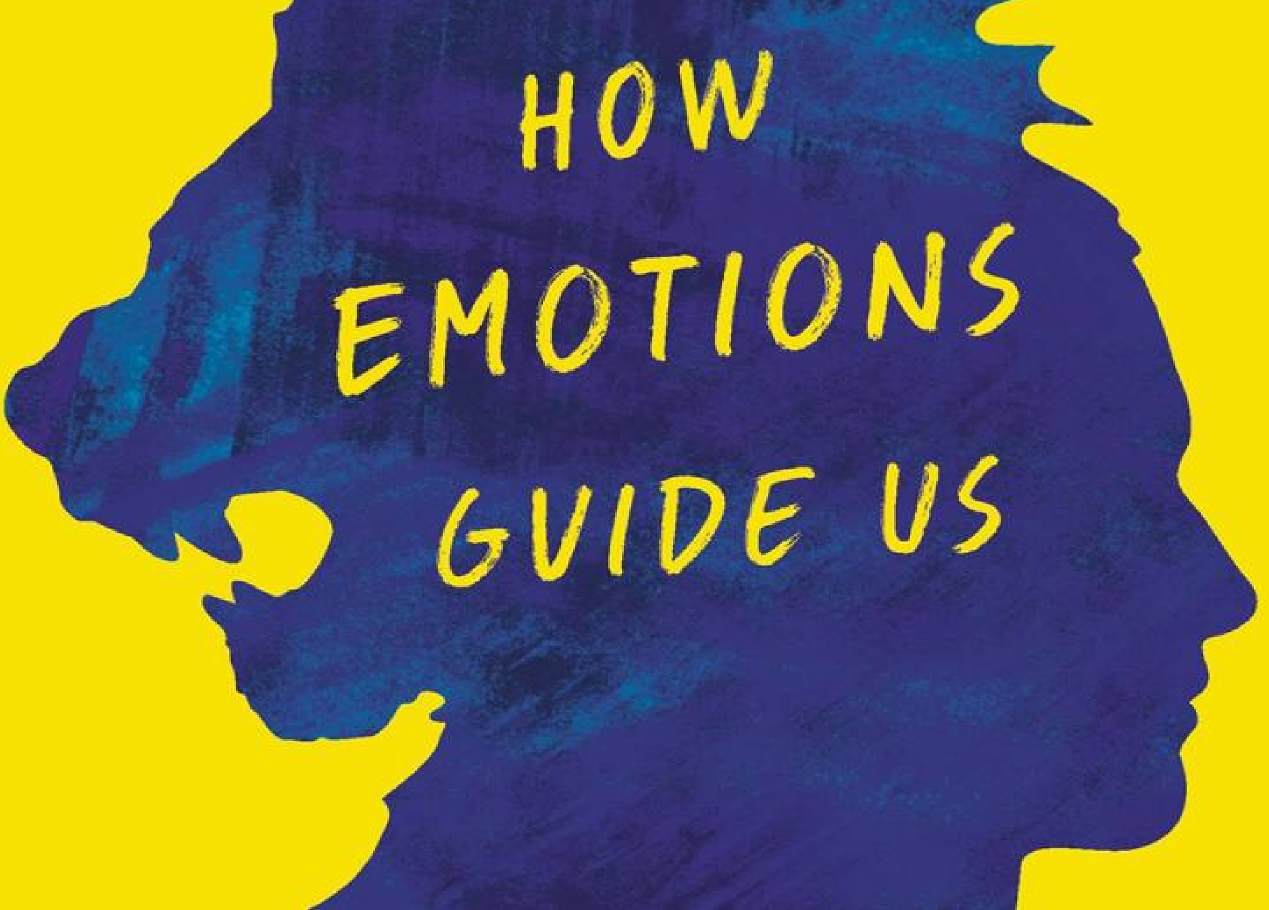
DAVID J. ANDERSON

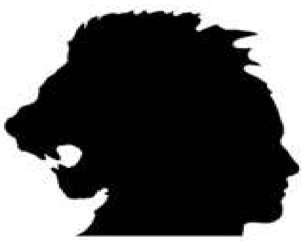
THE NATURE OF THE BEAST



HOW EMOTIONS GUIDE US

THE NATURE OF THE BEAST

How Emotions Guide Us



DAVID J. ANDERSON

BASIC BOOKS
New York

Basic Books
Hachette Book Group
1290 Avenue of the Americas, New York, NY 10104
www.basicbooks.com

