Over the last 50 years, research into a chemical called dopamine has revealed how different brain processes work. Now a lab in Houston is linking this single neurotransmitter to social phenomena—from stock market bubbles to the development of trust—and recasting the focus of neuroscience in the process.

BY JONAH LEHRER
END MONTAGUE is getting frustrated. He’s trying to show me his newest brain scanner, a gleaming white fMRI machine that looks like a gurney-toting tanning bed. The door, however, can be unlocked only by a fingerprint scan, which isn’t recognizing Montague’s fingers. Again and again, he inserts his palm under the infrared light, only to get the same beep of rejection. Montague is clearly growing frustrated—“I can’t get into my own scanning room!” he yells, at no one in particular—but he also appreciates the irony. A pioneer of brain imaging, he oversees one of the premier fMRI setups in the world, and yet he can’t even scan his own hand. “I can image the mind,” he says. “But apparently my thumb is beyond the limits of science.”

Montague is director of the Human Neuroimaging Lab at Baylor College of Medicine in downtown Houston. His lab recently moved into a sprawling, purpose-built space, complete with plush carpets, fancy ergonomic chairs, matte earth-toned paint and rows of overzealous computer monitors. (There are still some technical kinks being worked out, hence the issue with the hand scanner.) If it weren’t for the framed sagittal brain image, the place could pass for a well-funded Silicon Valley startup.

The centerpiece of the lab, however, isn’t visible. Montague has access to five state-of-the-art fMRI machines, which occupy the perimeter of the room. Each of the scanners is hidden behind a thick concrete wall, but when the scanners are in use—and they almost always are—the entire lab seems to quiver with a high-pitched buzz. Montague, though, doesn’t seem to mind. “It’s not the prettiest sound,” he admits. “But it’s the sound of data.”

Montague, who is uncommonly handsome, with a strong jaw and a Hollywood grin, first got interested in the brain while working in the neuroscience lab of Nobel Laureate Gerald Edelman as a post-doc. “I was never your standard neuroscientist,” he says. “I spent a lot of time thinking about how the brain should work, if I had designed it.” For Montague the cortex was a perfect system to model, since its in-comprehensible complexity meant that it depended on some deep, underlying order: “You can’t have all these cells interacting with each other unless there’s some logic to the interaction,” he says. “It just looked like noise, though—no one could crack the code.” That’s what Montague wanted to do.

The human brain, however, is an incredibly well-encrypted machine. For starters it’s hard to even know what the code is. Our cells express themselves in so many different ways. There’s the language of chemistry, with brain activity measured in squirts of neurotransmitter and kinase enzymes. And then there’s the electrical conversation of the cortex, so that each neuron acts like a biological transistor, emitting a binary code of action potentials. Even a silent cell is conveying some sort of information—the absence of activity is itself a form of activity.

Montague realized that if he was going to solve the cipher of the mind, he would need a cryptographic key, a “cheat sheet” that showed him a small part of the overall solution. Only then would he have any chance of connecting that chemistry to the electricity, or understand how the signals of neurons represented the world, or how some system of cells caused human nature. “There are so many different ways to describe what the brain does,” Montague says. “You can talk about what a particular cell is doing, or look at brain regions with fMRI, or observe behavior. But how do these things connect? Because you know they are connected; you just don’t know how.”

That’s when Montague discovered the powers of dopamine, a neurotransmitter in the brain. His research on the singular chemical has drawn tantalizing connections between the peculiar habits of our neurons and the peculiar habits of real people, so that the various levels of psychological description—the macro and the micro, the behavioral and the cellular—no longer seem so distinct.

What began as an investigation into a single neurotransmitter has morphed into an exploration of the social brain: Montague has pioneered research that allows him to link the obscure details of the cortex to all sorts of important phenomena, from stock market bubbles to cigarette addiction to the development of guilt. “We are profoundly social animals,” he says. “You can’t really understand the brain until you understand how these social behaviors happen, or what happens when they go haywire.”

And yet even as Montague attempts to answer these incredibly complex questions, his work remains rooted in the molecular details of dopamine. No matter what he’s talking about—and he likes to opine on everything from romantic love to the neural correlates of the Coca-Cola logo—his sentences are sprinkled with the jargon of a neural cryptographer. The brain remains a black box, an encrypted beast, but the transactions of dopamine are proving to be the Rosetta Stone, the missing link that just might allow the code to be broken.

