3. Behind the Curtain

Music and the Mind Machine

For cognitive scientists, the word *mind* refers to that part of each of us that embodies our thoughts, hopes, desires, memories, beliefs, and experiences. The brain, on the other hand, is an organ of the body, a collection of cells and water, chemicals and blood vessels, that resides in the skull. Activity in the brain gives rise to the contents of the mind. Cognitive scientists sometimes make the analogy that the brain is like a computer's CPU, or hardware, while the mind is like the programs or software running on the CPU. (If only that were literally true and we could just run out to buy a memory upgrade.) Different programs can run on what is essentially the same hardware—different minds can arise from very similar brains.

Western culture has inherited a tradition of dualism from René Descartes, who wrote that the mind and the brain are two entirely separate things. Dualists assert that the mind preexisted, before you were born, and that the brain is not the seat of thought—rather, it is merely an instrument of the mind, helping to implement the mind's will, move muscles, and maintain homeostasis in the body. To most of us, it certainly feels as though our minds are something unique and distinctive, separate from just a bunch of neurochemical processes. We have a feeling of what it is like to be me, what it is like to be me reading a book, and what it is like to think about what it is like to be me. How can *me* be reduced so unceremoniously to axons, dendrites, and ion channels? It feels like we are something more.

But this feeling could be an illusion, just as it certainly feels as though the earth is standing still, not spinning around on its axis at a thousand miles per hour. Most scientists and contemporary philosophers believe that the brain and mind are two parts of the same thing, and some believe that the distinction itself is flawed. The dominant view today is that that the sum total of your thoughts, beliefs, and experiences is represented in patterns of firings—electrochemical activity—in the brain. If the brain ceases to function, the mind is gone, but the brain can still exist, thoughtless, in a jar in someone's laboratory.

Evidence for this comes from neuropsychological findings of regional specificity of function. Sometimes, as a result of stroke (a blockage of blood vessels in the brain that leads to cell death), tumors, head injury, or other trauma, an area of the brain becomes damaged. In many of these cases, damage to a specific brain region leads to a loss of a particular mental or bodily function. When dozens or hundreds of cases show loss of a specific function associated with a particular brain region, we infer that this brain region is somehow involved in, or perhaps responsible for, that function.

More than a century of such neuropsychological investigation has allowed us to make maps of the brain's areas of function, and to localize particular cognitive operations. The prevailing view of the brain is that it is a computational system, and we think of the brain as a type of computer. Networks of interconnected neurons perform computations on information and combine their computations in ways that lead to thoughts, decisions, perceptions, and ultimately consciousness. Different subsystems are responsible for different aspects of cognition. Damage to an area of the brain just above and behind the left ear—Wernicke's area—causes difficulty in understanding spoken language; damage to a region at the very top of the head—the motor cortex—causes difficulty moving your fingers; damage to an area in the center of the brain—the hippocampal complex—can block the ability to form new memories, while leaving old memories intact. Damage to an area just behind your forehead can cause dramatic changes in personality—it can rob aspects of you from you. Such localization of mental function is a strong scientific argument for the involvement of the brain in thought, and the thesis that thoughts come from the brain.

We have known since 1848 (and the medical case of Phineas Gage) that the frontal lobes are intimately related to aspects of self and personality. Yet even one hundred and fifty years later, most of what we can say about personality and neural structures is vague and quite general. We have not located the "patience" region of the brain, nor the "jealousy" or "generous" regions, and it seems unlikely that we ever will. The brain has regional differentiation of structure and function, but complex personality attributes are no doubt distributed widely throughout the brain.

The human brain is divided up into four lobes—the frontal, temporal, parietal, and occipital—plus the cerebellum. We can make some gross generalizations about function, but in fact behavior is complex and not readily reducible to simple mappings. The frontal lobe is associated with planning, and with self-control, and with making sense out of the dense and jumbled signals that our senses receive—the so-called "perceptual organization" that the Gestalt psychologists studied. The temporal lobe is associated with hearing and memory. The parietal lobe is associated with motor movements and spatial skill, and the occipital lobe with vision. The cerebellum is involved in emotions and the planning of movements, and is the evolutionarily oldest part of our brain; even many animals, such as reptiles, that lack the "higher" brain region of the cortex still have a cerebellum. The surgical separation of a portion of the frontal lobe, the prefrontal cortex, from the thalamus is called a lobotomy. So when the Ramones sang "Now I guess I'll have to tell 'em/That I got no cerebellum" in their song "Teenage Lobotomy" (words and music by Douglas Colvin, John Cummings, Thomas Erdely, and Jeffrey Hyman) they were not being anatomically accurate, but for the sake of artistic license, and for creating one of the great rhymes in rock music, it is hard to begrudge them that.

Musical activity involves nearly every region of the brain that we know about, and nearly every neural subsystem. Different aspects of the music are handled by different neural regions—the brain uses functional segregation for music processing, and employs a system of feature detectors whose job it is to analyze specific aspects of the musical signal, such as pitch, tempo, timbre, and so on. Some of the music processing has points in common with the operations required to analyze other sounds; understanding speech, for example, requires that we segment a flurry of sounds into words, sentences, and phrases, and that we be able to understand aspects beyond the words, such as sarcasm (isn't *that* interesting). Several different dimensions of a musical sound need to be analyzed—usually involving several quasi-independent neural processes—and they then need to be brought together to form a coherent representation of what we're listening to.