First Edition: March 2022

Published by Basic Books, an imprint of Perseus Books, LLC, a subsidiary of Hachette Book Group, Inc. The Basic Books name and logo is a trademark of the Hachette Book Group.

Library of Congress Cataloging-in-Publication Data

Names: Anderson, David J., 1956– author.
Title: The nature of the beast : how emotions guide us / David J. Anderson.
Description: First edition. | New York : Basic Books, 2022. | Includes

bibliographical references and index.
Identifiers: LCCN 2021034465 | ISBN 9781541674639 (hardcover) | ISBN 9781541674646 (ebook)
Subjects: LCSH: Emotions. | Affective neuroscience.
Classification: LCC BF511 .A47 2022 | DDC 152.4—dc23

LC record available at https://lccn.loc.gov/2021034465

ISBNs: 9781541674639 (hardcover), 9781541674646 (ebook)

E3-20220112-JV-NF-ORI

Contents

BASIC
BOOKS

For my parents, Jim and Helene,
who married science and the humanities

Introduction

A CHILD PUMMELS A CLASSMATE IN A SCHOOL-YARD. Insults traded in a tavern escalate into a bar fight. Two rams butt horns on a mountain precipice. A pride of lionesses takes down a Cape buffalo. A malevolent mob storms the Capitol. A mass shooter slaughters scores of concertgoers from a hotel window.

All these examples are naked acts of aggression. They are overt behaviors that can be recognized and recorded in words or on video. However, at least in the case of humans, they are also accompanied by something we cannot see: an internal, subjective experience of emotion by the person committing the act. Witness any of the above situations, and you might assume that the assailants feel anger, rage, fury. But unless they specifically tell us how they feel, how can we ever really know what emotion they are experiencing—or, indeed, if they are experiencing any emotion at all? And what about the experience of animals, whom we cannot even ask?

Clearly, there is some connection between aggressive behavior and things like anger and rage. It is difficult to imagine being driven to commit such violent acts without an underlying emotional drive. Yet our intuition and experience tell us that anger and aggression are not the same thing. After all, one can certainly feel angry without physically expressing that feeling as aggression. Is anger generated separately and independently from aggressive behavior in the brain, or are they different manifestations of the same process? Does anger cause aggression, does aggression cause anger, or is there no causal relationship between the two? Ultimately, these questions can only be attended to after we've answered a more fundamental one: What exactly are emotions, and what do they do for us?

The problem is that, despite centuries of inquiry on the topic, we don't know. Not only that, but scientists cannot even agree on what an answer should look like. There are easily half a dozen different perspectives people bring to the study of emotion: psychological, cognitive, sociological, anthropological, philosophical, and neuroscientific.

Researchers in these disciplines don't speak the same language or understand things in the same terms. A psychologist seeks to explain emotions in terms of human drives, needs, and conflicts. A neuroscientist seeks to explain them in terms of patterns of brain activity. Some people want to explain a particular emotion, like sadness or fear; others want to understand in a general way what makes emotion different from other kinds of brain processes. Not only are we the proverbial blind men grasping different parts of the elephant and trying to describe what we are holding, but we don't even have the same word for "elephant."

Given this intellectual diversity and lack of consensus, it is not surprising that there have been many books espousing new theories of emotion. Most of them—certainly most of the theories that make their way into the public conversation—originate in psychology. Such theories are interesting and powerful, but they are often very abstract and hard to falsify by experiment. The perspective of this book is very different. I believe that neuroscience can offer us a way of thinking about emotions that is objective and empirical, and that gives us the tools to begin studying questions about emotion that have been so difficult that some have written them off as unsolvable.

