CHAPTER 1

What Don’t We Know about Emotions?

The greatest enemy of knowledge is not ignorance, it is the illusion of knowledge.
—Ralph J. Boorstin

If you are like most people, you feel convinced that, because you have emotions, you know a lot about what emotions are, and how they work. We believe you are almost certainly wrong. In the field of emotion, as in most fields, familiarity is not the same as expertise. After all, you have a heart, but that doesn’t make you an expert on hearts. You leave that to your cardiologist.

Yet the science of emotion is fraught with this problem: everyone seems to think they know what an emotion is. This is because we all have strong, and typically unjustified, intuitive beliefs about emotions. For instance, some people are absolutely certain that animals have emotions; others are absolutely certain that animals could not have emotions. Neither camp can usually give you convincing reasons for their beliefs, but they stick to them nonetheless.

We cannot emphasize enough the pervasive grip that our commonsense view of emotions has on how we (that is, researchers in the field) frame our scientific questions. We need to free ourselves of our commonsense assumptions—or at least question all of them—if we want to ask the right questions in the first place. This chapter introduces the topics of this book through this important premise and concludes by listing what we ideally would want from a mature science of emotion, and what entries in this list we will tackle in this book.

We wrote this book for two overarching aims. The first aim is to motivate the topic of emotion, to note that it is of great interest not only to laypeople but also to many scientific fields of study, and that it is a
very important topic as well. At the same time, we emphasize that we currently know remarkably little about it yet—in particular, we know a lot less than we think we know. This is good news for scientists: there is work to be done, interesting and important work.

The second aim is to provide a summary of what we do know and to sketch a framework within which to understand those empirical findings and within which to formulate new questions for the future. This process is in practice very piecemeal: we need to have a little bit of data even to begin thinking about what emotions are, but then we discover problems with the way prior experiments were done and interpreted. In the dialectic of actual scientific investigation, both conceptual framework and empirical discovery are continuously revised, and inform each other. However, we have not written our book this way. Instead, we begin with some of the foundations for a science of emotion (chapter 2)—what kinds of ontological and epistemological commitments it requires, what kind of structure an explanation takes—and then work our way toward a list of features or properties of emotions (chapter 3), which then finally are the things we look for, and discover, through empirical research (chapters 4–9). We return to the foundations and the questions again in chapters 10 and 11 by contrasting our views with those of others, and by suggesting some experiments for the future.

**Emotions According to Inside Out**

What is it about emotions that we would like to understand? And what do we think we understand, but in fact don’t (or are mistaken about)? Because emotions are ubiquitous in our lives, and integral to our experience of the world, it is dangerously easy to come up with simplistic views that do not stand up to closer scrutiny, and instead impede scientific progress because they create “the illusion of knowledge.”

The film *Inside Out*, which won the 2016 Academy Award for Best Animated Feature, as well as a Golden Globe, provides a good example of many common but incorrect assumptions about emotion. As you watch the film, you get a fanciful view of how emotions are supposed to work inside a twelve-year-old girl, how those emotions are supposed to be integrated with memory and personality, and how they are supposed
to be expressed as behavior. If *Inside Out*’s view of emotion were right, you would be tempted to conclude that we understand an enormous amount about how emotions work—and, more generally, about how the mind and brain work. But *Inside Out*’s view of how emotions work is wrong. In examining what, exactly, is wrong with it, we can highlight some of the gaps in our current understanding of emotion. If you’ve seen the film and you already find the view of emotion portrayed by *Inside Out* silly, you are ahead of the game—but bear with us as we use it as an example for uncovering problematic beliefs about emotion.

*Inside Out*’s view of emotion takes as its starting premise the idea that all our emotions boil down to a few primary ones: in the film, they are joy, anger, fear, sadness, and disgust. These five emotions are animated as different characters, charming little homunculi that live in the brain of the little girl and fight with each other for control of her behavior and mental state. These homunculi sit at a control panel and watch the outside world on a screen. They react to the outside world, and in response they manipulate levers and switches that control the little girl’s behavior. They are also affected by memories that are symbolized by transparent marbles; moreover, a series of theme parks provide a mental landscape symbolizing different aspects of the girl’s personality. The five emotion characters fight over access to the memory marbles and struggle to keep the girl’s theme-park attractions open for business.

From the film’s point of view, the five emotions are the dominant force controlling the little girl’s thoughts, memories, personality, and behavior; thinking, reasoning, and other cognitive activities are relegated to a sideshow. Truly, the little girl is an entirely emotional being. These details of the movie may not represent the way you think about emotions, but they characterize how many people do.