Listening to music starts with subcortical (below-the-cortex) structures—the cochlear nuclei, the brain stem, the cerebellum—and then moves up to auditory cortices on both sides of the brain. Trying to follow along with music that you know—or at least music in a style you're familiar with, such as baroque or blues—recruits additional regions of the brain, including the hippocampus—our memory center—and subsections of the frontal lobe, particularly a region called inferior frontal cortex, which is in the lowest parts of the frontal lobe, i.e., closer to your chin than to the top of your head. Tapping along with music, either actually or just in your mind, involves the cerebellum's timing circuits. Performing music—regardless of what instrument you play, or whether you sing, or conduct—involves the frontal lobes again for the planning of your behavior, as well as the motor cortex in the parietal lobe just underneath the top of your head, and the sensory cortex, which provides the tactile feedback that you have pressed the right key on your instrument, or moved the baton where you thought you did. Reading music involves the visual cortex, in the back of your head in the occipetal lobe. Listening to or recalling lyrics invokes language centers, including Broca's and Wernicke's area, as well as other language centers in the temporal and frontal lobes.

At a deeper level, the emotions we experience in response to music involve structures deep in the primitive, reptilian regions of the cerebellar vermis, and the amygdala—the heart of emotional processing in the cortex. The idea of regional specificity is evident in this summary but a complementary principle applies as well, that of distribution of function. The brain is a massively parallel device, with operations distributed widely throughout. There is no single language center, nor is there a single music center. Rather, there are regions that peform component operations, and other regions that coordinate the bringing together of this information. Finally, we have discovered only recently that the brain has a capacity for reorganization that vastly exceeds what we thought before. This ability is called neuroplasticity, and in some cases, it suggests that regional specificity may be temporary, as the processing centers for important mental functions actually move to other regions after trauma or brain damage.

It is difficult to appreciate the complexity of the brain because the numbers are so huge they go well beyond our everyday experience (unless you are a cosmologist). The average brain consists of one hundred billion (100,000,000,000) neurons. Suppose each neuron was one dollar, and you stood on a street corner trying to give dollars away to people as they passed by, as fast as you could hand them out—let's say one dollar per second. If you did this twenty-four hours a day, 365 days a year, without stopping, and if you had started on the day that Jesus was born, you would by the present day only have gone through about two thirds of your money. Even if you gave away *hundred*-dollar bills once a second, it would take you thirty-two years to pass them all out. This is a lot of neurons, but the real power and complexity of the brain (and of thought) come through their connections.

Each neuron is connected to other neurons—usually one thousand to ten thousand others. Just four neurons can be connected in sixty-three ways, or not at all, for a total of sixty-four possibilities. As the number of neurons increases, the number of possible connections grows exponentially (the formula for the way that n neurons can be connected to each other is $2^{(n*(n-1)/2)}$):

For 2 neurons there are 2 possibilities for how they can be connected
For 3 neurons there are 8 possibilities
For 4 neurons there are 64 possibilities
For 5 neurons there are 1,024 possibilities
For 6 neurons there are 32,768 possibilities

The number of combinations becomes so large that it is unlikely that we will ever understand all the possible connections in the brain, or what they mean. The number of combinations possible—and hence the number of possible different thoughts or brain states each of us can have—exceeds the number of known particles in the entire known universe.

Similarly, you can see how it is that all the songs we have ever heard—and all those that will ever be created—could be made up of just twelve musical notes (ignoring octaves). Each note can go to another note, or to itself, or to a rest, and this yields twelve possibilities. But each of those possibilities yields twelve more. When you factor in rhythm—each note can take on one of many different note lengths—the number of possibilities grows very, very rapidly.

Much of the brain's computational power comes from this enormous possibility for interconnection, and much of it comes from the fact that brains are parallel processing machines, rather than serial processors. A serial processor is like an assembly line, handling each piece of information as it comes down the mental conveyor belt, performing some operation on that piece of information, and then sending it down the line for the next operation. Computers work like this. Ask a computer to download a song from a Web site, tell you the weather in Boise, and save a file you've been working on, and it will do them one at a time; it does things so fast that it can seem as though it is doing them at the same time—in parallel—but it isn't. Brains, on the other hand, can work on many things at once, overlapping and in parallel. Our auditory system processes sound in this way—it doesn't have to wait to find out what the pitch of a sound is to know where it is coming from; the neural circuits devoted to these two operations are trying to come up with answers at the same time. If one neural circuit finishes its work before another, it just sends its information to other connected brain regions and they can begin using it. If late-arriving information that affects an interpretation of what we're hearing comes in from a separate processing circuit, the brain can "change its mind" and update what it thinks is out there. Our brains are updating their opinions all the time—particularly when it comes to perceiving visual and auditory stimuli—hundreds of times per second, and we don't even know it.

Here's an analogy to convey how neurons connect to each other. Imagine that you're sitting home alone one Sunday morning. You don't feel much of one way or another—you're not particularly happy, not particularly sad, neither angry, excited, jealous, nor tense. You feel more or less neutral. You have a bunch of friends, a network of them, and you can call any of them on the phone. Let's say that each of your friends is rather one dimensional and that they can exert a great influence on your mood. You know, for example, that if you telephone your friend Hannah she'll put you in a happy mood. Whenever you talk to Sam it makes you sad, because the two of you had a third friend who died and Sam reminds you of that. Talking to Carla makes you calm and serene, because she has a soothing voice and you're reminded of the times you sat in a beautiful forest clearing with her, soaking up the sun and meditating. Talking to Edward makes you feel energized; talking to Tammy makes you feel tense. You can pick up your telephone and connect to any of these friends and induce a certain emotion.

You might have hundreds or thousands of these one-dimensional friends, each capable of evoking a particular memory, experience, or mood state. These are your connections. Accessing them causes you to change your mood, or state. If you were to talk to Hannah and Sam at the same time, or one right after the other, Hannah would make you feel happy, Sam would make you feel sad, and in the end you'd be back to where you were—neutral. But we can add an additional nuance, which is the weight or force-of-influence of these connections—how close you feel to an individual at a particular point in time. That weight determines the amount of influence the person will have on you. If you feel twice as close to Hannah as you do to Sam, talking to Hannah and Sam for an equal amount of time would still leave you feeling happy, although not as happy as if you had talked to Hannah alone—Sam's sadness brings you down, but only halfway from the happiness you gained from talking to Hannah.