But first, a little background. Neuroscientists study diverse animal species, including humans, to understand how the neurons and circuits within their brains give rise to internal drives and behavior. Neuroscientists use a variety of methods to measure and manipulate the activity of neurons in the brain. They also build computer models of brain function based on these data. This allows for an intricate, causal understanding of brain function at its most basic level. Neuroscience offers the hope that, when we understand the brain in sufficient detail, we will be able to explain how its activity gives rise to both behavior *and* emotion.

Perhaps you already agree with me about the value of neuroscience. Maybe you even believe I am behind the times and that this question has already been settled. For example, you may have read that fear is produced by activity in a brain structure called the amygdala. Aren't there many studies in which neuroscientists put people in brain scanners and show that when they're afraid, their amygdala lights up (that is, becomes more active)? If that is the case for fear, then surely something similar must also be true for anger. So isn't the only thing left to do to figure out where anger and rage "live" in the brain? Can't we just put people in brain scanners, have someone watch what happens in their brain when they get angry, and then repeat the same process for different emotions until we have mapped them all?

In a word, no. Brain-scanning experiments (technically known as functional magnetic resonance imaging, or fMRI) do not directly visualize electrical activity in the brain; rather, they image blood flow to a particular area. Consequently, they provide only a very coarse-grained view of brain activity. More important, such experiments only provide correlative data. If we observe activity somewhere in a person's brain while they say they are afraid or angry, we cannot know whether the brain activity is causing the emotion or the emotion is causing the brain activity. Furthermore, relying on what a subject says they are feeling—that is, verbal report—is not necessarily an accurate way to assess that person's subjective feelings. A subject can mischaracterize or even lie about what they feel. What's more, it is quite difficult, in practice, to induce an authentic, bona fide emotion in a human subject in a brain scanner because the subjects know they are in an experiment and have all kinds of distractions, like the noise in the scanner and people running around in white lab coats. Finally, early fMRI studies of fear in different laboratories produced inconsistent results, as psychologist and author Lisa Feldman Barrett has shown (although more recent studies have produced more consistent results).

These and other factors have led in recent years to a sort of debunking of simplistic neuroscientific explanations of emotion based on human brain-scanning experiments. For example, Feldman Barrett argued in a 2015 *New York Times* opinion piece that neural activity in the amygdala is not in fact the source of fear; instead, fear and other emotions are diffusely distributed across the brain rather than localized to any single brain region. In another op-ed, Feldman Barrett argued further that anger in humans comes in so many different forms that looking for brain activity that singularly represents this emotion is, effectively, an exercise in futility. From this perspective, efforts to understand anger—or any emotion, for that matter—at the level of neuroscience seem fundamentally flawed, and the nature of emotion itself seems to remain beyond our grasp.

What I hope to show you is that such dismissals are too quick. Neuroscience has something real to tell us about how emotions work; we've just been going about it in the wrong way. Over the past two decades, that has begun to change with the development of revolutionary new techniques for understanding brain function in so-called model organisms, animals that are bred in the laboratory for research and are amenable to genetic manipulations, like mice and fruit flies. These methods use genes and light to mark, map, measure, and manipulate specific types of neurons in the brain. Unlike brain-scanning experiments, which measure neural activity indirectly, as blood flow to the brain, these new methods can directly measure electrical activity in individual neurons and can map their direct connections to particular cells in other brain regions.

These techniques allow specific sets of neurons to be turned on or off at will, to determine how that affects specific behaviors. Unlike brain-scanning experiments, such experiments can distinguish cause from effect. Fittingly, I refer to such experiments as *causal neuroscience*. They will help us arrive at a deeper conceptual understanding of emotion, but they have practical implications as well. Distinguishing cause and effect is crucial if you want to, for example, find the correct brain targets for new psychiatric drugs or therapies utilizing deep brain stimulation.