THE IMPORTANCE OF DOPAMINE was discovered by accident. In 1954 James Olds and Peter Milner, two neuroscientists at McGill University, decided
to implant an electrode deep into the center of a rat's brain. The precise placement of the electrode was largely happenstance: At the time the geography of the mind remained a mystery. But Olds and Milner got lucky. They inserted the needle right next to the nucleus accumbens (NAc), a part of the brain dense with dopamine neurons and involved with the processing of pleasurable rewards, like food and sex. Olds and Milner quickly discovered that too much pleasure can be fatal. After they ran a small current into the wire, so that the NAc was continually excited, the scientists noticed that the rodents lost interest in everything else. They stopped eating and drinking. All courtship behavior ceased. The rats would just curl in the corner of their cage, transfixed by their bliss. Within days all of the animals had perished. They had died of thirst.

It took several decades of painstaking research, but neuroscientists eventually discovered that the rats were suffering from an excess of dopamine. The stimulation of the brain triggered a massive release of the neurotransmitter, which overwhelmed the rodents with ecstasy. In humans addictive drugs work the same way: A crack addict who has just gotten a fix is no different from a rat in electrical trap. This, then, became the dopaminergic cliché—it was the chemical explanation for sex, drugs, and rock 'n' roll.

But that view of the neurotransmitter was vastly oversimplified. What wasn't yet clear was that dopamine is also a profoundly important source of information. It doesn't merely let us take pleasure in the world; it allows us to understand the world.

The first experimental insight into this aspect of the dopamine system came from the pioneering research of Wolfram Schultz, a neuroscientist at Cambridge University. He was originally interested in the neurotransmitter because of its role in triggering Parkinson's disease, which occurs when dopamine neurons begin to die in a part of the brain that controls bodily movement. Schultz recorded from cells in the monkey brain, hoping to find those cells involved in the production of movement. He couldn't find anything. "It was a classic case of experimental failure," he says. But after years of searching in vain, Schultz started to notice something odd about those dopamine neurons: They began to fire just before the monkeys got a reward. (Originally, the reward was a way of getting the monkeys to move.) "At first I thought it was unlikely that an individual cell could represent anything so complicated," Schultz says. "It just seemed like too much information for one neuron."

After hundreds of experimental trials, Schultz began to believe his own data: He realized that he had found, by accident, the reward mechanism at work in the primate brain. "Only in retrospect can I appreciate just how lucky we were," he says. After publishing a series of landmark papers in the mid-1980s, Schultz set out to decipher this reward circuitry in exquisite detail. How, exactly, did these single cells manage to represent a reward? His experiments observed a simple protocol: He played a loud tone, waited for a few seconds, and then squirted a few drops of apple juice into the mouth of a monkey. While the experiment was unfolding, Schultz was probing the dopamine-rich areas of the monkey brain with a needle that monitored the electrical activity inside individual cells. At first the dopamine neurons didn't fire until the juice was delivered; they were responding to the actual reward. However, once the animal learned that the tone preceded the arrival of juice—this requires only a few trials—the same neurons began firing at the sound of the tone instead of the sweet reward. And then eventually, if the tone kept on predicting the juice, the cells went silent. They stopped firing altogether.

When Schultz began publishing his data, nobody quite knew what to make of these strange neurons. "It was very, very tough to figure out what these cells were encoding," Schultz says. He knew that the cells were learning something about the juice and the tone, but he couldn't figure out what they were learning it. The core remained impenetrable.

At the time Montague was a young scientist at the Salk Institute, working in the neurobiology lab of Terry Sejnowski. His approach to the brain was rooted in the abstract theories of computer science, which he hoped would shed light on the software used by the brain. Pete Dayan, a colleague of Montague's at Salk, had introduced him to a model called temporal difference reinforcement learning (TDRL). Computer scientists Rich Sutton and Andrew Barto, who both worked on models of artificial intelligence, had pioneered the model. Sutton and Barto wanted to develop a "neuronlike" program that could learn simple rules and behaviors in order to achieve a goal. The basic premise is straightforward: The software makes predictions about what will happen—about how a checkers game will unfold for example—and then compares these predictions with what actually happens. If the prediction is right, that series of pre-

"WE ARE PROFOUNDLY SOCIAL ANIMALS. YOU CAN'T REALLY UNDERSTAND THE BRAIN UNTIL YOU UNDERSTAND HOW THESE SOCIAL BEHAVIORS HAPPEN, OR WHAT HAPPENS WHEN THEY GO HAYWIRE."
dictions gets reinforced. However, if the prediction is wrong, the software reevaluates its representation of the game.