Let's say that all of these people can talk to one another, and in so doing, their states can be modified to some extent. Although your friend Hannah is dispositionally cheery, her cheerfulness can be attenuated by a conversation she has with Sad Sam. If you phone Edward the energizer after he's just spoken with Tense Tammy (who has just gotten off the phone with Jealous Justine), Edward may make you feel a new mix of emotions you've never experienced before, a kind of tense jealousy that you have a lot of energy to go out and do something about. And any of these friends might telephone you at any time, evoking these states in you as a complex chain of feelings or experiences that has gone around, each one influencing the other, and you, in turn, will leave your emotional mark on them. With thousands of friends interconnected like this, and a bunch of telephones in your living room ringing off the hook all day long, the number of emotional states you might experience would indeed be quite varied.

It is generally accepted that our thoughts and memories arise from the myriad connections of this sort that our neurons make. Not all neurons are equally active at one time, however—this would cause a cacophony of images and sensations in our heads (in fact, this is what happens in epilepsy). Certain groups of neurons—we can call them networks—become active during certain cognitive activities, and they in turn can activate other neurons. When I stub my toe, the sensory receptors in my toe send signals up to the sensory cortex in my brain. This sets off a chain of neural activations that causes me to experience pain, withdraw my foot from the object I stubbed it against, and that might cause my mouth to open involuntarily and shout "&%@!"

When I hear a car horn, air molecules impinging on my eardrum cause electrical signals to be sent to my auditory cortex. This causes a cascade of events that recruits a very different group of neurons than toe stubbing. First, neurons in the auditory cortex process the pitch of the sound so that I can distinguish the car horn from something with a different pitch like a truck's air horn, or the air-horn-in-a-can at a football game. A different group of neurons is activated to determine the location from which the sound came. These and other processes invoke a visual orienting response—I turn toward the sound to see what made it, and instantaneously, if necessary, I jump back (the result of activity from the neurons in my motor cortex, orchestrated with neurons in my emotional center, the amygdala, telling me that danger is imminent).

When I hear Rachmaninoff's Piano Concerto no. 3, the hair cells in my cochlea parse the incoming sound into different frequency bands, sending electrical signals to my primary auditory cortex—area A1—telling it what frequencies are present in the signal. Additional regions in the temporal lobe, including the superior temporal sulcus and the superior temporal gyrus on both sides of the brain, help to distinguish the different timbres I'm hearing. If I want to label those timbres, the hippocampus helps to retrieve the memory of similar sounds I've heard before, and then I'll need to access my mental dictionary—which will require using structures found at the junction between the temporal, occipetal, and parietal lobes. So far, these regions are the same ones, although activated in different ways and with different populations of neurons, that I would use to process the car horn. Whole new populations of neurons will become active, however, as I attend to pitch sequences (dorsalateral prefrontal cortex, and Brodmann areas 44 and 47), rhythms (the lateral cerebellum and the cerebellar vermis), and emotion (frontal lobes, cerebellum, the amygdala, and the nucleus accumbens—part of a network of structures involved in feelings of pleasure and reward, whether it is through eating, having sex, or listening to pleasurable music).

To some extent, if the room is vibrating with the deep sounds of the double bass, some of those same neurons that fired when I stubbed my toe may fire now—neurons sensitive to tactile input. If the car horn has a pitch of A440, neurons that are set to fire when that frequency is encountered will most probably fire, and they'll fire again when an A440 occurs in Rachmaninoff. But my inner mental experience is likely to be different because of the different contexts involved and the different neural networks that are recruited in the two cases.

My experience with oboes and violins is different, and the particular way that Rachmaninoff uses them may cause me to have the opposite reaction to his concerto than I have to the car horn; rather than feeling startled, I feel relaxed. The same neurons that fire when I feel calm and safe in my environment may be triggered by the calm parts of the concerto.

Through experience, I've learned to associate car horns with danger, or at least with someone trying to get my attention. How did this happen? Some sounds are intrinsically soothing while others are frightening. Although there is a great deal of interpersonal variation, we are born with a predisposition toward interpreting sounds in particular ways. Abrupt, short, loud sounds tend to be interpreted by many animals as an alert sound; we see this when comparing the alert calls of birds, rodents, and apes. Slow onset, long, and quieter sounds tend to be interpreted as calming, or at least neutral. Think of the sharp sound of a dog's bark, versus the soft purring of a cat who sits peacefully on your lap. Composers know this, of course, and use hundreds of subtle shadings of timbre and note length to convey the many different emotional shadings of human experience.

In the "Surprise Symphony" by Haydn (Symphony no. 94 in G Major, second movement, andante), the composer builds suspense by using soft violins in the main theme. The softness of the sound is soothing, but the shortness of the pizzicato accompaniment sends a gentle, contradictory message of danger, and together they give a soft sense of suspense. The main melodic idea spans barely more than half an octave, a perfect fifth. The melodic contour further suggests complacency—the melody first goes up, then down, then repeats the “up” motif. The parallelism implied by the melody, the up/down/up, gets the listener ready for another “down” part. Continuing with the soft, gentle violin notes, the maestro changes the melody by going up—just a little—but holds the rhythms constant. He rests on the fifth, a relatively stable tone harmonically. Because the fifth is the highest note we’ve encountered so far, we expect that when the next note comes in, it will be lower—that it will begin the return home toward the root (or tonic), and “close the gap” created by the distance between the tonic and the current note—the fifth. Then, from out of nowhere, Haydn sends us a loud note an octave higher, with the brash horns and timpani carrying the sound. He has violated our expectations for melodic direction, for contour, for timbre, and for loudness all at once. This is the “Surprise” in the “Surprise Symphony.”