Overwhelmingly, however, emotions are still broadly viewed and explained in psychological terms. To be clear, there is nothing inherently wrong with such explanations. From a utilitarian viewpoint, however, if psychological explanations were enough, then talk therapy would be sufficient to successfully treat most mental illnesses, including post-traumatic stress disorder (PTSD), attention deficit hyperactivity disorder (ADHD), major depression, bipolar disorder, and schizophrenia, to name just a few. Certainly, talk therapy can be useful for some patients. But in many cases—particularly those involving severe mental illnesses—it is not. At that point drugs are used (albeit often together with talk therapy). The problem is that we don’t have good drugs to treat or cure most psychiatric disorders, and the drugs that we do have often cause side effects that can be so unpleasant and debilitating that many people stop taking them and suffer the consequences (like the brilliant novelist David Foster Wallace, author of *Infinite Jest*, who discontinued his depression meds because of their side effects and wound up committing suicide).

The sad fact is that there hasn’t been a fundamentally new psychiatric drug approved in the last 50 years. All the “new” drugs being released are just variants on the same basic theme—for example, selective serotonin reuptake inhibitors (SSRIs) like Prozac, Paxil, and Lexapro. The reason is that most of the drugs that we do have—like SSRIs—were discovered by chance. Such lucky accidents don’t happen very often, and many people suffer while we are waiting for the next one to occur. We need to be able to have a way to develop new psychiatric therapies by design, through an understanding of underlying disease mechanisms.

Causal neuroscience offers that hope. For example, if the activity of a certain type of neuron is correlated with anxiety, it could mean either that the neuron's activity causes anxiety or that anxiety causes the neuron to be active. If turning that neuron off makes an anxious animal more "chill" while activating the neuron makes the animal more anxious, it would suggest that those neurons cause anxiety. If such manipulations have no effect, it would suggest that the neuron's activity is a consequence, not a cause, of the animal's anxiety state. That outcome matters if you are trying to decide which type of neuron to study in order to search for a new treatment for anxiety disorders.

ADMITTEDLY, THESE NEW METHODS of causal neuroscience are difficult to apply in humans, for both technical and ethical reasons. Our brains are large, complex organs, and it is challenging to find ways to reliably stimulate or inhibit tiny, specific areas that might isolate specific brain functions and activity. Furthermore, such methods are *invasive*: they require surgery to open up the brain and insert electrodes, optical fibers, or other equipment. In humans, medical ethics dictate that brain surgery can only be performed to treat an illness, such as epilepsy, and any recordings of brain activity have to be restricted to the affected region. Therefore, neurosurgeons can't just stick an electrode anywhere they want to in a healthy person's brain, start stimulating, and see what happens. That makes it impossible to do a systematic, brain-wide search for regions controlling different emotional feelings. In addition, causal neuroscience studies usually require heritable modifications of an animal's genes, something that is currently prohibited in humans. It also involves injections of inactivated viruses into the brain in order to genetically modify the neurons of interest. In humans, this would only be allowed to treat a specific illness, such as brain cancer.

For these reasons, if we want to use these new methods to achieve a *causal* understanding of how emotions like fear and anger are generated by the brain and linked to behaviors like aggression—one that will impact human health—then we need to work on animal models. Already, the application of causal neuroscience in animals has had an enormous impact on our understanding of brain processes such as vision, perception, learning, memory, and motor function, to name just a few. There is every reason to think that they should have a similar impact on our understanding of emotion and of the relationship of emotion to action.

