously subversive idea.

"I was talking about executive control, and how information about rewards and rules encoded in the prefrontal cortex can lead to rational, goal-directed behavior," Miller reflects. "Then they interplayed this lawyer raving that it was all 'poppycock. Blah, blah, and that's what's wrong with neuroscientists is they think everything has a correlate in the brain, which leaves no room for free will, and if there's no free will, there can be no law! Because law is all about choice—choosing to be good or bad.'

"Well, of course mental events have correlates in the brain!" Miller blurts. "Unless you believe the mind is separate from biology somehow. And more to the point, neural correlates in the brain do not banish free will at all!" Given that there are so many ways of achieving a goal in this world, we need a brain that doesn't lock us

up into one path of behavior. Free will is intact, precisely because every mental choice has a correlate in a brain that has the flexibility to confront these choices. "One has the responsibility to choose among them—this is what free will is all about. This is the essence of free will." Unwittingly, German radio chose a neuroscientist who in barely a decade has fit several significant pieces into the puzzle of how complex volitional behavior emerges from interactions between millions of neurons in the PFC. With a kind of relentless logic has Miller offered up one blockbuster experiment after another.

Now forty-three, as the Picower Professor of Neuroscience, Miller has a fistful of awards and his own laboratory at MIT. While his Web site photo shows him as a demonic figure backlit in a fiery red light, head shaved and do-ragged, face goateed, eyes burning embers, Miller is no iconoclast in his experimental techniques. He utilizes the same method of exploring the PFC as Goldman-Rakic and her mentor Walle Nauta: sending ultrathin electrodes into the lateral PFCs of monkeys to record firing from hundreds of individual neurons at once. But for this Miller takes a random approach. "We don't search for neurons that are engaged in the task, we just drop our electrodes down and record anything we find," he says.

Ear-stud bling and pirate beard aside, Miller is a precision thinker; the word "exactly" peppers his conversation. When he arrived at MIT in 1995, human imaging studies were beginning to take off, but he chose to stay with the old, uncool electrophysiology, because for examining the secret life of neurons, no imaging system was exacting enough. "Single neurons are the basic level of coding," Miller says. "I wanted to stay at that level because I'm interested in knowing exactly how information is processed and understanding the details of neural mechanisms that underlie executive control."

In 1997, Miller presented proof of Fuster's polymorph PFC cells. He and his team taught monkeys to mentally integrate the arbitrary relationships between objects pictured on a computer screen and their locations. When he recorded from almost two hundred neurons in the monkeys' lateral PFCs just after the images had vanished from the screen, he saw that many cells fired for a composite "what-and-where" construct in the monkeys' memory.[4] Such neurons are analogous to those in people that encode the memory of exactly what that golf ball nestled in that particular patch of rough looks like.

Miller also found that if the task only required the monkey to remember where the object was, the PFC neurons fired only for location. If the monkey needed only to recall the object's appearance, the same neurons fired for that image. The properties of many of the cells switched back and forth depending on the job requirement, suggesting, as Fuster proposed, that PFC neurons could change their tune depending on the score. About 50 percent of the neurons were cells that encode "what-and-where" relationships. "When object and location information are used together, as is typically the case in the real world," he says, information about these attributes converges in the PFC. "What-and-where" cells were the initial confirmation of Miller's growing belief that executive processes depend on the PFC's ability to fuse in one's mental universe uncommon relationships between disparate things.

"We are always figuring out relationships through experience and putting them together into a little model, sets of rules, logic or principles as needed to guide us through various situations," he says. How we do this "figuring out" depends on our capacity to forge from among wide varieties of information and mental representations the relationships that are new and arbitrary, relationships that evolution hasn't had time to program into our brains. Miller also saw that these neurons are distributed in both the dorsolateral and ventrolateral PFC (see figure 1 on page xi). "There may be gradients in the PFC where the ventrolateral is more 'what' and the dorsal part is more 'where.' But there is lots of overlap. And this is crucial, because it's the overlap that allows the prefrontal cortex to put together these arbitrary contingencies we need to learn new behavior."

Miller's discovery spurred him on to further challenge "temporary-storage unit" models of PFC function in which the brain's discrete sensory systems—vision, touch, hearing, and other parts of the posterior cortex—provide the PFC with raw material for short-term processing. This is the idea that the back brain "comes up with an answer," Miller puts it. "Then it's simply shoved up to the PFC and held online for a few seconds. We're showing that the PFC does something more, that it actually constructs the relationships needed by complex behavior."

To pursue this idea further, Miller's team then trained monkeys

to remember pairs of associated images. An image of a house, say, would be paired with a picture of a flag. "If I tell you to remember a house and flag, and I say 'House,' you're supposed to remember to say 'Flag.' The monkey was doing the same thing but in the visual domain," Miller adds. Showing the monkey the house image, the investigators detected a rising activity in PFC cells that reflected not the house the monkey just saw but the flag image the monkey was anticipating seeing.

Clearly then, the PFC doesn't just receive inputs from back-brain visual systems and hold them online, but plays a command role in selectively extracting them from storage chambers and loading them in anticipation. Prefrontal cortex neurons generate prospective codes that allow us to prepare for events to come. "The PFC can play a role in anticipating things," says Miller, "and anticipation is what voluntary behavior and executive control is all about. You anticipate achieving some goal—preparing a fine meal or graduating from college—and yet you must be able to come up with the plans to achieve that goal." In the real world, sought-after goals are rarely achieved moments after we conceive of them. When we decide to go to the beach, we may realize we need our sunglasses. We have to recall (mostly unconsciously) what they look like and (more consciously) where we last put them. This ability to recall stored information in anticipation of its use, this prospective memory, Miller showed, involves PFC neurons that code for the "memory" of the anticipated, the expected but not yet occurring reality.