This Haydn symphony violates our expectations of how the world works. Even someone with no musical knowledge or musical expectations whatsoever finds the symphony surprising because of this timbral effect, switching from the soft purring of the violins to the alert call of horns and drums. For someone with a musical background, the symphony violates expectations that have been formed based on musical convention and style. Where do surprises, expectations, and analyses of this sort occur in the brain? Just how these operations are carried out in neurons is still something of a mystery, but we do have some clues.

Before going any farther, I have to admit a bias in the way I approach the scientific study of minds and brains: I have a definite preference for studying the mind rather than the brain. Part of my preference is personal rather than professional. As a child I wouldn’t collect butterflies with the rest of my science class because life—all life—seems sacred to me. And the stark fact about brain research over the course of the last century is that it generally involves poking around in the brains of live animals, often our close genetic cousins, the monkeys and apes, and then killing (they call it “sacrificing”) the animal. I worked for one miserable semester in a monkey lab, dissecting the brains of dead monkeys to prepare them for microscopic examination. Every day I had to walk by cages of the ones that were still alive. I had nightmares.

At a different level, I've always been more fascinated by the thoughts themselves, not the neurons that give rise to them. A theory in cognitive science named functionalism—which many prominent researchers subscribe to—asserts that similar minds can arise from quite different brains, that brains are just the collection of wires and processing modules that instantiate thought. Regardless of whether the functionalist doctrine is true, it does suggest that there are limits to how much we can know about thought from just studying brains. A neurosurgeon once told Daniel Dennett (a prominent and persuasive spokesperson for functionalism) that he had operated on hundreds of people and seen hundreds of live, thinking brains, but he had never seen a thought.

When I was trying to decide where to attend graduate school, and who I wanted to have as a mentor, I was infatuated with the work of Professor Michael Posner. He had pioneered a number of ways of looking at thought processes, among them mental chronometry (the idea that much can be learned about the organization of the mind by measuring how long it takes to think certain thoughts), ways to investigate the structure of categories, and the famous Posner Cueing Paradigm, a novel method for studying attention. But rumor had it that Posner was abandoning the mind and had started studying the brain, something I was certain I did not want to do.

Although still an undergraduate (albeit a somewhat older one than usual), I attended the annual meeting of the American Psychological Association, which was held in San Francisco that year, just forty miles up the road from Stanford, where I was finishing up my B.A. I saw Posner's name on the program and attended his talk, which was full of slides containing pictures of people's brains while they were doing one thing or another. After his talk was over he took some questions, then disappeared out a back door. I ran around to the back and saw him way ahead, rushing across the conference center to get to another talk. I ran to catch up to him. I must have been quite a sight to him! I was out of breath from running. Even without the panting, I was nervous meeting one of the great legends of cognitive psychology. I had read his textbook in my first psychology class at MIT (where I began my undergraduate training before transferring to Stanford); my first psychology professor, Susan Carey, spoke of him with what could only be described as reverence in her voice. I can still remember the echoes of her words, reverberating through the lecture hall at MIT: “Michael Posner, one of the smartest and most creative people I’ve ever met.”

I started to sweat, I opened my mouth, and . . . nothing. I started “Mmm . . .” All this time we were walking rapidly side by side—he’s a fast walker—and every two or three steps I’d fall behind again. I stammered an introduction and said that I had applied to the University of Oregon to work with him. I’d never stuttered before, but I had never been this nervous before. “P-p-p-professor P-p-posner, I hear that you’ve shifted your research focus entirely to the b-b-brain—is that true? Because I really want to study cognitive psychology with you,” I finally told him.

“Well, I am a little interested in the brain these days,” he said. “But I see cognitive neuroscience as a way to provide constraints for our theories in cognitive psychology. It helps us to distinguish whether a model has a plausible basis in the underlying anatomy.”

Many people enter neuroscience from a background in biology or chemistry and their principal focus is on the mechanisms by which cells communicate with each other. To the cognitive neuroscientist, understanding the anatomy or physiology of the brain may be a challenging intellectual exercise (the brain scientists’ equivalent of a really complicated crossword puzzle), but it is not the ultimate goal of the work. Our goal is to understand thought processes, memories, emotions, and experiences, and the brain just happens to be the box that all this happens in. To return to the telephone analogy and conversations you might have with different friends who influence your emotions: If I want to predict how you’re going to feel tomorrow, it will be of only limited value for me to map the layout of the telephone lines connecting all the different people you know. More important is to understand their individual proclivities: Who is likely to call you tomorrow and what are they likely to say? How are they apt to make you feel? Of course, to entirely ignore the connectivity question would be a mistake too. If a line is broken, or if there is no evidence of a connection between person A and person B, or if person C can never call you directly but can only influence you through person A who can call you directly—all this information provides important constraints to a prediction.

This perspective influences the way I study the cognitive neuroscience of music. I am not interested in going on a fishing expedition to try every possible musical stimulus and find out where it occurs in the brain; Posner and I have talked many times about the current mad rush to map the brain as just so much atheoretical cartography. The point for me isn't to develop a map of the brain, but to understand how it works, how the different regions coordinate their activity together, how the simple firing of neurons and shuttling around of neurotransmitters leads to thoughts, laughter, feelings of profound joy and sadness, and how all these, in turn, can lead us to create lasting, meaningful works of art. These are the functions of the mind, and knowing where they occur doesn't interest me unless the where can tell us something about how and why. An assumption of cognitive neuroscience is that it can.

My perspective is that, of the infinite number of experiments that are possible to do, the ones worth doing are those that can lead us to a better understanding of how and why. A good experiment is theoretically motivated, and makes clear predictions as to which one of two or more competing hypotheses will be supported. An experiment that is likely to provide support for both sides of a contentious issue is not one worth doing; science can only move forward by the elimination of false or untenable hypotheses.