But now we arrive at a significant problem. How do you measure emotions in animals? Most people use the word "emotion" in everyday speech to refer to "feelings." Feelings are subjective experiences that we become consciously aware of through introspection. Scientifically, the only way to assess subjective feelings is by verbal report: the researcher asks the subject how they feel, and the subject describes their feelings. Since animals can't talk, we have no way of knowing what they are feeling—or indeed, if they are feeling anything at all (in the sense that we experience feelings). Subjective feelings are a manifestation of conscious awareness, and there is currently no way to objectively determine whether a non-human animal is conscious. Therefore, if we consider emotions exclusively as "feelings," we cannot know whether they are an attribute of animals. As the Nobel Prize–winning Dutch ethologist Niko Tinbergen (one of my scientific heroes) wrote: "Hunger, like anger, fear and so forth, is a phenomenon that can be known only by introspection. When applied to another species, it is merely a guess about the possible nature of the animal's subjective state."[1]

From this perspective, the observation that an animal is fighting doesn't necessarily mean that it is also experiencing what we would call "anger." The fact that it is freezing doesn't necessarily mean that it is experiencing what humans label as "fear." Action is one thing; emotion is something else altogether. In this view, the observation that an animal is exhibiting a particular behavior doesn't necessarily mean that it has any emotions at all.

Those of you who (like myself) are pet owners may find this view patently absurd. Most of us feel strongly that we can intuit how an animal is feeling just by looking at it. For example, I'm pretty sure I can tell whether my cat is happy or alarmed by looking at her body language and her face. If she looks content or frightened, then it seems obvious that she must be *experiencing* those emotions (i.e., that she is aware of them), just as we would. If you think this way, you are in good company. Charles Darwin, the great naturalist (and another one of my scientific heroes), wrote in his 1872 book *The Expression of the Emotions in Man and Animals* that "even insects express anger, terror, jealousy and love by their stridulation [rubbing their wings together to produce sound]."[2]

The assumption that animals feel the way we do seems right and true, particularly when we think about the reaction of our pet dog or cat to events that we view as fearful or threatening. But what of the reaction of a fish, a fly, or a bee to a threatening event? Should we attribute emotions to those animals as well, as Darwin assumed we could? Or should we remain agnostic until we can find a more objective way to determine whether a particular animal species has emotions?

Darwin's assumption was fine for his purpose, which was to explain the evolutionary benefit of specific "emotional" behaviors that people and some animals share, like why our eyes widen when we are afraid.[i] However, for a hard-nosed neuroscientist like me, this assumption is problematic for several reasons. First, if we define emotions as feelings, then, as Tinbergen said, we can't objectively know if an animal has any emotions at all. Second, if we simply assume, like Darwin, that all animals have emotions, then in order to infer what *kind* of emotion a given creature is feeling, we have to attribute to that animal the same emotions we would feel under similar conditions. But animals are not little people in furry costumes, and so our intuition may mislead us. For example, if I see my cat roll on her back with her paws in the air when I come home from work, I infer that she is happy to see me—because *I* would be happy to see me if I were the cat locked in the house alone all day. However, I have no independent, objective way of knowing how or what my cat is feeling, other than by observing her behavior. I can't both explain her behavior by assuming I know her feelings *and* decide what she is feeling by observing her behavior—that's circular logic. Maybe she has simply learned that she can train me to rub her belly by rolling on her back with her paws in the air.

To compound this problem, a given behavior may reflect one of several possible underlying emotions, which can be difficult to distinguish. An animal may be immobile because it is frozen in fear—or because it is asleep. It may attack another because it is threatened, because it is establishing dominance, or because it wants to eat it. Similarly, when a male animal of a given species mounts another male, is that a homosexual behavior that reflects love or affiliation, or is it a dominance display? Trying to infer the particular emotion state that an animal is in, just from observing its behavior, can be difficult.

Finally—and most important, from the perspective of this book—many animal behaviors that *look* "emotional" to us, through our anthropomorphic lens, may simply be automatic, genetically pre-programmed responses hardwired to specific external sensory cues. Such behaviors are essentially reflexes, much like the type you experience when a doctor raps your knee with a little hammer and your leg extends. As Max Planck Institute cyberneticist Valentino Braitenberg has shown, it is easy to program a four-wheeled vehicle to behave in a way that deceives people into thinking it is exhibiting an emotion (attraction or disgust) when in fact its movements are simply being controlled by sensors that are wired to move the wheels clockwise or counterclockwise, like a Mars rover (more on this in Chapter 2). So if a mouse moves away from a hot surface, does that necessarily mean that it is in a state of pain? Or is it just exhibiting a reflex? Even in humans, the rapid withdrawal of your hand from a hot stovetop is controlled by a reflex in your spinal cord that bypasses your brain altogether. (The pain you feel is something different; that happens afterward, in your brain.) By the same token, a mouse or a fly that jumps or freezes in response to a threat may just be exhibiting a reflex, without any accompanying internal state of fear.