So what’s wrong with the film’s creative, engaging metaphor? Let’s unpack a few of the key ideas about emotions that *Inside Out* showcases, highlight the errors in their underlying assumptions, and try to articulate the scientific questions that they raise. Although science may not yet have the answers, the exercise will help us frame the issues.

**Idea 1. There are a few primary emotions.** The prevailing view, enshrined in many psychology textbooks, is that there is a small set of
“primary” or “basic” emotions: as we already mentioned, these are joy, anger, fear, sadness, and disgust, according to *Inside Out*. Different scientific emotion theories offer a big range in the number of basic emotions—anywhere from two to eleven! A second type of emotion is often called “social” or “moral” emotion and typically includes shame, embarrassment, pride, and others. These social emotions are thought to be more essentially tied to social communication than the basic emotions are. But although there are multiple schemes, many classic emotion theories tend to share the idea of a fixed, and relatively small, set of emotions that correspond to the words we have for emotions in English.

The idea of a small set of basic emotions was most notably introduced by the psychologist Paul Ekman, based largely on data from his studies of emotional facial expressions in humans. Ekman argued that facial expressions of basic emotions can be recognized across all human cultures (Ekman 1994); he studied them even among tribes in New Guinea. Ekman’s set of basic emotions includes happiness, surprise, fear, anger, disgust, and sadness (although contempt is also sometimes included). The neurobiologist Jaak Panksepp similarly proposed a set of basic emotions, derived from his observations of animal behavior: seeking, rage, fear, lust, care, panic, and play (Panksepp 1998). These emotion theories have much to recommend them and stimulated entire lines of important research. But they also suggest two questionable background assumptions (which Ekman and Panksepp themselves may or may not have held).

*Questionable assumption 1: Emotions (at least the “primary” ones) are irreducible.* A presumption that often accompanies the idea of a small set of primary emotions is that they are irreducible units. According to this assumption, emotions like “fear” or “anger” cannot be broken down into further components that are still emotional. The psychologist Lisa Feldman Barrett has argued strongly against this assumption, pointing out that it requires belief in some kind of mysterious “essences” of emotions—the belief that there is something irreducible that makes each primary emotion the emotion that it is (Feldman Barrett 2017a). This central assumption underlies the representation of each of the primary emotions in *Inside Out* as a distinct *character*. 

For general queries, contact webmaster@press.princeton.edu
“Joy” and “fear” do not merge with each other; they are each unique individuals. They have stable, fixed identities and functions, and do not share components (for example, in the movie’s metaphorical language, they do not share internal organs, limbs, and such).

Yet there is scant scientific evidence that “joy,” “fear,” or “anger” are irreducible and do not share component parts. Equally plausible is an alternative view in which each of these emotions is made up of a collection of components, or building blocks, some of which are shared by other emotions. Initial doubts such as these lead to the following set of scientific questions that can serve as a starting point for further investigation:

“Are different emotion states composed of features or dimensions that are shared, to variable extents, across multiple emotions? Are some emotions composed of, or based on, combinations of other more basic emotions?”

Questionable assumption 2: the primary emotions correspond to those for which we have names in English.

Related to questionable assumption 1 is the idea that words like “happiness,” “fear,” “anger,” and so forth in fact pick out scientifically principled categories of emotion. It is easy to see why this is unlikely to be the case. For one, we had these words for emotions long before there was any science of emotion—so why would one expect them to align well with scientific emotion categories? For another, different cultures have different words for emotions, and many of these turn out to be extraordinarily difficult to translate. In German, the word “Schadenfreude” denotes the emotion we feel when we feel happy about somebody else’s misfortune. Should that be a primary emotion, just because there’s a common word for it in German? There are many more such examples, entertainingly cataloged in Tiffany Watt Smith’s book, *The Book of Human Emotions* (Smith 2016). This poses some important scientific questions:

“How should we taxonomize emotions? How many emotions are there, and what names should we give to them? Are there different emotions in different cultures? Are there different emotions in different species? Can we use a word like ‘fear’ to refer to the same...
type of emotion state in a person, a dog, and a cat? How and when in evolution did emotions first arise, and how did they diversify?"

Given how little we yet know about these questions, and given that there are good reasons to believe our current emotion categories ("happiness," "sadness," and such) will need to be revised, we will say little in this book about specific emotions. We will refer to some emotions (notably "fear") by way of example. And we will sketch how a future science of emotion might give us better categories or dimensions by which to taxonomize emotions. But this book is primarily about emotions in general, not about specific emotion categories.