Montague was entranced by these software prototypes. "It was just so clearly the most efficient way to learn," he says. The problem was that TDRL remained purely theoretical, a system both elegant and imaginary. Even though computer scientists had begun to adapt the programming strategy for various practical purposes, such as running a bank of elevators or determining flight schedules, no one had found a neurological system that worked like this.

But one spring day in 1991, Dayan burst into Montague's small office. "He was very excited and showed these figures from some new paper in my face," Montague remembers. "He kept on saying to me, "What does this look like? What does this look like?"" The figures were from Schultz's experiments with dopamine neurons, and they showed how these cells reacted to the tone and the juice. "I thought he had faked the data," Montague says. "Dayan was a big prankster, and I assumed he'd photoscopied some of our own figures from TDRL just to tease me. It looked too good to be true." Montague immediately realized that he and Dayan could make sense of Schultz's mysterious neurons. They knew these dopamine cells were doing; they had seen this code before.

"The only reason we could see it so clearly," Montague says, "is because we came at it from this theoretical angle. If you were an experimentalist seeing this data, it would have been extremely confusing. What if the bell is those cells doing? Why aren't they just responding to the juice?" That same day Montague and Dayan began writing a technical paper that laid out their insight, explaining how these neurons were making precise predictions about future rewards. But the paper—an awkward mix of Schultz's dopamine recordings and equations borrowed from computer science—went nowhere. "We wrote that paper 11 times," Montague says. "It got bounced from every journal. I came this close to leaving the field. I realized that neuroscience just wasn't ready for theory, even if the theory made sense."

Nevertheless, Montague and Dayan didn't give up. They published their ideas in obscure journals, like Advances in Neural Information Processing Systems. When the big journals rejected their interpretation of monkey neurons, they instead looked at the nervous systems of honeybees, which relied on a version of TDRL when foraging for nectar. That paper got published in Nature in 1995. "We had to drag the experimentalists kicking and screaming," Montague says. "They just didn't understand how these funny-looking equations could explain their data. They told us, 'We need more data.' But what's the point of data if you can't figure it out?"

The crucial feature of these dopamine neurons, say Montague and Dayan, is that they are more concerned with predicting rewards than with the rewards themselves. Once the cells memorize the simple pattern—a loud tone predicts the arrival of juice—they become exquisitely sensitive to variations on the pattern. If the cellular predictions proved correct and the primates experienced a surge of dopamine, the prediction was reinforced. However, if the pattern was violated—if
the tone sounded but the juice never arrived—then the monkey's dopamine neurons abruptly decreased their firing rate. This is known as the "prediction error signal." The monkey got upset because its predictions of juice were wrong.

What's interesting about this system is that it's all about expectation. Dopamine neurons constantly generate patterns based upon experience: if this, then that. The exocophony of reality is distilled into models of correlation. And if these predictions ever prove incorrect, then the neurons immediately readjust their expectations. The discrepancy is internalized; the anomaly is remembered. "The accuracy comes from the mismatch," Montague says. "You learn how the world works by focusing on the prediction errors, on the events that you didn't expect." Our knowledge, in other words, emerges from our cellular mistakes. The brain learns how to be right by focusing on what it got wrong.

Despite his frustrations with the field, Montague continued to work on dopamine. In 1997 he published a Science paper with Dayan and Schultz whose short title was audaciously grand: "A Neural Substrate of Prediction and Reward." The paper has since been cited more than 1,200 times, and it remains the definitive explanation of how the brain parses reality into a set of accurate expectations, which are measured out in short bursts of dopamine. A crucial part of the cellular code had been cracked.

But Montague was getting restless. "I wanted to start asking bigger questions," he says. "Here's this elegant learning system, but how does it fit with the rest of the brain? And can we take this beyond apple juice?"

At first glance the dopamine system might seem largely irrelevant to the study of human behavior. Haven't we evolved beyond the brutish state of "reward harvesting," where all we care about is food and sex? Dopamine might explain the simple psychology of a lizard, or even a monkey sipping juice, but it seems a stretch for it to explain the Promethean mind of a man. "One of the distinguishing traits of human beings is that we chase ideas, not just primary rewards," Montague says. "What other animal goes on hunger strike? Or abstains from sex? Or blows itself up in a cafe in the name of God?" These unique aspects of human cognition seem impossible to explain with neurons that track and predict rewards. After all, these behaviors involve the rejection of rewards. We are shunning off tempting treats because of some abstract belief or goal.