So if neuroscience is able to offer a better way of thinking about emotions than we currently have, as I believe it is, then we now know what it must do. First, it should operationally redefine "emotion" in a way that does not require attributing feelings to animals. Second, it needs to be able to distinguish whether a given animal's behavior expresses any emotion at all or is just an automatic reflex. Third, it must offer a way to determine what *kind* of emotion the animal is having, without falling back on attributing our own subjective human experiences to it. Finally, it must be able to show us that learning about how emotions work in animal brains can tell us something about how they work in our own brains.

TO GIVE YOU A glimpse of what I'll be talking about, let's start with the first and toughest question: Is there a way to objectively identify instances of emotional expression in animals without attributing human-like feelings to them? As I often do when I get stuck thinking about brains and behavior, I turn to my cats for inspiration: Serafina, a delicate, sensitive calico with an inquisitive, searching face that looks almost human, and Buster, a gray tabby who was rescued as a kitten from the bushes behind my research lab at Caltech and has a yellow-eyed, inscrutable stare and an energy and fearlessness that betray his feral origins. Serafina has never adjusted to Buster, whom we took in about a year after we adopted her. Despite his desperate attempts to win her affections, she studiously avoided him. When he would pester her too persistently, she would hiss and swat him away, often drawing blood with her claws. Then she would groom herself and return to her usual meditative state as Buster escaped to another room.

Buster, on the other hand, would never hurt a fly. Even when he is roughhousing with me, his claws remain sheathed and he only play-bites. Serafina, on the other hand, quickly gets irritated if I play too roughly with her and lacerates my hands with her razor-sharp teeth and claws. However, there was one occasion on which Buster was utterly transformed. Buster is an indoor cat, and one day he confronted through our glass back door a large, unfamiliar gray tomcat who had strayed into our yard. As they stood nose to nose, Buster released a deep moan from his throat, his back arched, and the fur on his tail puffed up to several times its normal thickness. He remained like that, staring out the window and moaning, until the tomcat left. Shortly after that, Serafina wandered into the room near Buster, and he suddenly wheeled and attacked her with a fury and viciousness that I had never seen before, leaving a nasty scratch on her nose that took weeks to heal.

Clearly, Buster’s attack on Serafina was more than just a reflexive, automatic response to the tomcat. Apparently Buster had been put into some kind of amped-up state that persisted even after the tomcat wandered away, and which he eventually released by attacking Serafina. Was Buster *aware* of the fact that he was in this state? Did he experience a subjective feeling, akin to what we humans experience as anger or rage, during this episode? Maybe he did, or maybe he didn’t—there was no way to know. I realized, however, that this question didn’t matter if what I wanted to understand was how the brain produced that state. By analogy, a rock sitting in the sun is hot, and the same rock in the middle of the night is cool: it’s in a type of physical state that is characterized by how much heat energy it contains. I’m pretty sure that the rock isn’t subjectively aware of its heat, but we can measure it with a thermometer nevertheless.

Unfortunately, in the case of emotion there’s no such device. It’s not like I can go to the hardware store, purchase a Rage-O-Meter for $9.95 plus tax, and stick it up Buster’s rectum. (*That* would be a challenge!) Yet the experience with the tomcat had clearly caused *something* to happen in Buster’s brain and body that persisted long after the tomcat disappeared, and that fundamentally altered his behavior toward Serafina. How could I scientifically study that slippery “something” without attributing subjective feelings to my cat? (I should note that when I go home and take off my scientist hat, I treat my cats as if they *do* have subjective feelings, and I believe that dogs and many other mammalian species do as well. However, when I go into the lab, I have to check my beliefs at the door and maintain scientific objectivity.)