Idea 2. Emotions are rigidly triggered by specific external stimuli. In the film Inside Out, all five emotion characters sit lazily around the control panel watching a screen that projects the outside world into the little girl’s mind, and are aroused into action only when an appropriate stimulus or circumstance appears. In the film, some stimuli do not activate a given emotion at all (for example, the “anger” character often sits dozing in his chair and does not react unless something maddening happens to the girl), while other stimuli activate multiple emotions. If the depiction from the film were accurate, we could easily figure out the emotion states of other people (and presumably other animals) by a straightforward list of rules that link specific stimuli to specific emotions in a characteristic and inflexible manner. This picture assumes that emotions are far simpler and more automatic than we in fact now know them to be. According to Idea 2, emotions would be just like reflexes. Some things will make you happy, others will make you sad, and some will trigger a specific mix of emotions, according to a set of rules.

Questionable assumption 3. Emotions are like reflexes. The movie gets it right that emotions are often triggered by stimuli in our surroundings. But what determines which emotions are triggered by which stimuli and under which circumstances? Why would seeing a dog trigger only a minimal emotional response in some people, and strong fear or happiness (emotional responses of opposite valence)
in others? What accounts for the extraordinary flexibility with which many different stimuli, depending on the context and depending on the person, can elicit emotions? One can pose the following scientific questions:

“What determines whether an external stimulus will evoke an emotion or not, and what determines the kind of emotion evoked? What role do development and learning play in determining an organism’s response to a given stimulus? How does this process differ from simpler stimulus-response mappings, such as a reflex?”

Idea 3. Emotions control our behavior. The film portrays the emotion characters as controlling the little girl’s behavior by operating joysticks on the control panel. The little girl is but a hapless puppet, with emotions determining her behavior. This central visual metaphor encapsulates the title of the movie: our behavior is controlled, from the “inside out,” by our emotions. This feature is the counterpart to 2 above, with respect to the behavioral output rather than the stimulus input.

Questionable assumption 4. Specific emotions cause fixed and specific behaviors. Our subjective experience of emotion leads to the intuition that our emotions cause our behavior: I cry because I feel sad. Yet not all emotion theorists agree with this assumption. Indeed, the nineteenth-century American psychologist William James argued, counterintuitively, that emotions are a consequence, not a cause, of behavior: I feel afraid because I run from the bear, I do not run because I feel afraid (James 1884). Yet James already had doubts that just observing bodily reactions was sufficient to identify specific emotion categories. If it were true that specific emotions cause fixed and specific behaviors, we could infallibly deduce a person’s emotions just from watching their behavior. If so, then taken together with questionable assumption 3, we wouldn’t need emotions at all to explain behavior, there would simply be a set of rules linking stimuli to behavior. That was the view that behaviorism advocated in the earlier twentieth century. One reason for the demise of behaviorism was that people realized that mappings from stimuli to behavior were far too complicated, and too dependent on
context, inference, and learning, to be formulated as rules. Emotions, in our view, are internal states that afford a flexible mapping to behavior, as we will detail throughout this book. This leads to the following scientific questions:

"Do internal emotion states cause behavior, or are they merely an accompaniment to behavior? Or might emotions actually be a consequence of behavior? What exactly are the causal links between stimuli, emotions, and behavior? How could we identify emotions in the absence of behavior? After all, we can be angry without punching somebody or showing any other easily detectable behavior."

Idea 4. Different emotions are located in different, discrete brain regions. The beguiling picture of emotions as walking, talking cartoon characters in *Inside Out* is closely aligned with the belief that different emotions must correspond to anatomically distinct modules in the brain. Is there a place in the brain for fear, for example? This is a question that has received a lot of attention, including serious scientific investigation!

Questionable assumption 5. Specific emotions occur in specific brain structures. The era of functional neuroimaging with fMRI, as well as the study of patients with focal brain lesions, has led to the idea that emotions are generated in localized brain structures. For example, findings on the amygdala (a brain region studied in both of our laboratories to which we will return in some detail in later chapters) have led to the popular view that "fear is in the amygdala." Yet more recent work clearly shows that this view cannot be right; indeed, that it does not even make sense, and that emotions depend on a much more distributed set of brain regions. This leads to the following scientific questions:

"How is the processing of emotion carried out across the brain? Are there identifiable functional neural substrates that organize or implement specific emotion states? Or is any given emotion state produced in such a highly distributed manner that it is impossible to assign a function in emotion to any brain region or neuronal cell population? Would it ever be possible to predict what emotion an
individual is experiencing purely by examining activity in his/her brain?"