Another quality of a good experiment is that it is generalizable to other conditions—to people not studied, to types of music not studied, and to a variety of situations. A great deal of behavioral research is conducted on only a small number of people ("subjects" in the experiment), and with very artificial stimuli. In my laboratory we use both musicians and nonmusicians whenever possible, in order to learn about the broadest cross section of people. And we almost always use real-world music, actual recordings of real musicians playing real songs, so that we can better understand the brain's responses to the kind of music that most people listen to, rather than the kind of music that is found only in the neuroscientific laboratory. So far this approach has panned out. It is more difficult to provide rigorous experimental controls with this approach, but it is not impossible; it takes a bit more planning and careful preparation, but in the long run, the results are worth it. In using this naturalistic approach, I can state with reasonable scientific certainty that we're studying the brain doing what it normally does, rather than what it does when assaulted by rhythms without any pitch, or melodies without any rhythms. In an attempt to separate music into its components, we run the risk—if the experiments are not done properly—of creating sound sequences that are very unmusical.

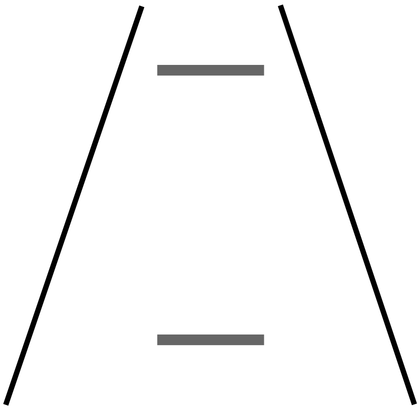
When I say that I am less interested in the brain than in the mind, this does not mean that I have no interest in the brain. I believe that we all have brains, and I believe brains are important! But I also believe similar thoughts can arise from different brain architectures. By analogy, I can watch the same television program on an RCA, a Zenith, a Mitsubishi, even on my computer screen with the right hardware and software. The architectures of all these are sufficiently distinct from one another that the patent office—an organization charged with the responsibility of deciding when something is sufficiently different from something else that it constitutes an invention—has issued different patents to these various companies, establishing that the underlying architectures are significantly different. My dog Shadow has a very different brain organization, anatomy, and neurochemistry from mine. When he is hungry or hurts his paw, it is unlikely that the pattern of nerve firings in his brain bears much resemblance to the pattern of firings in my brain when I'm hungry or stub my toe. But I do believe that he is experiencing substantially similar mind states.

Some common illusions and misconceptions need to be set aside. Many people, even trained scientists in other disciplines, have the strong intuition that inside the brain there is a strictly isomorphic representation of the world around us. (*Isomorphic* comes from the Greek word *iso*, meaning "same," and *morphus*, meaning "form.") The Gestalt psychologists, who were right about a great many things, were among the first to articulate this idea. If you look at a square, they argued, a square-shaped pattern of neurons is activated in your brain. Many of us have the intuition that if we're looking at a tree, the image of the tree is somewhere represented in the brain as a tree, and that perhaps seeing the tree activates a set of neurons in the shape of a tree, with roots at one end and leaves at the other. When we listen to or imagine a favorite song, it feels like the song is playing in our head, over a set of neural loudspeakers.

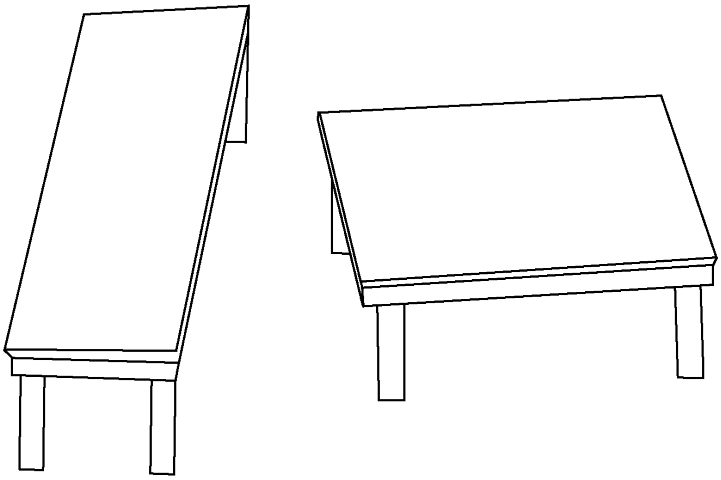
Daniel Dennett and V. S. Ramachandran have eloquently argued that there is a problem with this intuition. If a mental picture of something (either as we see it right now or imagine it in memory) is itself a picture, there has to be some part of our mind/brain that is seeing that picture. Dennett talks about the intuition that visual scenes are presented on some sort of a screen or theater in our minds. For this to be true, there would have to be someone in the audience of that theater watching the screen, and holding a mental image inside his head. And who would that be? What would that mental image look like? This quickly leads to an infinite regress. The same argument applies to auditory events. No one argues that it doesn't feel like we have an audio system in our minds. Because we can manipulate mental images—we can zoom in on them, rotate them, in the case of music we can speed up or slow down the song in our heads—we're compelled to think there is a home theater in the mind. But logically this cannot be true because of the infinite regress problem.

We are also under the illusion that we simply open our eyes and—we see. A bird chirps outside the window and we instantly hear. Sensory perception creates mental images in our minds—representations of the world outside our heads—so quickly and seamlessly that it seems there is nothing to it. This is an illusion. Our perceptions are the end product of a long chain of neural events that give us the illusion of an instantaneous image. There are many domains in which our strongest intuitions mislead us. The flat earth is one example. The intuition that our senses give us an undistorted view of the world is another.