People pull up prospective codes for things that are not part of their actual remembered future: fantasies of winning the lottery or a Nobel Prize, acquiring a Lamborghini, conducting the Berlin Philharmonic, hitting a grand slam home run in Yankee Stadium. Different cues will elicit anticipation of delights in an unreal, alternate universe, constructed nonetheless into a powerfully detailed script from a wealth of hyperemotional imagery. With one caveat, Miller adds: if an activity is grabbing your attention now, the fantasy anticipation drama will not be running. Because the PFC is primarily an in-the-now processing unit, it is calibrated for present action or whatever is currently topmost in priority. But, he continues, "If you are not doing anything important, there's always gonna be this mode the PFC is in—anticipating things."

After isolating PFC neurons that hold arbitrary but convergent

points of information, Miller's group next discovered neurons that encode rules. "Let me tell you a little about the back-and-forth cell," he says eagerly. Miller's lab taught a monkey simple rules using sets of pictures—analogous in human terms, say, to stop at red, go at green. The investigators picked new sets of pictures each day, so the monkey quickly had to relearn which picture meant go right, which meant go left. Not only did 40 percent of the lateral PFC neurons they recorded come to represent these contingencies, but a neuron only responded when the picture A meant, say, go right but not when A meant something else. Or only responded when B meant go left and not something else.

And it took the monkey only ten minutes to switch the rules. In that time, the neural activity in the PFC changed to reflect the rule changes. Here was evidence of the rapid-fire plasticity of a PFC rewiring itself to integrate relationships about a "seeing then doing" rule from information that is processed largely in separate systems in the posterior brain. And doing it with minimal training. The PFC neurons were showing off their agile, quick-break abilities to get a new rule into play.

People constantly learn arbitrary relationships, rules as elementary as stop at red, go at green. We are not born knowing the rules, but pick up protocols to play whatever "game" to maximal effect. Dining at a restaurant is one of Miller's favorite examples. You know the rules: how to access the menu, choose your drinks, order from servers, pay the check, and tip. While memories about dining in one particular restaurant on one particular evening are probably stored elsewhere, the PFC extracts the general features of previous restaurant visits and procedures to give you a general set of behaviors for eating out tonight. And it alters these rules so they can be customized for a bayou-side catfish joint or a four-star Chez Something-or-other.

This experiment showed PFC cells encoding concrete rules, where the rule is always tied to a specific stimulus—red means stop; green says go. Miller next sought the neural correlates of more abstract rules. Humans, and perhaps monkeys, engage in behaviors where the rules are more free-floating. A human calibrates his judgment and embarks on a course of action based on such concepts as "truth," "justice," or "fair play," even though they're not tied to a concrete agent. Would PFC neurons encode for these rules as well?

In this experiment, monkeys viewed two pictures, one after the other. If the "same" rule was in effect, the monkey indicated "same." Conversely, if the "different" rule was in effect, the monkey had to respond only if the two pictures were different. Miller's group trained the monkeys until they were adept enough to make judgments about "same" or "different" even if they were seeing the picture sets for the first time. Recording from neurons in the lateral PFC, they found that up to 50 percent of them conveyed information about the "same" or "different" rule. In fact, more neurons were concerned with the abstract rules of the game than with working memory. This suggests that rule-encoding and rule-representing is perhaps an even more fundamental PFC function than is working memory.

"The definition of an abstract rule," Miller declares, "is something that can be applied to a new experience for which there are no preexisting associations." The genius for fast, efficient, abstract rule-encoding frees an organism from getting stuck in the same old associations or rote behavior. It permits shortcut learning, enabling a smart animal to maximize his advantage in a new situation—think on one's feet—whether it is an engineer refitting the building codes of a site to the architect's revised plan, or a courtroom lawyer revising her examination style after a witness's sudden revelation during a trial. By their freelance nature, PFC neurons can encode for a virtually limitless numbers of rule-representations.

Continuing to explore the neural substrates of rules of the game, the lab looked at category-making. How do we fundamentally organize objects and experiences—apples versus oranges, raw versus cooked, liberal versus conservative, growth versus income stocks? Actually, what don't we categorize? How does the brain create category boundaries as the landscape of experience changes? Miller found that individual cells, "category neurons" as it were, in the monkeys' PFC become tuned to the concept of "cat" and other cells to the concept of "dog."

What grabbed everybody's attention was the design of the experiment. Miller's team collaborated with his MIT colleague Tomaso Poggio, whose lab created a computer-graphics 3-D morphing design program straight out of the *Terminator* and *Matrix* FX vocabularies. The experimenters took three prototype cats (a house cat, a cheetah, and a tiger) and three prototype dogs (a pointer, a St.

Bernard, and a German shepherd) and, digitally "melting" cat and dog characteristics together, generated animated composite images that were combinations of many possible feline-canine arrangements, from nearly pure cat to nearly pure dog. By blending differing concentrations of cat and dog in series of images, they could vary the "catness" or "dogness" of an image and push the limits on category boundaries.[5]

Watching the image on the screen morph from a cheetah to some indefinable entity, then resolve into a St. Bernard, it was hard to pinpoint exactly when the creature was no longer a "cat" and now a "dog." I was worse at it than the two lab monkeys, but then again I hadn't trained like they did. Working for months, the monkeys, who had never seen a live cat or dog, learned that any image that was more than 50 percent dog was dog; any image more than 50 percent cat was cat. The monkeys had become skilled enough to tell when an image was 60 percent cat and 40 percent dog. Since the program generated many new cat-dog chimeras during the experiment, the animals weren't just rote-memorizing specific image mixes.

Beforehand, Miller wasn't sure what he'd see going on among the PFC neurons. "We knew categories had to be represented somewhere in the brain, because monkeys use category information to guide their behavior. But I thought it was possible, even likely, that we would not find evidence for category representation at the single-neuron level," he admits. "I thought it was likely that to represent a 'cat' category, there might be neurons to encode whiskers, ears, tails, neurons for overall shape. I suspected that somehow, at some high level, all these neurons might respond at the same time to amount up to the category, cat."