As we will explain later, modern neuroscience approaches have given us a view of brain function that reconciles a dichotomy inherent in these questions. It will turn out that there are no macroscopic brain structures dedicated specifically to emotions (fear is not “in the amygdala”), but that there is specificity nonetheless. The specificity is at the level of circuits and cell populations, a level of organization that requires modern neuroscience tools to visualize. We spend some time in chapters 4 and 5 explaining these neuroscience tools, since their logic is required to reformulate the questions about emotion.

Idea 5. Emotions are conscious homunculi. The movie illustrates beautifully the idea that the brain is a machine with a little person (or persons) inside, who views the outside world, reacts to it, and then transfers those reactions to us. In other words, our subjective experience of emotion is created and embodied by the subjective experience of a miniature version of ourselves in our brain, a so-called homunculus (box 1.1). (As an aside, it is also interesting that this view, of little emotion homunculi within ourselves, to some extent relieves us of full responsibility for our emotional behavior—as when we say, “my anger made me do it.”)

Questionable assumption 6. Emotions are purely subjective experiences. How the brain creates an internal representation of the external world, and translates that representation into thoughts, feelings, and action, is a central open question in neuroscience. We know for sure that there is no little person sitting inside the brain looking at a screen and pulling on joysticks. The only things that have access to the patterns of neuronal activity in the brain are other neurons in the brain. How neurons “decode” the information represented by other groups of neurons and pass that information on to yet further groups of neurons so as to organize and express thoughts, emotions, and actions, is a deep mystery that we are far from solving. This leads to the following scientific questions:
“How exactly do emotions arise in the brain? Can we separate the subjective, conscious experience of emotions from the existence of emotion states per se? Do emotions always have to be conscious? If so, how should we study them in animals, who may or may not be conscious and, in any case, cannot tell us how they feel?”

As we elaborate in the next section, we believe it is critical to distinguish between emotions as internal functional states, and conscious experiences of emotions (often called “feelings”). Emotions and feelings are not the same thing, although they are of course closely related. Most of this book is about emotions, not about feelings. We review some of the work on feelings near the end of this book.

**Box 1.1. The fallacy of the homunculus.**

A homunculus, literally “little person,” refers to the idea that inside your brain there is a separate observer, something that can watch and interpret the activity of all the other brain regions in the same way that an external scientist might be able to record from your brain and make sense of its processing. The idea of a homunculus has a long history in psychology and the philosophy of mind. It fundamentally arises from a confusion between different levels of description. On the one hand, we know that humans and animals have emotions (and many other mental states). On the other hand, we know that these mental states are produced by the brain. It is therefore tempting to conclude that emotions must literally be found in the brain if we only look with sufficiently microscopic tools.

But producing emotions is not the same as having an emotion. By analogy, there are many places in the brain that participate in producing vision, from the retina to the thalamus, to the cortex. But you cannot find vision in any one of these regions, nor does any of them have the experience of seeing. Or to take one more example: you can drive a car. So who or what does the driving? You can no more be driving by yourself (without a car) than a car can drive by itself (unless perhaps it’s a self-driving car). And you
can’t take apart the car to look for where the “driving” really is located. Driving, vision, and emotion are system properties: they are not properties of any of the constituent parts, but all the parts work together to generate the property.

The most common aspect of emotion where a homunculus fallacy often arises is with respect to the conscious experience of emotion (or, for that matter, the conscious experience of anything else). Unlike the little characters that *Inside Out* put inside the mind of a girl, there are no homunculi in the brain for experiencing your emotions. There are brain systems that make you have a conscious experience of emotion. But the conscious experience of the emotion is a global property of a person (or animal), and the mechanisms whereby it is produced do not themselves have that property.

**Toward a Science of Emotion**

Without further reflection, it might seem that it should be straightforward to investigate emotions, and to discover how emotions work in the brain. But the assumptions and questions sketched in the first part of this chapter show us that a science of emotion faces some difficult challenges looming ahead. A science of emotion needs to examine most of our initial intuitions about emotions, sharpen vague questions so that they can be experimentally investigated, and confront both empirical and conceptual problems.