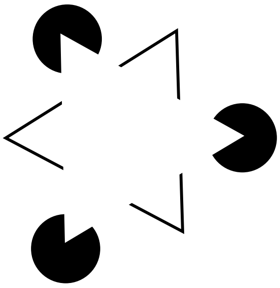
It has been known at least since the time of Aristotle that our senses can distort the way we perceive the world. My teacher Roger Shepard, a perception psychologist at Stanford University, used to say that when functioning properly, our perceptual system is supposed to distort the world we see and hear. We interact with the world around us through our senses. As John Locke noted, everything we know about the world is through what we see, hear, smell, touch, or taste. We naturally assume that the world is just as we perceive it to be. But experiments have forced us to confront the reality that this is not the case. Visual illusions are perhaps the most compelling proof of sensory distortion. Many of us have seen these sorts of illusions as children, such as when two lines of the same length appear to be different lengths (the Ponzo illusion).



Roger Shepard drew an illusion he calls “Turning the Tables” that is related to the Ponzo. It’s hard to believe, but these tabletops are identical in size and shape (you can check by cutting out a piece of paper or cellophane the exact shape of one and then placing it over the other). This illusion exploits a principle of our visual system’s depth perception mechanisms. Even knowing that it is an illusion does not allow us to turn off the mechanism. No matter how many times we view this figure, it continues to surprise us because our brains are actually giving us misinformation about the objects.



In the Kaniza illusion there appears to be a white triangle lying on top of a black-outlined one. But if you look closely, you'll see that there are no triangles in the figure. Our perceptual system completes or "fills in" information that isn't there.



Why does it do this? Our best guess is that it was evolutionarily adaptive to do so. Much of what we see and hear contains missing information. Our hunter-gatherer ancestors might have seen a tiger partially hidden by trees, or heard a lion's roar partly obscured by the sound of leaves rustling much closer to us. Sounds and sights often come to us as partial information that has been obscured by other things in the environment. A perceptual system that can restore missing information would help us make quick decisions in threatening situations. Better to run now than sit and try to figure out if those two separate, broken pieces of sound were part of a single lion roar.

The auditory system has its own version of perceptual completion. The cognitive psychologist Richard Warren demonstrated this particularly well. He recorded a sentence, "The bill was passed by both houses of the legislature," and cut out a piece of the sentence from the recording tape. He replaced the missing piece with a burst of white noise (static) of the same duration. Nearly everyone who heard the altered recording could report that they heard both a sentence and static. But a large proportion of people couldn't tell where the static was! The auditory system had filled in the missing speech information, so that the sentence seemed to be uninterrupted. Most people reported that there was static and that it existed apart from the spoken sentence. The static and the sentence formed separate perceptual streams due to differences in timbre that caused them to group separately; Bregman calls this streaming by timbre. Clearly this is a sensory distortion; our perceptual system is telling us something about the world that isn't true. But just as clearly, this has an evolutionary/adaptive value if it can help us make sense of the world during a life-or-death situation.

According to the great perception psychologists Hermann von Helmholtz, Richard Gregory, Irvin Rock, and Roger Shepard, perception is a process of inference, and involves an analysis of probabilities. The brain's task is to determine what the most likely arrangement of objects in the physical world is, given the particular pattern of information that reaches the sensory receptors—the retina for vision, the eardrum for hearing. Most of the time the information we receive at our sensory receptors is incomplete or ambiguous. Voices are mixed in with other voices, the sounds of machines, wind, footsteps. Wherever you are right now—whether you're in an airplane, a coffee shop, a library, at home, in a park, or anywhere else—stop and listen to the sounds around you. Unless you're in a sensory isolation tank, you can probably identify at least a half-dozen different sounds. Your brain's ability to make these identifications is nothing short of remarkable when you consider what it starts out with—that is, what the sensory receptors pass up to it. Grouping principles—by timbre, spatial location, loudness, and so on—help to segregate them, but there is still a lot we don't know about this process; no one has yet designed a computer that can perform this task of sound source separation.

The eardrum is simply a membrane that is stretched across tissue and bone. It is the gateway to hearing. Virtually all of your impressions of the auditory world come from the way in which it wiggles back and forth in response to air molecules hitting it. (To a degree, the pinnae—the fleshy parts of your ear—are also involved in auditory perception, as are the bones in your skull, but for the most part, the eardrum is the primary source of what we know about what is out there in the auditory world.) Let's consider a typical auditory scene, a person sitting in her living room reading a book. In this environment, let's suppose that there are six sources of sound that she can readily identify: the whooshing noise of the central heating (the fan or blower that moves air through the ductwork), the hum of a refrigerator in the kitchen, traffic outside on the street (which itself could be several or dozens of distinct sounds comprising different engines, brakes squeaking, horns, etc.), leaves rustling in the wind outside, a cat purring on the chair next to her, and a recording of Debussy preludes. Each of these can be considered an auditory object or a sound source, and we are able to identify them because each has its own distinctive sound.

Sound is transmitted through the air by molecules vibrating at certain frequencies. These molecules bombard the eardrum, causing it to wiggle in and out depending on how hard they hit it (related to the volume or amplitude of the sound) and on how fast they're vibrating (related to what we call pitch). But there is nothing in the molecules that tells the eardrum where they came from, or which ones are associated with which object. The molecules that were set in motion by the cat purring don't carry an identifying tag that says cat, and they may arrive on the eardrum at the same time and in the same region of the eardrum as the sounds from the refrigerator, the heater, Debussy, and everything else.