Montague's insight, however, was that ideas are just like apple juice. From the perspective of the brain, an abstraction can be just as rewarding as the tone that predicts the reward. Evolution essentially bootstrapped our penchant for intellectual concepts to the same reward circuits that govern our animal appetites. "The guy who's on hunger strike for some political cause is still relying on his midbrain dopamine neurons, just like a monkey getting a treat," Montague says. "His brain simply values the cause more than it values dinner."

According to Montague, the reason abstract thoughts can be so rewarding, is that the brain relies on a common neural currency for evaluating alternatives. "It's clear that you need
"WE WROTE THAT PAPER 11 TIMES; IT GOT BOUNCED FROM EVERY JOURNAL. I REALIZED THAT NEUROSCIENCE JUST WASN'T READY FOR THEORY, EVEN IF THE THEORY MADE SENSE."

some way to compare your options, even if your options come from very different categories," he says. By representing everything in terms of neurons firing rates, the human brain is able to choose the abstract thought over the visceral reward, as long as the abstraction excites our cells more than apple juice. That’s what makes ideas so powerful: No matter how esoteric or ethereal they get, they are ultimately fed back into the same system that makes us want sex and sugar. As Montague notes, “You don’t have to dig very far before it all comes back to your lemons.”

In recent years Montague has shown how this basic computational mechanism is a fundamental feature of the human mind. Consider a paper on the neural foundations of trust, recently published in Science. The experiment was born out of Montague’s frustration with the limitations of conventional (fMRI) “the most unrealistic element of fMRI experiments is that we could only study the brain by itself,” Montague says. “But when are brains ever by themselves?” And so Montague pioneered a technique known as hyper-scanning, allowing subjects in different fMRI machines to interact in real time. His experiment revolved around a simple economic game in which getting the maximum reward required the strangers to trust one another. However, if one of the players grew especially selfish, he or she could always steal from the pot and erase the tenacious bond of trust. By monitoring the players’ brains, Montague was able to predict whether or not someone would steal money several seconds before the theft actually occurred. The secret was a cortical area known as the caudate nucleus, which closely tracked the payouts from the other player. Montague noticed that whenever the caudate exhibited reduced activity, trust tended to break down.

But what exactly is the caudate computing? How do we decide whom to trust with our money? And why do we sometimes decide to stop trusting those people? It turned out that the caudate worked just like the reward cells in the monkey brain. At first the caudate didn’t get excited until the subjects actually trusted one another and garnered their separate rewards. But over time this brain area started to expect trust, so it fired long before the reward actually arrived. Of course, if the bond was broken—if someone cheated and stole money—then the neurons stopped firing; social assumptions were proven wrong. (Montague is currently repeating this experiment with a collaborating lab in China so that he can detect the influence of culture on social interactions.) The point, he says, is that people were using this TDRL strategy—a strategy that evolved to help animals find edible rewards—to model another mind. Instead of predicting the arrival of juice, the neurons were predicting the behavior of someone else’s brain.

A FEW YEARS AGO, Montague was reviewing some old papers on TDRL theory when he realized that the system, while effective and efficient, was missing something important. Although dopamine neurons excelled at measuring the mismatch between their predictions of rewards and those that actually arrived—these errors provided the input for learning—they’d learn much quicker if they could also incorporate the prediction errors of others. Montague called this a “predictive error learning signal,” since the brain would be benefiting from hypothetical scenarios. “You’re updating your expectations based not just on what happened, but on what might have happened if you’d done something differently.” As Montague saw it, this would be a very valuable addition to our cognitive software.

“I just assumed that evolution would use this approach, because it’s too good an idea not to use,” he says.

The question, of course, is how to find this “what if” signal in the brain. Montague’s clever solution was to use the stock market. After all, Wall Street investors are constantly comparing their actual returns against the returns that might have been, if only they’d sold their shares before the crash or bought Google stock when the company first went public. The experiment went like this: Each subject was given $100 and some basic information about the “current” state of the stock market. After choosing how much money to invest, the players watched nervously as their investments either rose or fell in value. The game continued for 20 rounds, and the subjects got to keep their earnings. One interesting twist was that instead of using random simulations of the stock market, Montague relied on distillations of data from famous historical markets. Montague had people “play” the Dow of 1929, the Nasdaq of 1998, and the S&P 500 of 1987, so the neural responses of investors reflected real-life bubbles and crashes.