Let’s take a closer look at one of the major sources of conceptual confusion in emotion science. There is an assumption that different words, concepts, or types of data must refer to distinct things. We will argue instead that one and the same thing can be described with very different words and measured with very different types of data. Consider the thought-provoking image on the next page (figure 1.1), produced by neuroscientist Rebecca Saxe at MIT and published in *Smithsonian Magazine* (December 2015). Saxe got a mother and her infant child to go into an MRI scanner and obtained these images showing their brains. Saxe writes:
While they lie there, the scanner builds up a picture of what’s inside their skulls. Often MR images are made for physicians, to find a tumor or a blocked blood vessel. Scientists also make the images, to study brain function and development. In my lab, at MIT, we use MRI to watch blood flow through the brains of children; we read them stories and observe how their brain activity changes in reaction to the plot. By doing so, we’re investigating how children think about other people’s thoughts.

To some people, this image was a disturbing reminder of the fragility of human beings. Others were drawn to the way that the

**FIGURE 1.1.** MRI scan of a mother kissing her child. Reproduced with permission, copyright Rebecca Saxe.
two figures, with their clothes and hair and faces invisible, became universal, and could be any human mother and child, at any time or place in history. Still others were simply captivated by how the baby’s brain is different from his mother’s; it’s smaller, smoother and darker—literally, because there’s less white matter. Here is a depiction of one of the hardest problems in neuroscience: How will changes in that specific little organ accomplish the unfolding of a whole human mind?

As for me, I saw a very old image made new. The Mother and Child is a powerful symbol of love and innocence, beauty and fertility. Although these maternal values, and the women who embody them, may be venerated, they are usually viewed in opposition to other values: inquiry and intellect, progress and power. But I am a neuroscientist, and I worked to create this image; and I am also the mother in it, curled up inside the tube with my infant son. (http://www.smithsonianmag.com/science-nature/why—captured-MRI-mother-child-180957207/)

As you were reading the above quote, you probably felt a tension between the colder, internal glimpse of two physical bodies shown in the MRI scan and your realization that these are two real people engaged in an affectionate emotional behavior. The MRI scan shows only tissue contrast, revealing bones, fluid, muscle, and brain. At the same time, we know that this is a mother and child—they are people, with thoughts and emotions. Both our everyday view of people and the view made possible with the MRI are of the same thing.

This is perhaps the most critical realization for a science of emotion (indeed, for a science of the mind in general). You can feel emotions. You can infer that other people are having emotions from their behavior. And you can image and record traces of emotions in the brain. These are very different types of data, very different sources of evidence about an emotion. And indeed, they need to be kept separate if one is studying them in their own right—as we shall see, feeling an emotion, having an emotion state, and attributing emotions to another person engage distinct processes in your brain. Nonetheless, your experience of your own emotion, your attribution of an emotion to another person
you might see laughing or crying, and the neuroscientist’s investigation of an emotion from neurobiological data are not about three different things. They are ultimately all about one and the same thing, an emotion state. You can infer the emotion state in another person from observing their behavior, you can investigate the neural mechanisms of the emotion state through neuroscience experiments, and the emotion state may cause you yourself to have a conscious experience of the emotion. The behavioral observation, neurobiological measurement, and personal experience each can provide evidence for one and the same thing: an emotion state.

To flesh this out a little further, let’s view emotion from four different perspectives: the perspective of the behavioral biologist who might be carefully watching the behavior of an animal in the wild or the laboratory (or, for that matter, watching the behavior of a human being); the psychologist concerned with having people talk about and rate their conscious emotional experiences; the psychologist measuring emotional responses in the body, such as changes in heart rate or facial expression (common approaches in the psychology of emotion); and the neurobiologist who is studying (or even manipulating) the function of neurons in the brain (figure 1.2). All four perspectives can be perfectly objective and have an established and agreed-upon methodology—but they are rather different data and often do not use the same language to describe the concepts and methods that relate their data to emotions. Yet all four investigate emotion.

Of those four perspectives, it is especially neuroscience that can show you things you could never get from your everyday knowledge of emotions. What kind of drug will work best for curing depression? Why do some people fear dogs whereas others love them? And, first and foremost, what are the underlying mechanisms that generate emotions—how do neurobiological events in the brain cause tears to run down our face when we are in a state of sadness, and how does this emotion state change much of the rest of our behavior, our attention, our memory, our decision-making? These and many other questions like them are important for treating psychiatric illnesses, for understanding everyday human cognition and behavior, and for understanding the cognition and behavior of other animals. You cannot get at them by just thinking...
about your feelings. The aim of a neuroscience of emotion should be to make transparent how and why specific emotions have the features that they do: to explain them through their underlying mechanisms (this is a topic we discuss in detail in chapter 4).