That's not what they found. Recording from around four hundred cells, they observed nearly one-third to be specifically category-responsive, those firing to all-cat images until the image morphed up to the edge of the cat-dog boundary; others firing to all-dog images until the image approached the dog-cat boundary. Once the image crossed the species boundary, firing activity changed abruptly. "It was sharp," describes Miller. "One window opening, another closing—just like that."

Cat-category neurons responded to every manifestation of cats. So two cats could look very different from one another and the

PFC cells still treated them as "cat." Or one dog might resemble a Doberman and another a dachshund and the PFC would say they're both dogs. "In the end that makes sense to me," Miller offers. "Because when you walk into a room, you instantly recognize a table, a chair. With enough experience, a category gets encoded on the single-neuron level, allowing you very rapidly, efficiently, and effortlessly to organize and conceptualize the things around you."

When we perceive things noncategorically, objects or events can change gradually, shade or evolve smoothly from one to another. The sharp boundary effect, the sudden switch-off between dog neurons and cat neurons, however, fits our experience of category-making. "So, as with that sharp behavioral boundary where in our minds we know something is either/or, we expected to see some sort of sharp boundary in neural activity. And that is exactly what we found," he continues. Pondering the "street" implications, one might see the beginnings of an explanation for why political and ethnic problems are so intractable. People easily form sharp, arbitrary category boundaries between "us" and "them," categories that are fed by emotional wellsprings to the extent they are hard to unwire. It may be more difficult to break down a category boundary than to build it.

But do category-forming neurons exist solely in the PFC? Another place to check for these cells is the inferior temporal cortex (ITC), that region just above and behind the ear involved in the high-level processing of visual recognition and visual memory (see figure 2 on page x). If you take any neuroscientist off the street who's familiar with memory and ask him where categories are going to reside, Miller says, he'll tell you it'll most likely be the inferior temporal cortex. When Miller and his crew "marched back" in the cortex, comparing neuron firing in the PFC and the ITC while a monkey played the cat and dog game, they saw some neurons that conveyed implicit information about the category, as in the PFC. But just as often, ITC neurons conveyed information about cats and dogs as individual animals. If the monkey viewed two very different-looking cats, ITC neurons might convey information about them being in the category of cat, but also might prefer to fire for certain cats over others.

But that's not all. Yes, long-term memories for abstract cate-

gories of, say, cat families may be stored in the temporal lobe, but they are mixed in with all sorts of other information about fur, whiskers, paws, a grab bag of physical attributes of what individual cats look and act like. Prefrontal cortex neurons, on the other hand, convey the categorical equivalence and ignore differences of appearance. "That's a definition of a category: I can treat a tiger and a house cat as both cats even though they look different from one another. We only see that equivalence across changes in physical appearance in the PFC. I didn't expect to see such striking differences in the two areas," he exclaims. "And that the final level of abstraction only takes place at the level of the PFC."

A PFC/ITC category-generating network makes sense operationally. Information about the physical appearance of things is fairly immutable, fairly hardwired in the temporal cortex. Individuals always look pretty much the same, with a little variation over time. But categories are more ephemeral. "I could have a transportation category," muses Miller, "and can instantly generalize upon and modify it to include a new form of transportation, such as the Segway scooter, that two-wheeled 'human transporter.' High-level abstract categories and concepts need to be more dynamic and fluid."

If categories were stored back in the temporal cortex, along with information about the physical appearance of the Segway scooter, one would need information about its motor, wheelbase, steering mechanism, and battery connections stored alongside the details of every possible kind of transportation. By waiting until the last possible stage of processing to encode the abstract category, you can be much more open-ended with what you regard as a member of that category. It's a brain being efficient after millions of years of evolution.

Another question needed to be answered: that, having evolved on a savannah where dog and catlike predators roamed, did the monkeys have some "genetic memory" of cat and dog? To test the genetic memory hypothesis, they reassigned the cat and dog stimuli to three new arbitrary categories that had nothing to do with cat versus dog. "The monkeys learned these arbitrary categories just as easily as the original cats and dogs," says Miller. "Further, all the PFC shifted to reflect the new arbitrary categories. This is a strong hint that even the original categories were learned and arbitrary to the monkeys."

Miller replaced cats and dogs with numbers of dots—abstract entities. The number experiment, dubbed rather inaccurately by the media "Monkey See, Monkey Count," involved another Herculean training and design effort that included preventing the monkey from cheating. The animal might simply memorize all the possible combinations and patterns of images. With a hundred versions of each number-dot picture—five hundred new stimuli every day—the monkey couldn't possibly memorize them all.

Judging the relative quantity of low numbers is highly adaptive. Many animals do it. It's a way of quickly categorizing and making sense of quantity. The task was limited to no more than five numbers, because after five or six items the monkey's ability to categorize number quantity drops off. In fact, without the verbal encoding power of counting 1, 2, 3, 4 . . . 100 . . . 234, and so on, a human's ability to conceptualize number quantity is not much better. If humans are prevented from verbally counting, they show the same drop-off in ability to measure quantity at about 5 or 6. And despite the headlines, monkeys can't count.[6]

Recording from lateral PFC neurons, the investigators again found that about a third were tuned to a specific number. Firing intensity, furthermore, progressively declined as the number of dots moved away from the neuron's "favorite number." Overall, the neural patterns seemed to form a "bank of overlapping filters." The neurons "knew" that the quantity of three is closer to four than it is to one. "The results from our cat and dog and number studies are remarkably parallel. The final representation of the abstract concept seems only to come to fruition in the PFC," says Miller. But there were subtle differences between number and cat/dog processing.