Imagine that you stretch a pillowcase tightly across the opening of a bucket, and different people throw Ping-Pong balls at it from different distances. Each person can throw as many Ping-Pong balls as he likes, and as often as he likes. Your job is to figure out, just by looking at how the pillowcase moves up and down, how many people there are, who they are, and whether they are walking toward you, away from you, or are standing still. This is analogous to what the auditory system has to contend with in making identifications of auditory objects in the world, using only the movement of the eardrum as a guide. How does the brain figure out, from this disorganized mixture of molecules beating against a membrane, what is out there in the world? In particular, how does it do this with music?

It does this through a process of feature extraction, followed by another process of feature integration. The brain extracts basic, low-level features from the music, using specialized neural networks that decompose the signal into information about pitch, timbre, spatial location, loudness, reverberant environment, tone durations, and the onset times for different notes (and for different components of tones). These operations are carried out in parallel by neural circuits that compute these values and that can operate somewhat independently of one another—that is, the pitch circuit doesn't need to wait for the duration circuit to be done in order to perform its calculations. This sort of processing—where only the information contained in the stimulus is considered by the neural circuits—is called bottom-up processing. In the world and in the brain, these attributes of the music are separable. We can change one without changing the other, just as we can change shape in visual objects without changing their color.

Low-level, bottom-up processing of basic elements occurs in the peripheral and phylogenetically older parts of our brains; the term *low-level* refers to the perception of elemental or building-block attributes of a sensory stimulus. High-level processing occurs in more sophisticated parts of our brains that take neural projections from the sensory receptors and from a number of low-level processing units; this refers to the combining of low-level elements into an integrated representation. High-level processing is where it all comes together, where our minds come to an understanding of form and content. Low-level processing in your brain sees blobs of ink on this page, and perhaps even allows you to put those blobs together and recognize a basic form in your visual vocabulary, such as the letter *A*. But it is high-level processing that puts together three letters to let you read the word ART and to generate a mental image of what the word means.

At the same time as feature extraction is taking place in the cochlea, auditory cortex, brain stem, and cerebellum, the higher-level centers of our brain are receiving a constant flow of information about what has been extracted so far; this information is continually updated, and typically rewrites the older information. As our centers for higher thought—mostly in the frontal cortex—receive these updates, they are working hard to predict what will come next in the music, based on several factors:

- what has already come before in the piece of music we're hearing;
- what we remember will come next if the music is familiar;
- what we expect will come next if the genre or style is familiar, based on previous exposure to this style of music;
- any additional information we've been given, such as a summary of the music that we've read, a sudden movement by a performer, or a nudge by the person sitting next to us.

These frontal-lobe calculations are called top-down processing and they can exert influence on the lower-level modules while they are performing their bottom-up computations. The top-down expectations can cause us to misperceive things by resetting some of the circuitry in the bottom-up processors. This is partly the neural basis for perceptual completion and other illusions.

The top-down and bottom-up processes inform each other in an ongoing fashion. At the same time as features are being analyzed individually, parts of the brain that are higher up—that is, that are more phylogenetically advanced, and that receive connections from lower brain regions—are working to integrate these features into a perceptual whole. The brain constructs a representation of reality, based on these component features, much as a child constructs a fort out of Lego blocks. In the process, the brain makes a number of inferences, due to incomplete or ambiguous information; sometimes these inferences turn out to be wrong, and that is what visual and auditory illusions are: demonstrations that our perceptual system has guessed incorrectly about what is out-there-in-the-world.

The brain faces three difficulties in trying to identify the auditory objects we hear. First, the information arriving at the sensory receptors is undifferentiated. Second, the information is ambiguous—different objects can give rise to similar or identical patterns of activation on the eardrum. Third, the information is seldom complete. Parts of the sound may be covered up by other sounds, or lost. The brain has to make a calculated guess about what is really out there. It does so very quickly and generally subconsciously. The illusions we saw previously, along with these perceptual operations, are not subject to our awareness. I can tell you, for example, that the reason you see triangles where there are none in the Kaniza figure is due to perceptual completion. But even after you know the principles that are involved, it is impossible to turn them off. Your brain keeps on processing the information in the same way, and you continue to be surprised by the outcome.

Helmholtz called this process “unconscious inference.” Rock called it “the logic of perception.” George Miller, Ulrich Neisser, Herbert Simon, and Roger Shepard have described perception as a “constructive process.” These are all ways of saying that what we see and hear is the end of a long chain of mental events that give rise to an impression, a mental image, of the physical world. Many of the ways in which our brains function—including our senses of color, taste, smell, and hearing—arose due to evolutionary pressures, some of which no longer exist. The cognitive psychologist Steven Pinker and others have suggested that our music-perception system was essentially an evolutionary accident, and that survival and sexual-selection pressures created a language and communication system that we learned to exploit for musical purposes. This is a contentious point in the cognitive-psychology community. The archaeological record has left us some clues, but it rarely leaves us a "smoking gun" that can settle such issues definitively. The filling-in phenomenon I've described is not just a laboratory curiosity; composers exploit this principle as well, knowing that our perception of a melodic line will continue, even if part of it is obscured by other instruments. Whenever we hear the lowest notes on the piano or double bass, we are not actually hearing 27.5 or 35 Hz, because those instruments are typically incapable of producing much energy at these ultralow frequencies: Our ears are filling in the information and giving us the illusion that the tone is that low.

We experience illusions in other ways in music. In piano works such as Sindig's "The Rustle of Spring" or Chopin's Fantasy-Impromptu in C-sharp Minor, op. 66, the notes go by so quickly that an illusory melody emerges. Play the tune slowly and it disappears. Due to stream segregation, the melody "pops out" when the notes are close enough together in time—the perceptual system holds the notes together—but the melody is lost when its notes are too far apart in time. As studied by Bernard Lortat-Jacob at the Musée de l'Homme in Paris, the Quintina (literally "fifth one") in Sardinian a capella vocal music also conveys an illusion: A fifth female voice emerges from the four male voices when the harmony and timbres are performed just right. (They believe the voice is that of the Virgin Mary coming to reward them if they are pious enough to sing it right.)