But although this book will focus on neuroscience, our hope is that our broad and functionally based approach will contribute to an integrated science of emotion, a science that investigates emotions through behavior, psychology, and neurobiology. Such a science of emotion should also aim to investigate emotions across species, from worms and insects, to mollusks and fish, to birds and reptiles, to mice and dogs, to monkeys and to people. It would identify specific instances of emotions

For general queries, contact webmaster@press.princeton.edu
with particular states of the brain; that is, how they are in fact instantiated in people and animals. But it would also explain why this brain activity instantiates a particular type of emotion in people and animals, and how one might imagine that such an emotion could be instantiated otherwise. Could a nonbiological robot (of the right sort) in principle have emotions? Answering these questions requires a framework that allows us to operationalize emotions in all these different instances. We describe such a framework in chapter 2.

We will speak frequently of “emotion states,” or, for shorthand, simply “emotions.” But of course, emotions are anything but static states. We refer to them as “states” to keep things simple, but with the realization that they are complex processes that vary in time. We will take apart emotion states and talk about processes in later chapters. But the fact is that we know very little yet about this level of description—so we will keep referring to “emotion states,” again, with the full acknowledgment that emotions are in fact temporally extended processes.

**Emotions Are Decoupled Reflexes**

To see where emotions fit into the picture of complex behavior, let’s start by considering a simpler behavior. The knee-jerk reflex, for example, which is a behavior produced when the doctor taps your relaxed knee tendon to cause your leg to move. We call that behavior simple for two reasons. First, it is simple because your knee doesn’t jerk to pretty much anything except that specific stimulus. It doesn’t move if the neurologist shows you a picture of a little hammer. It doesn’t move if you just think of a hammer. It also doesn’t move if you tap your shoulder. Second, the converse is also true—the tapping of the knee doesn’t do much other than move the knee. Your arms don’t move, nor do other parts of your body. So the link from stimulus to behavior is very narrow, encapsulated, and modular. In fact, we now know that you can get rid of all of the brain, but as long as the spinal cord is still intact, you will still have the knee-jerk reflex.

Like other reflexes, the knee-jerk is a rigid, narrow, and automatic stimulus-response mapping. You don’t feel like you decided to move your knee, you had no urge to move it, you’re not thinking or planning
to move your knee, and in fact it is hard not to move your knee. Such responses have a purpose, but they face a major problem: the world is so complex and changeable that it would take a continuous supplementation of more and more reflexes to deal with all possible kinds of stimuli and adaptive behaviors. You would need yet a different reflex to respond to the ground under your foot, a cup in your hand, food on your lips, and so forth. You would also need some mechanism by which these reflexes could communicate with one another, so that your overall behavior becomes coordinated. Something more flexible is needed for an animal to survive in the world. Enter emotions.

As we will describe in chapter 3, emotions may have evolved out of reflexes, but they show properties that go well beyond what reflexes can accomplish. We may think of the most basic set of these properties as “emotion building blocks,” which share some minimal features that distinguish emotions from reflexes (we give a list of these in chapter 3). These features of emotions can be used to infer emotions in many animal species. For instance, flies already show behaviors that meet some of the criteria for emotion building blocks, such as persistence through time (Gibson et al. 2015). There is an evolutionary story to tell about how emotions became more elaborated and eventually took on the set of properties that characterize emotions as we see them in humans. That story needs to be a functional account of what it is that emotions do, what function they accomplish in the life of an animal.

An example of such a story can be seen with the emotion of disgust. Disgust evolved so that animals could avoid poisonous or contaminated foods. That might seem like a problem for reflexes to solve: just link the taste of the poison to the behavior of spitting out the food and you are done. Well, it’s more complicated than what reflexes can solve. There are many different poisons and contaminants. Some taste bad, some don’t. Others might look, feel, or smell contaminated. There are also poisons and contaminants that animals learn about by watching what happens to other individuals (or, in humans, having other people tell you what might happen if you ate them). This kind of learning can be seen in young children for instance— their food preferences change considerably as they mature. Little children often put things in their mouths that we adults find disgusting, but later learn to avoid such foods.
Similarly, even in adults, there can be profound learning of disgust. For example, we may get violently ill from food poisoning and thereafter no longer enjoy that food. Finally, in some situations we may have to overcome our feelings of disgust and ingest something we would otherwise avoid in order to survive: think of a lost explorer who resorts to drinking his own urine in order to avoid death by dehydration (a strategy that only works for a short while, in case you were curious). So there are clearly both innate and learned aspects to disgust, and the disgust state exerts a strong effect on our behavior but can, to a degree, also be overcome. Reflexes just don’t have the flexibility to work across these varied cases; they are too narrowly tuned to specific stimuli, and too rigid and uncontrollable in how they cause behaviors.