"We did the number experiment because numbers, although very abstract, are an example of a genetic memory," says Miller. "Many, maybe most, animals can make small number judgments without explicit training. What we found recently is that the innate memories for small numbers seems to be stored a small area in the parietal cortex and is then 'loaded' into the PFC when needed. This is in contrast to the cat and dog categories, which did not seem to be explicitly represented in sensory cortex and underscores the importance of the PFC in learning arbitrary rules."

Your Inner Proust

How the PFC distributes working-memory computations across its various territories is the subject of much research. The ventrolateral PFC is "like day and night to the dorsolateral," asserts Michael Petrides, the director of the Cognitive Neuroscience Unit of the Montreal Neurological Institute at McGill University. Petrides, among others, proposes the existence of "multiple executive processing modules" within the human PFC. With the arrival of functional magnetic resonance imaging (fMRI) techniques in the mid-1990s, he began digitally capturing the contributions of these different subsectors, and devised a model of the lateral PFC as a two-stage mental processor with mastery over the flow of experience during time.

For him, there is a "looming dichotomy" between what the "upper" and "lower" lateral prefrontal areas do. "I'm claiming there is a big chasm—as big a divide as between North America and South America—where dorsolateral areas do something predominately different from the ventrolateral ones." In Petrides's model, the ventrolateral PFC, Brodmann areas 47/12 and 45, constitute a kind of search engine for retrieving specific memories and data from archives in the posterior brain, particularly when the information is embedded in an ambiguous context or interleaved with many other memories. When we need to recall something in particular—Who was that man at the party?—the mid-ventrolateral PFC is recruited to the search.

A person with damage to the ventrolateral PFC has lost this capacity to initiate an archival search for that one piece of information—where the mind serves as a heat-seeking missile targeting the exact name, number, image, or idea. "If you are a patient with frontal damage," Petrides says, "you are not amnesic. You're also just as smart as anybody else, but you make mistakes; you fail to retrieve information, not because the information is not there, but because you lack the basic executive processes that enable you to go back into your memory traces when the retrieval is tricky and highly ambiguous."

This active, precision-targeted recall is distinct from varieties of remembering that do not require the ventrolateral PFC's talents. Petrides continues, "If I meet you for the first time, then later run

into you, I will recognize you. My posterior visual, auditory, and multisensory areas will process this information, as long as they have good interaction with my hippocampus and other limbic areas. So if I see you again, similar images will be reactivated in posterior cortical areas, and they will be sufficient for me to recognize you. For this I don't need my frontal cortex." In people suffering from damage to the ventrolateral PFC, this passive recognition memory system is usually intact.

This nonprefrontal, associational memory system can be activated by a single, momentary concordance. "I could be having a nice martini somewhere in downtown Montreal," Petrides goes on, "and I start thinking about a particularly strong experience of a meeting in Colorado where we had a great fish dinner. I then immediately remember that you were there. Or I might have been watching TV and there was something about Denver in the news. That immediately takes me back to the meeting, and I remember the restaurant where we had the fine seafood dinner and what great fun it was, and suddenly the whole image springs out in my mind of you and others sitting around the table. That is what happens most of the time. Our memories are being reenacted, retrieved because one sees things again, and so new traces touch old traces. One association triggers the second and so on.

"And yet, as soon as someone asks, 'Was she at the Colorado meeting?' in this active kind of memory, I can't merely let memories link to other memories. I initiate a search of my memory traces for the specific pieces of information I want." The strength of an active initiatory memory system subserved by the ventrolateral PFC and its networks has implications for the contours of the individual self. Take Marcel Proust in *À la recherche du temps perdu*. Obviously Proust, intensely attuned to his associational memories, also had an extraordinary search engine for his archives of lost time. Proust's "faithful guardians of the past" could roam his back-brain libraries for "texts" of amazing specificity, and retrieve them in consummate detail. Such gifts are invaluable to any art or calling. Take the physician who, with scant evidence of symptoms, can search for and summon up the probable diagnosis of a exotic skin disease based on data about it stored in her long-term memory.

Your Inner Palm Pilot

On the other side of Petrides's lateral prefrontal divide is the mid-dorsolateral PFC, Brodmann areas 46 and 9. As the PFC's executive suite, the dorsolateral is a specialized place where memories and events, once stored and interpreted in the posterior cortical areas, are now summoned to be sorted, monitored, and recalibrated. The mid-dorsolateral PFC's job is to attend to and manipulate the status of mental events on our assembly line—actions we intend to take, plans we expect to execute. It is home base for our "memory of the future."

"At any moment in our lives many things compete for our current awareness," Petrides explains. "A person might be holding online a number of relevant events and cannot afford to ignore any of them, but must monitor them all. He must keep track of which events have happened, which events are yet to happen." This manipulative ability gives us tremendous flexibility. As the beginning of strategic planning, it is essential for making creative designs. "I wake up and set up six or seven intentions," says Petrides: "'This morning I have to call Mrs. X, or be in my office at three o'clock, when she will call. I also have to make sure I contact those four others about the party today.' In the mid-dorsolateral area, neurons have coded these intentions. So as I go through the day and do many different things, those neurons continue to code them."

After he calls Mrs. X and checks off that as "done," his mid-dorsolateral PFC recodes the neurons. The mid-dorsolateral PFC thus reorders a series of events being held online. "How can I do any manipulation, if I cannot hold those relevant six things in my mind, if I cannot say A, B, C, and D are relevant components?" Petrides asks. "And that A has moved down to B's place in the list, and now C is in A's position? I must have a mind capable of attending simultaneously, keeping track, and prioritizing multiple representations."