In the Eagles' "One of These Nights" (the title song from the album of the same name) the song opens with a pattern played by bass and guitar that sounds like one instrument—the bass plays a single note, and the guitar adds a glissando, but the perceptual effect is of the bass sliding, due to the Gestalt principle of good continuation. George Shearing created a new timbral effect by having guitar (or in some cases, vibrophone) double what he was playing on the piano so precisely that listeners come away wondering, "What is that new instrument?" when in reality it is two separate instruments whose sounds have perceptually fused. In "Lady Madonna," the four Beatles sing into their cupped hands during an instrumental break and we swear that there are saxophones playing, based on the unusual timbre they achieve coupled with our (top-down) expectation that saxophones should be playing in a song of this genre.

Most contemporary recordings are filled with another type of auditory illusion. Artificial reverberation makes vocalists and lead guitars sound like they're coming from the back of a concert hall, even when we're listening in headphones and the sound is coming from an inch away from our ears. Microphone techniques can make a guitar sound like it is ten feet wide and your ears are right where the soundhole is—an impossibility in the real world (because the strings have to go across the soundhole—and if your ears were really there, the guitarist would be strumming your nose). Our brains use cues about the spectrum of the sound and the type of echoes to tell us about the auditory world around us, much as a mouse uses his whiskers to know about the physical world around him. Recording engineers have learned to mimic those cues to imbue recordings with a real-world, lifelike quality even when they're made in sterile recording studios.

There is a related reason why so many of us are attracted to recorded music these days—and especially now that personal music players are common and people are listening in headphones a lot. Recording engineers and musicians have learned to create special effects that tickle our brains by exploiting neural circuits that evolved to discern important features of our auditory environment. These special effects are similar in principle to 3-D art, motion pictures, or visual illusions, none of which have been around long enough for our brains to have evolved special mechanisms to perceive them; rather, they leverage perceptual systems that are in place to accomplish other things. Because they use these neural circuits in novel ways, we find them especially interesting. The same is true of the way that modern recordings are made.

Our brains can estimate the size of an enclosed space on the basis of the reverberation and echo present in the signal that hits our ears. Even though few of us understand the equations necessary to describe how one room differs from another, all of us can tell whether we're standing in a small, tiled bathroom, a medium-sized concert hall, or a large church with high ceilings. And we can tell when we hear recordings of voices what size room the singer or speaker is in. Recording engineers create what I call "hyperrealities," the recorded equivalent of the cinematographer's trick of mounting a camera on the bumper of a speeding car. We experience sensory impressions that we never actually have in the real world.

Our brains are exquisitely sensitive to timing information. We are able to localize objects in the world based on differences of only a few milliseconds between the time of arrival of a sound at one of our ears versus the other. Many of the special effects we love to hear in recorded music are based on this sensitivity. The guitar sound of Pat Metheny or David Gilmour of Pink Floyd use multiple delays of the signal to give an otherwordly, haunting effect that triggers parts of our brains in ways that humans had never experienced before, by simulating the sound of an enclosed cave with multiple echoes such as would never actually occur in the real world—an auditory equivalent of the barbershop mirrors that repeated infinitely.

Perhaps the ultimate illusion in music is the illusion of structure and form. There is nothing in a sequence of notes themselves that creates the rich emotional associations we have with music, nothing about a scale, a chord, or a chord sequence that intrinsically causes us to expect a resolution. Our ability to make sense of music depends on experience, and on neural structures that can learn and modify themselves with each new song we hear, and with each new listening to an old song. Our brains learn a kind of musical grammar that is specific to the music of our culture, just as we learn to speak the language of our culture.

Noam Chomsky's contribution to modern linguistics and psychology was proposing that we are all born with an innate capacity to understand any of the world's languages, and that experience with a particular language shapes, builds, and then ultimately prunes a complicated and interconnected network of neural circuits. Our brain doesn't know before we're born which language we'll be exposed to, but our brains and natural languages coevolved so that all of the world's languages share certain fundamental principles, and our brains have the capacity to incorporate any of them, almost effortlessly, through mere exposure during a critical stage of neural development.

Similarly, it seems that we all have an innate capacity to learn any of the world's musics, although they, too, differ in substantive ways from one another. The brain undergoes a period of rapid neural development after birth, continuing for the first years of life. During this time, new neural connections are forming more rapidly than at any other time in our lives, and during our midchildhood years, the brain starts to prune these connections, retaining only the most important and most often used ones. This becomes the basis for our understanding of music, and ultimately the basis for what we like in music, what music moves us, and how it moves us. This is not to say that we can't learn to appreciate new music as adults, but basic structural elements are incorporated into the very wiring of our brains when we listen to music early in our lives.

Music, then, can be thought of as a type of perceptual illusion in which our brain imposes structure and order on a sequence of sounds. Just how this structure leads us to experience emotional reactions is part of the mystery of music. After all, we don't get all weepy eyed when we experience other kinds of structure in our lives, such as a balanced checkbook or the orderly arrangement of first-aid products in a drugstore (well, at least most of us don't). What is it about the particular kind of order we find in music that moves us so? The structure of scales and chords has something to do with it, as does the structure of our brains. Feature detectors in our brains work to extract information from the stream of sounds that hits our ears. The brain's computational system combines these into a coherent whole, based in part on what it thinks it ought to be hearing, and in part based on expectations. Just where those expectations come from is one of the keys to understanding how music moves, when it moves us, and why some music only makes us want to reach for the off button on our radios or CD players. The topic of musical expectations is perhaps the area in the cognitive neuroscience of music that most harmoniously unites music theory and neural theory, musicians and scientists, and to understand it completely, we have to study how particular patterns of music give rise to particular patterns of neural activations in the brain.