An alternative is to evolve another layer of control on top of reflexes: to decouple them from their rigid stimulus-response arcs, and instead to incorporate features of them into a different kind of architecture, a central emotion state. We believe that emotions like disgust evolved as central states whose very function is to orchestrate and flexibly regulate such complicated stimulus-response mappings. The emotion state can still be thought of as a functional module similar to a reflex, but, unlike reflexes, that module is portable across a huge range of different situations, many of which we learn about. The scientific study of disgust nowadays views this emotion as having evolved from the need to avoid contamination with pathogens or toxins, which evolved into more complex and social forms of disgust (Rozin 1996) (figure 1.3). But although portable across very different domains, the emotion state of disgust remains concerned with a specific function: the avoidance of passive stimuli that are potentially harmful (the avoidance of actively harmful stimuli might map onto a different emotion, closer to “fear”).

The evolutionary invention of emotions as flexible central states resulted in additional benefits. Because emotions need to interface with so many different kinds of stimuli, and stimulus modalities, and because they have to interface with many aspects of motor control, they are also poised to interact with many other central processes. Thus, emotions influence attention, memory, and other cognitive processes—a rapidly growing field in human cognitive neuroscience today (Pessoa 2013). Indeed, it is essential for emotions to interact
**FIGURE 1.3.** Functional varieties of disgust. The scheme on the right conceives of disgust as a state evolved for specific functions related to avoidance of potential poisons or contaminants. The different boxes as we go from the top to the bottom of the figure illustrate that disgust has been elaborated from ancestral functions to avoid poison, to more social ones. There is continuing debate about how these functions evolved; this scheme assumes a biological evolution, but Paul Rozin and Jonathan Haidt have argued that cultural evolution may explain some of the functions of disgust. Reproduced with permission from Tybur et al. 2013.
with all of these other processes in order for them to carry out their functional role. For instance, your disgust for a particular food needs to be encoded in memory and needs to redirect your attention. If all you had to work with was the emotion state, without any connection to other cognitive processes, it would bring us back to the encapsulation of reflexes, and their pitfalls.

There are many other properties of emotions that distinguish them from reflexes, a topic we consider in more detail in chapter 3. Of course, reflexes and emotions do not exhaust all of the mechanisms that explain complex behavior. You also deliberate and plan volitional actions, for example. Most of the time, your behavior is produced by a rich and coordinated mixture of many reflexes, emotions, and planned actions. While there is no definitive dividing line, we should thus situate emotions at an intermediate level of complexity in behavioral control, between reflexes and deliberative behavior.

Emotions “decouple” stimuli from responses, as compared to reflexes (Scherer 1994), but they are still fairly automatic compared to deliberative, planned actions. They occupy a particular level of control in the behavior of higher animals: above reflexes, but below planned action. While this helps to situate how emotions fit into the architecture of central states that regulate behavior, it does not yet distinguish emotions from many other central states. For instance, states like motivation, drive, and arousal also fit with much of what we have been saying about emotion. These other kinds of states are considered further in chapter 5. Further delineating emotions requires a more fine-grained look at their properties: what kinds of operating characteristics do we find for prototypical emotion states, and how might this help us to understand the underlying mechanisms, and the functions that they implement? These are topics we take up in the next two chapters.

We have so far sketched only the roughest outline of a science of emotion and what it brings to the table, making three important points. First, emotions need to be studied behaviorally, psychologically, and neurobiologically. A science of emotion needs to be interdisciplinary in the types of methods and data that it considers. Second, emotions function in the flexible control of behavior and are generated by mechanisms within the brain. This generates a list of features of emotions...
that distinguishes them from reflexes and from several other types of internal states, and that we can test for experimentally (the topic of chapter 3). And third, emotions do not only influence behavior, they also influence many other cognitive processes like attention and memory (and other emotions). Implicit in these three points is that many animals have emotions, not only humans, and, in our view, not only mammals. Everything we have written so far about emotions is applicable to animals and may be investigated in them just as in humans. A science of emotion should thus be interdisciplinary not only in its methods but also in the species it studies.

Questions We Will Not Answer in This Book

In addition to understanding the fundamental principles that are common to all emotion states, there are more detailed questions that a science of emotion will eventually want to address. We will not give answers to these questions in this book, but we will discuss which of them might be answered by empirical investigation, and which might just be semantic issues that are in fact not necessary to answer. Some of these questions return to those motivated by our list of questionable assumptions earlier in this chapter.