People with damaged mid-dorsolateral PFC areas fail at keeping track—whether of objects in a sequence, numbers, or abstract ideas. These patients do not lose their capacity to speak or remember, but inevitably their lives collapse. "They are intelligent on educational tests, but not street smart," says Petrides. "To be street smart means that right now you are attending to one thing, while at the same

time there are little lightbulbs keeping track of the three other things that also need to happen. So at the appropriate time you can quickly turn off A and turn your attention to B and C. Patients with mid-dorsolateral PFC injury cannot do that kind of prioritizing or self-ordering." They cannot easily get out of groove A. This dorsolateral ability no doubt evolutionarily preceded techniques humans have invented to enhance this executive function: from the invention of writing itself, to its simple subrule, taught in grade school, to "outline" our papers. It is the mental construct behind, perhaps, everything from the Dewey decimal system to esoteric electronic stratagems for organizing vast amounts of material along space and time continuums.

Area 46 may be an organizational sector. But is this dorsal/ventral dichotomy as clear-cut as Petrides has postulated? The Cambridge neuroscientist Adrian Owen, a sometime Petrides collaborator, agrees that the mid-ventrolateral PFC retrieves specific memories. "If I ask you to remember the number 7946, the ventrolateral is involved in retrieving that number later on," he says. "But if I ask you whether there were any *even* numbers in that sequence, then you must introspect, work though the contents of your memory and decide, yes, there were even numbers: four and six." This, Owen thinks, requires the participation of the mid-dorsolateral PFC.

In imaging studies of the frontal lobe, Owen has tried to "really crack open this ventrolateral/dorsolateral thing." He now thinks a "gradient" exists in levels of processing complexity, with a gradual change from more basic memory processing at the ventrolateral level to higher-level processing at the dorsolateral stage. "The dorsolateral PFC," he suspects, "is involved in identifying potential strategies to facilitate, or make memory most efficient. The dorsal/ventral distinction, then, would be for me now, one of levels of abstraction." So then is the dorsolateral PFC involved in computing more intentional, and therefore conscious, processes? To Owen conscious versus unconscious may not be real distinctions. "When somebody is looking at the contents of their memory, are they aware they are doing it? Or with higher-level thinking, when you figure out a way to approach a problem and then set about doing it, how aware are you of your scheming? Who says to himself, 'Well, now I'm going to start strategizing . . . '?"

In a test of shape memory, Owen used abstract designs rather

than familiar shapes, to avoid a situation where people would say, "I remembered the square." This situation is not unlike Miller's need to keep the monkeys from "cheating" at number tests. Yet the subjects reported afterward that they remembered the shapes nonetheless, by creating strategies such as "more shapely" or "less shapely." They were not aware of creating schemes to facilitate recalling these abstract images. On the scanners, Owen saw elevated dorsolateral PFC activity. So, he argues, this region may serve "to identify order in the world. It says, 'Yes, I can use shapeliness to facilitate memory!' This strategizing is not something we necessarily do in a conscious, self-motivated way, but it is the way this brain region is set up to maximize effectiveness."

Owen also has "some really nice data" to foil Mr. Homunculus. "This model avoids that little brain-within-a-brain problem in the sense that the dorsolateral PFC relatively automatically identifies high-level structures in the information it is processing." Owen suspects this automatic-ordering faculty is almost always based on past experience. A square is always a square with certain geometric properties, but that's not true of all objects we see and think of as shapes. We may try to organize new shapes to fit into categories we've seen in the past or are familiar with. Owen's idea—that there is an innate bias toward ordering the novel and random flow of the external world—concurs with Miller's category-building neurons. Just how we compose these organizational strategies is highly idiosyncratic, intensely personal.

Take a chess player doing a spatial memory test: this person will tend to refer to objects in space in terms of chess positions. "The chess player is well practiced at spatial thinking. But I'm not a chess player," says Owen, who is in fact a lead singer and bass player in the rock band YouJumpFirst. "I don't see chess positions anywhere in the world! Now, if both of us are looking at the same spatial problem, our visual cortices will do exactly the same thing. But the 'strategy' we'd bring to bear on that problem would be entirely different. Because people do organize reality differently, it's difficult to find specificity in the prefrontal cortex. We talk about PFC function in terms of 'manipulation' and 'monitoring' strategies, but what are these? We all in the field think we know what we're talking about, but none of us actually believes that this is solely what this region does."

But when does the PFC "know" that it needs dorsolateral intervention in order to conduct a more complex manipulation? And is there a limit to the strategic operations it can keep online at any given time? The Stanford psychologist John Gabrieli, using fMRI to explore working memory, was struck by how even modest tasks, such as remembering a string of letters or digits for a short time, engaged huge portions of the PFC. He was also impressed by the limitations of working-memory capacity—that it can only hold around seven bits of information at once. This basic unit capacity is actually less—more like three or four bits. But you can manipulate a couple of these units to add up to the "magic seven." "You're kind of saying to yourself: 'those 3' and then 'these 4.' You are juggling two things, and that suggests it doesn't have to be that hard a task before a lot of the dorsolateral PFC is involved," he says. Perhaps that's why phone numbers are generally seven digits, plus an area code.

"You begin to wonder what's going on in more complex executive operations. When does a quantitative thing become a qualitative thing? Once you get past about three items," Gabrieli adds, "it's as if your brain says, 'Okay, now I need to turn on this other computer.' That is true of almost everything in life. If you carry one or two shopping bags, it may seem as if the third is just another one. But that's when you start dropping things. What difference will one more make? It's a qualitative increase; at one point it becomes the final straw. If you have to manage enough information at once, it's simply that managing it becomes a dorsolateral executive process."

The Brain's Conflict Monitor

A growing consensus thus implicates the dorsolateral PFC as involved in identifying potential strategies to facilitate working memory. But "who" alerts the PFC to summon its special talents? To explore this key issue we need to visit another subsector of the PFC, the anterior cingulate cortex. Before that, however, we need to acquaint ourselves with the Stroop test.