The scientists immediately discovered a strong neural signal that drove many of the investment decisions. The signal was predictive learning. Take, for example, this situation. A player has decided to wager 10 percent of her total portfolio in the market, which is a rather small bet. Then she watches as the market rises dramatically in value. At this point, the regret signal in the brain—a swell of activity in the ventral caudate, a reward area rich in dopamine neurons—lights up. While people enjoy their earnings, their brain is fixated on the profits they missed, figuring out the difference between the actual return and the best return “that could have been.” The more we regret a decli
sion, the more likely we are to do something different the next time around. As a result investors in the experiment naturally adapted their investments to the ebb and flow of the market. When markets were booming, as in the Nasdaq bubble of the late 1990s, people perpetually increased their investments.

But fickle learning isn’t always adaptive. Montague argues that these computational signals are also a main cause of financial bubbles. When the market keeps going up, people are naturally inclined to make larger and larger investments in the boom. And then, just when investors are most convinced that the bubble isn’t a bubble—many of Montague’s subjects eventually put all of their money into the booming market—the bubble bursts. The Dow sinks, the Nasdaq collapses. At this point investors race to dump any assets that are declining in value, as their brain realizes that it made some very expensive prediction errors. That’s when you get a financial panic.

Such fickle-error learning signals aren’t relevant only for stock market investors. Look, for instance, at addiction. Dopamine has long been associated with addictive drugs, such as cocaine, that overexcite these brain cells. The end result is that addicts make increasingly reckless decisions, forgetting long-term goals for the sake of an intensely pleasurable short-term fix. “When you’re addicted to a drug, your brain is basically convinced that this expensive white powder is worth more than your marriage or life,” Montague says. In other words addiction is a disease of valuation: Dopamine cells have lost track of what’s really important.

Montague wanted to know which part of the dopamine system was distorted in the addicted brain. He began to wonder if addiction was, at least in part, a disease of fickle learning. Addicted smokers will continue to smoke even when they know it’s bad for them. Why can’t they instead revise their models of reward?

Last year Montague decided to replicate his stock market study with a large group of chronic smokers. It turned out that smokers were perfectly able to compute a “what if” learning signal, which allowed them to experience regret. Like nonsmokers they realized that they should have invested differently in the stock market. Unfortunately, this signal had no impact on their decision making, which led them to make significantly less money during the investing game. According to Montague, this data helps explain why smokers continue to smoke even when they regret it. Although their dopamine neurons correctly compute the rewards of an extended life versus a hit of nicotine—they are, in essence, asking themselves, “What if I don’t smoke this cigarette?”—their brain doesn’t process the result. That feeling of regret is conveniently ignored. They just keep on lighting up.

Montague explored the confidence of a scientist used to confirming his hypotheses. He buzzes with ideas for new experiments—“I get bored rather easily,” he says—and his lab is constantly shifting direction, transitioning from dopamine to neuroeconomics to social neuroscience. Montague is currently consumed with questions about how people interact when they’re part of a group. “A mob or a market is not just a collection of discrete individuals,” he says. “It’s something else entirely. You would do things in a group that you would never do by yourself. But what’s happening in your brain? We’ve got all these sociological studies but no hard data.” Montagge’s been warned that the project is too complicated, that social interactions are too subtle and complex to pick up in a scanner, but he’s convinced otherwise. “If I’d listened to the naysayers,” he says, “I’d still be studying honeybees.”

Montague’s experiments take advantage of his unique fMRI setup. He has four people negotiate with one another as they decide how much to offer someone else during an investing game. While the group is bickering, Montague is monitoring the brain activity of everyone involved. He’s also instructed the group with a computer player that has been programmed to act just like a person with borderline personality disorder. The purpose of this particular experiment is to see how “one bad apple” can lead perfect strangers to also act badly.

While Montague isn’t ready to share the results—“he’s still gathering data”—what he’s found so far is, he says, “stunning, shocking even… For me the lesson has been that people act very badly in groups. And now we can see why.”

Such exuberance is well earned. In the space of a few short years, Montague has taken his theoretical model of learning—a model he borrowed from some old computer science textbooks—and shown that it’s an essential part of the human brain. He’s linked the transactions of a single neurotransmitter to a dizzying array of behaviors, so that it’s now possible to draw a straight line between monkeys craving juice and stock market bubbles. A neurotransmitter that wasn’t supposed to matter is now out most important clue into the secret messages of the mind and the breakdown of social graces. The code hasn’t been broken. But for the first time, it’s getting cracked.