The key intellectual adjustment involved letting go of the idea that emotions consist exclusively of subjective feelings. Rather, they are *internal, central states* of the brain that can exist independently of whether the owner of that brain has any conscious awareness of them or not. We know anecdotally that even we humans can sometimes have emotions that we are not consciously aware of but which our friends and loved ones can infer from our body language or facial expression. And there are laboratory studies that have provided evidence of unconscious emotions in humans as well. If emotions can exist independently of consciousness in people, then they may exist in animals as well, whether or not we consider those species to be "conscious." This doesn't mean that I think that cats, dogs, and other animals don't have subjective feelings; it just means that I don't need to answer that question scientifically in order to study how their brains generate emotion states.

But what exactly do I mean by a "state" of the brain? Brain states function to control the way that information from the outside world is interpreted by the brain and converted into action. For example, if you were starving, a plate of cold french fries in a puddle of congealed grease would elicit ravenous feeding, but if you were full after a meal, the same stimulus could make you recoil. Most animals would behave in exactly the same way—including fruit flies. You may have a subjective feeling of being starved or sated, but the feeling is not what controls your behavior. Your brain state controls your behavior, and your subjective feelings are your conscious experience of your brain at work—the brain's perception of its own internal state.

Some brain states can prevent you from responding to a stimulus at all. Sleep is such a state. When you are fast asleep, you don't hear subtle noises that you would otherwise easily detect if you were awake. If the noises get loud enough, of course, they can wake you up. But the point is that when you are in the sleep state, the same sensory stimulus is processed differently by the brain compared to when you are awake. And a person doesn't need to be consciously aware of the fact that they are asleep in order for that state to suppress their responses to external stimuli. Similarly, an animal doesn't need to have a subjective experience of thirst in order for its brain to tell its body to find and consume water when it is dehydrated.

In other words, emotions are internal states that control how the brain's input is converted into its output, like a supervisor directing workers how to connect calls at an old-fashioned telephone switchboard. Externally visible behavior is one such output, or "readout," of the internal emotion state. But there are other measurable readouts that can occur internally, such as changes in heart rate, blood pressure, or hormone levels. Subjective feelings—our conscious awareness of these internal states—are just another such readout. However, they are neither the only one nor the essential one. Therefore, by considering emotions as functional internal states, we can study how the brain controls emotions in animal models without having to assume or figure out whether animals do or don't have subjective feelings.

Neuroscience seeks to understand how internal emotion states are generated, or "implemented," by the brains of humans and other species. This implementation may involve special global patterns of electrical activity (for example, certain kinds of brain waves), increases in certain neurotransmitters or other brain chemicals, or activity in specific brain regions, types of neurons, and neuronal circuits. In principle, if we understood how the brain generates a particular emotion state, then we should be able to measure electrical activity or chemicals in an animal's brain and identify what emotion state it is in. Conversely, if we knew what emotion state an animal was in (from measuring its behavior and physiology), then we should be able to predict what its patterns of brain activity and chemistry should look like. But in order to do this, we first need to be able to determine whether a behaving animal is *in* any kind of emotion state at all—let alone what *kind* of emotion it is having.

Why is this even an issue? What is the alternative? Isn't it obvious that if an animal is freezing or running away, it is experiencing fear? While that may be our intuition, from a neuroscientific perspective it is actually not a foregone conclusion. The reason is that animal behaviors that may *look* superficially like they express an underlying emotion state may in fact just be automatic, hardwired reflexes. As I will describe in Chapter 2, it is possible to design a fairly simple robot that displays behaviors toward a stimulus (e.g., a light source) that look superficially like "love," "fear," and "hate." The robot is programmed by engineers to respond to the stimulus in a fixed way that depends on the robot's orientation to the