It's worthwhile to stop and admire the Stroop test. Cited and applied thousands of times during the past seventy years, it is the classic examination of attention and lack thereof in the prefrontal cortex. No one can talk about working memory and executive con-

trol without sooner or later encountering the Stroop. It works this way: a player is presented with words for colors (e.g., GREEN) printed in either the color the word indicates, or another color (e.g., RED). On command, the participant must either name the word, or disregard the word's meaning and name the word's color. What's so remarkable is that when the print color differs from the word's meaning—if, say, the word GREEN is printed in red ink—a person takes longer to say "red" than he does to say "green." To name the word GREEN as "red," the person fights to inhibit and suppress the stronger tendency to say "green." Or he makes an outright mistake and says "green." "Scratching the itch," as the neuroscientist Jonathan Cohen put it. The error is called "Stroop interference," or simply "the effect." Others have likened trying to name the word's color to wading knee-deep in mental sludge.[7]

We are programmed to read words for their meaning. Thus when asked to suppress this response in order to focus on a word's color, our minds balk at this violation of what we "always do." Thus the Stroop neatly demonstrates a core function of executive control: the ability to override a strong but wrong signal to select a weaker but right one. Patients with PFC impairments, including attention deficit problems, schizophrenia, and various injuries, struggle with the Stroop. The Stroop is sensitive to subtle changes in normal brains as well. Fatigue, loss of sleep, minor brain damage, and strange environments, such as high altitudes, increase one's error rate and the time it takes to name a word's color. To test mental flexibility, the Stroop has been given to people in all sorts of extreme states, including climbers nearing the 8,000-meter mark on Mt. Everest.

The Stroop has escaped the lab in other ways as well. Recently it was programmed into the MiniCog, a little handheld electronic device, used by NASA astronauts. Its developers claim that corporate strivers, as well as space walkers, can check on their prefrontal CEO abilities at any anxious moment by seeing how they score on the Stroop. A Web site advises stock market day traders to practice the Stroop. Since they face an "oppressive opponent within their own minds," the ad warns, they can better cope with the constant bombardment of distracting external stimuli by practicing the Stroop. You will learn to better "filter what your brain deems unimportant, based on criteria you have given it." John Ridley Stroop

published his invention and its first test results in the *Journal of Experimental Psychology* in 1935, the same year Carlyle Jacobsen announced the results of his chimp studies. Compared to Jacobsen, Stroop's name and experiment is far better known.[8]

In 1986, Jonathan Cohen, now director of the Center for the Study of Brain, Mind, and Behavior (as well as codirector of the new Institute in Neuroscience) at Princeton, was one of an exotic breed of young researchers captivated by the potential for applying connectionist computer modeling to the neurobiology of thought. At Carnegie Mellon he studied with the neural-net pioneer Jay McClelland. In McClelland's class, Cohen met Kevin Dunbar, whose previous work focused on the Stroop. Cohen vividly recalls sitting in McClelland's office when, Dunbar said, "If this connectionist stuff is so good, we should be able to model this Stroop finding."

"Neither of us had much experience in modeling," recalls Cohen, "but I wanted to try to build a model of the Stroop." In the mid-1980s McClelland and a few others were exploring parallel distributed processing architecture to simulate brain activity. They called these programs "connectionist" because, like actual neurons, their computerized simulated cells communicated with other simplified digital "neurons" in the model to create networks that in turn simulated brainlike behavior.

Over the course of 1987, Cohen and Dunbar went about designing a connectionist model of a neural network that could negotiate the Stroop test. They programmed in two processing pathways: one devoted to word, the other to color information. Both pathways would converge upon command to respond to a task demand: name the word or name the color. Like a human, the model had to select between the two competing processes—word or color. To mimic the human condition, the scientists strengthened the model's word pathway by "training" it more intensively than the color pathway. When they finally ran Stroop simulations, sure enough, the machine performed faster in "naming" the word than it did the word's color. Since in computer modeling everything is modifiable, they reset the program, overtraining the color pathway. Then when they ran the Stroop sim, the machine did better "naming" the color than it did the word.

"Out of this simple neural-net model leaped not only the fact that the relevant strength of the pathways could determine the

speed but that everything was subject to control. Realizing the model could account for the basic Stroop effect, I simulated new learning data, and showed it could account for other findings. It was not perfect. The model had idiosyncrasies and unexplained elements we were not comfortable with. But," Cohen adds, "it provided a conceptual grounding for me."

Around that time, Cohen attended a conference on the burgeoning ideas about prefrontal functions. People were talking about Goldman-Rakic's work and recent findings about working memory. There was palpable excitement about the notion of maintaining information online. Inhibition—the PFC's ability to curtail rote in favor of new behavior—was another theory. "With all this percolating in my mind," Cohen recalls, "I started thinking, 'Maybe the prefrontal cortex is involved in the Stroop effect.'"

In doing the Stroop correctly, you are maintaining in your mind the rule, representation, or strategy. Your brain chooses the desired but weaker interpretation, inhibiting the stronger but undesirable interpretation. You want color, not word. But still, it's ambiguous. "I suddenly realized that the PFC might be sitting there presenting this information, not just holding it online, but using it to literally guide how the rest of the system will perform. And that epiphany was basically ten more years of research!" Cohan admits. He suspected that the PFC was weighing competing representations and judging which among them to give the go-ahead signal. It was not unlike Fuster's idea of competing systems adjudicated by the PFC. But how did the PFC "know" about this conflict in the first place, in order to attend to it and steer the neurons toward the right "goal"? That question led Cohen to a "lower" part of the prefrontal system, the anterior cingulate cortex (ACC).

The ACC (see figure 1 on page ix) is the elephant to the neuroscientists' blind men; everybody's got a slightly differently take on it. The mental operations it putatively engages in include heightening skin, touch, and pain sensitivities—even emotional pain. Thought to be a regulator of positive mood, it has been dubbed the brain's "cheerleader." Some think the ACC functions as the brain's quality controller or "oops monitor." Studies find it more active when a person lies than when he or she tells the truth. When dysfunctional, it may play a role in depression, obsessive-compulsive disorder, anorexia, and attention deficit disorder.

Because of its position and extensive hookups, the ACC, Brodmann areas 32 and 24, has many strategic alliances with other city-states of the brain. It is intimately connected to limbic structures and the autonomic nervous system that oversees heart rate, blood pressure, metabolism, and other of the body's housekeeping functions. "The part of the ACC that lies in the rostral prefrontal cortex," says Cohen, "is clearly involved in higher-level executive functioning. Exactly what that functioning is, is what's really interesting. "In my thumbnail account, this strip of cortex's mandate is taking stock of the system's performance. Internal states are the focus," he stresses, totally oblivious to the jackhammers pounding away in the Princeton psych building's throes of renovation. "The ACC is about looking inward, into whether the thing you are doing now, or about to do, will lead to a good or bad outcome. This information could pertain to motor performance or autonomic inputs—your stomach growling tells you your behavior isn't satisfying a fuel need. It may pertain to how you perceive you are doing on an SAT test or job interview."

More important, the ACC, in Cohen's view, is a chief player in control processes, those neural mechanisms that help the PFC adjust to changing demands; to reconfigure the amount of attention needed to think something through efficiently. That the brain somehow detects and monitors its own inner performance was until recently not a direct object of inquiry. The control aspect was assumed. Somehow the brain "just knew" when to turn up or down its intensity level. But this explanation was dismayingly "homuncular."

Since medical school, Cohen had wondered: what happens when a person decides to attend to "this" as opposed to "that"? At first he suspected that monitoring was a lateral prefrontal job. But he and his collaborators decided to see if other brain areas signaled to the lateral PFC what it should be paying attention to, or how much attention it should allocate. With Cameron Carter, then at the University of Pittsburgh, Cohen developed a computer model of an ACC to strap onto his Stroop-playing virtual PFC machine. The idea was arduous in its development. A breakthrough came when the two men recalled a brain-mapping meeting where the host opened the session with a rhetorical question: looking at the thousands of imaging abstracts submitted, which area of the brain was most active? It turned out to be the ACC. Indeed, the ACC fired up

across many different studies—in response solely to the task's difficulty. The data were hinting that when the going gets rough, the ACC gets going.

Now armed with evidence that the ACC responded in some ill-defined sense to "brain sweat," Cohen and Carter were struck by reports that it might serve to monitor errors. As a task grows more difficult, a person makes more errors. But there was a knotty problem with the error theory. Much of the imaging data found the ACC to be active when subjects performed difficult tasks but made no errors, or no more than when the tasks were easy. The discordance between the error story and ACC activity piqued Cohen's interest. "Brain imaging has often been dissed for telling us things we knew," he says, "and here's a great example of it not only telling us something new about the ACC but providing evidence that forced us to think hard."

Subjects who performed flawlessly on the Stroop test still showed elevated ACC activity. Cohen asked himself, "What's going on in the Stroop that monitors difficulty?" It dawned on him: "Maybe it's not error but uncertainty." It made sense that a region surveilling errors should be most active when the task is most difficult and uncertain. "Difficulty and uncertainty are in part indexed by accuracy, at least your performance is," Cohen says. "Difficulty and uncertainty are what we ultimately came to articulate as conflict. Conflict is what's driving the ACC."

An attractive feature of the conflict hypothesis was that it might exorcize the homunculus once and for all. Conflict, not "My bad!" would be all the brain can know. There would be no "wrong" buzzer squawking, but a kind of neuronal dissonance, when the brain struggles to choose between two or more responses when it can only make one. Cohen was willing to go out on a limb and say, "When something is incompatible in the brain, there arises from the ACC a high activity that flashes 'conflict!'" Useful for gauging when the PFC needs to come online and when it doesn't, the ACC, then, could be the region that alerts the PFC to be alert for "incoming." Or to stand down.

So Cohen went back to plug an "ACC conflict monitoring unit" into his Stroop machine. With then postdoc Todd Braver, he set up the electronic units so that for each Stroop trial the "ACC" would gauge input from the rest of the nets and compute the amount of

prevailing conflict among the "response units." Running the model, the two found that the computer's "ACC" did indeed detect conflict during the color-naming condition of the Stroop test, even when the model performed correctly. Tweaking the difficulty of the computer trials, it became apparent that whenever the model took longer to respond, it was because competition persisted between the alternative responses. This confirmed their hunch that the ACC was monitoring conflict, crosstalk that arose in the model's "word-" and "color-naming" pathways before the model made its response.

A basic tenet of information processing, hammered out by the designers of parallel-processing computers decades earlier, is that a computer program needs excess control in situations where there is crosstalk. Most artificial intelligence (AI) programs have control-alerting functions. If reduction of crosstalk is a primary function of control in parallel computing systems, then a brain, too, might monitor for the presence of crosstalk to know where it needed to allocate control. Such a monitoring signal from the ACC monitor would note either the presence of conflict or that the coast was clear, and quiet down until it detected conflict again, whereupon it would again recruit various degrees of PFC involvement. "The notion of a loop using conflict monitoring to determine how active the PFC should or shouldn't be," says Cohen, "began to make sense."

In feedback-feedforward loops, frequency and time are critical elements. When a person does the Stroop, says Cohen, "you are sitting there, primed before the stimulus comes in, ready to trip the response and quickly say the word's color. The (red) word GREEN appears on the screen coupled with a little bit of noise in the system, which could mean that you got distracted by someone slamming a door down the hall or something. And you utter 'Green!' And you go: 'Oh, that's not what I meant!' And you may correct it on the next trial."

Cohen designed the computer model's response units to "hover" at the threshold, barely below the levels that noise wouldn't trip them over. When the model made a mistake, it was because the stronger support for the word GREEN caused the green unit to respond before an "attention unit" had a chance to kick in to suppress it in favor of the color RED unit. The computer went for GREEN. And the ACC/conflict unit signal appeared.

Over a series of trials, RED response units begin to accumulate "strength," eventually overtaking GREEN response units to trip the correct response. But during an intermediate period, where GREEN still had a minor advantage but RED won out, there was a warning of impending conflict. Did human EEG studies of the Stroop, Cohen wondered, show signs of a brain wave that fired before someone did a correct task, as in his machine? Lo and behold, in the literature was an electrophysiological response called the M2C, a firing pattern evident about 200 milliseconds prior to the stimulus. "Exactly where we predicted is a conflict signal." (Interestingly, one-fifth of a second is about the time it takes for a batter to resolve his conflict about whether he's looking at a fastball or breaking ball and make the appropriate choice to swing or not swing.)

To Cohen, this M2C and the error signal were identical. "We say both reflect the detection of conflict. Conflict precedes response, and if you answer correctly, by definition it gets resolved in favor of the correct response. You've suppressed the incorrect response, end of story. On the other hand, if conflict results in an error, you continue to process information, which in turn leads you to acquire information about the correct response that competes with the previously activated incorrect one. You see the conflict and you correct yourself."

Cohen and colleagues next attempted to simulate how the brain fine-tunes performance to lessen errors and maximize "winnings" over time. They used as a template a long-standing finding called the "Rabbitt effect," so named for the British psychologist Patrick M. A. Rabbitt. The Rabbitt effect is a fairly commonsensical feedback loop: after making an error, you tend to slow down, be more cautious, and so become more accurate on subsequent trials. Then the better you do, the faster you go, and consequently the more mistakes you make, so you slow down. And so on. (Statistics on motorists' frequency of getting speeding tickets may confirm the Rabbitt effect.)

Cohen's critics have noted that the Rabbitt effect was meant only to describe explicit, conscious errors. Certainly, some explicit knowledge of having made an error has nothing to do with conflict—just "I didn't get it right," which leads you to adjust your performance. Cohen, however, contends that the ACC's conflict monitoring yields the more accurate results over time. And it may do so implicitly, unconsciously. Also, in "Rabbitt fashion," people do better when

they have a string of hard tasks than if they have a string of easy ones followed by a hard task. Again the reasoning is commonsense: If the previous task was difficult, you will be more focused and conservative in your approach to the next one. If the previous series have been easy, you are lulled into complacency, slacken your focus, and so make an error.

The question for the Stroop computer was, again, one of feedback. Could the conflict signal alert the PFC sim to be more "on its toes" in the subsequent trials after an error? And slacken off after it seemed the trials were growing too easy? Cohen's group ran trial after trial to determine how much the model had to turn up or down the "alertness volume" to maximize its correct responses. They found that the control loop is sufficient to account for the compensations described in the Rabbitt effect. All your PFC needs to know from its ACC foreman is that it has gotten highly competitive out there in the testing environment, and that last task was a little tougher than it expected, so you should pay more attention next time.

Cohen's model began correcting itself, and veered away from errors. It performed the Stroop as well as a practiced human—except when the investigators intentionally tinkered with its parameters to simulate diseaselike deficits. The neural net was working as a flexible feedback system without human programmers' deus ex machina–like hectoring. Occasionally in its processing of information the sim tripped its digital switches too quickly, yielded the incorrect answer, and the machine registered the conflict signal right after the error. And occasionally, after it made a correct answer, the scientists noted the presence of the conflict signal. Through this conflict-monitoring feedback setup, Cohen saw himself effectively "chipping away at the homunculus."

Cohen and postdoc Matt Botvinick speculated that an expert might predict a person's future behavior largely on the basis of his ACC firing patterns. A period of high ACC activity should be followed by quicker and more accurate responses; low ACC activation, the opposite. Jamming of one's ACC signals, they mused, should disrupt these strategic behaviors; a person would make many errors, behave recklessly. Problems in summoning the PFC to intervene would abound. Are people with ACC defects neurobiologically incapable of detecting they had a conflict or made a mistake, much

less of doing anything to correct it? If so, might "normal" people with sensitive conflict monitors find inexplicable the "lack of control" exhibited by those who commit "errors" such as antisocial behavior and crime? So the uncomprehending person asks, "How could you keep on doing that stupid thing?" Such recrimination seems destined to be pointless if the subject has a disordered ACC/PFC system.

The Stroop-playing computer has no conscious awareness of its triumphs or failures; so do human ACC operations require consciousness? That is, must the ACC's call to "focus, focus, focus" be accompanied by an awareness that one is entering a thicket of conflict, or that one has made an error? "The quick answer," Cohen replies, "would be: not necessarily." The question may be unanswerable, in part because "conscious awareness" of error-making and corrections may be a secondary effect, "a story you're making up afterward 'to explain' how you were smart enough to correct yourself." You might say, "I did badly; I've got to do better."

In testing situations, however, the experimenter can often see these effects outside of people's consciousness. Consciousness can be "epiphenomenology," an illusion, a distortion in thinking the brain creates in order to portray its operations to us. Awareness may have an impact, but it doesn't mean the part of the system that's aware is actually driving the brain. "One nice feature of the model," Cohen states with satisfaction, "is that the ACC gives you a bit of an advantage without a hint of consciousness. Even if you haven't yet made an error, the mere presence of the conflict itself would be a good signal to your higher executive functions that you ought to adjust performance because you're likely to make an error if you leave the status quo.[9]

"That the ACC lies at the interface between the limbic and cognitive systems now makes sense," he says. "The limbic system is all about emotion, about placing motivational weight and significance on external and internal events." If the genius of the ACC is to gather information about the performance part of the system, it would also have to convey information about emotional states to the PFC—the system responsible for integrating feeling and knowledge, and driving motivational states. "The ACC," Cohen concludes, "turns out to be responding in a way we knew some part of the system had to."