1. How many emotions are there? Is the neural representation of a given emotion discrete and separate from that of other emotions, or does it share features or components with other emotions? If the latter, which features of neural representation are specific to a particular emotion, and which are shared? Are some emotions more basic (“primary”) than others? We will suggest that this question can be answered by situating emotions in a dimensional space and listing their functional properties. We also believe that many of the words we currently have for different emotions are not going to find a place in a mature science of emotion, which will need to revise the emotion categories that we talk about in everyday life.

2. What is the relationship between the brain representation of an emotion state and that of other related states such as arousal,
motivation, drive, mood, or affect? Are they related (nested or overlapping), or are they different? We will suggest that empirical data indeed show differences between these states, although this question is also partly semantic.

3. What determines the type of emotion that will be evoked by a given stimulus, or whether that stimulus evokes any emotion at all? Are our emotional responses at all “hardwired” by genetics, or are they mainly determined by experience? We will suggest that this question will require further experiments that compare emotions across people and animals, so that we can better understand the full range of emotion states and how they arose in evolution.

4. Are emotion states causative to behavior, or (as William James believed) are they consequences of (reactions to) behavior? Or is it only our subjective experience of emotion states that is consequential to behavior? We will argue for the view that emotions cause behavior, not the other way around—but we will also acknowledge that once an emotion has caused a behavior, that behavior can itself trigger further emotion states. So the answer to the original question is: both.

5. How is the “feeling” aspect of an emotion state encoded by the brain, and how does that differ from the “motive” aspect of the state (the part that controls our behavior)? We will argue that feelings are quite distinct from emotions and need to be kept distinct for a science of emotion to make progress. This is a strong view we will return to in the next chapters in some detail to make our argument. However, it is essentially a premise for the rest of the book rather than an answer to a question. We believe that if you accept this premise, there will be benefits for a science of emotion since it becomes more tractable.

6. How will we know when we have fully understood brain mechanisms of emotion? One test would be to be able to predict, with a high degree of precision, how experimentally perturbing the function of any arbitrarily selected brain region or group of neurons would affect any aspect of a given emotion state. We give examples of how this can already be done (in animals) in chapters 5–7. Another test would be our ability to engineer a machine that can
express an emotion state in a manner indistinguishable (by more than superficial visual criteria), from that in a human—that is, to build a robot that has emotions. We think this is possible in principle, even though it is currently science fiction.

**What Do We Want to Know about Emotions?**

This chapter gave more questions than answers, and discussed examples of what we don’t yet know. So what is it that we do want to know? We close with a summary of what one might ideally wish for in a science of emotion, and where we focus in this book. Figure 1.4 provides a list of desiderata—things that one would eventually like to have some answer to, even if the answer is that the original question was ill-posed. We do not address all of these desiderata in depth, although we comment on all of them at some point throughout the book.

The first broad desideratum lists some of the most basic phenomena that a science of emotion should explain to the layperson. As you can see from the entries in the figure, this makes a critically important point about the scope of this book. We aim to provide a framework for investigating and explaining emotional behaviors, in both humans and animals. But we bracket the topic of conscious experience, often called feelings. The reasons for this bracketing are entirely practical. We certainly think feelings are important, and we believe that both humans and animals have them. We leave them out in this book because they introduce difficulties—there is no agreement on how to measure feelings, especially in animals, and there is even less agreement on how to investigate them neuroscientifically. A similar bracketing of conscious experience has been taken for many other topics in neuroscience (for example, vision or memory), and yet these sciences have been highly successful. We are wagering that a science of emotion can also get a successful start without needing to tackle conscious experience of emotion at the outset.

The second broad desideratum lists some properties of every good science. Any framework for a science of emotion should provide consistent and objective criteria that produce results that make sense. Let’s unpack this important claim. Providing consistent and objective criteria means
that we agree on how to determine whether there’s an emotion or not, and that we do so consistently across cases. Thus, we should have some agreement on what kinds of measures or outcomes would rule in favor of a particular animal being in an emotion state or not. Moreover, we should apply similar criteria across animals and humans.

For general queries, contact webmaster@press.princeton.edu
We also need these criteria to produce results that make sense. Making sense means providing some correspondence with our prescientific intuitions. Of course, there can be disagreements, and of course there are problematic cases. But this doesn’t detract from the fact that there are clear cases on which most people agree—if this were not the case, we couldn’t be talking about emotion. Most people agree that humans, monkeys, dogs, and cats have emotions. Most people agree that cars and clouds and laptop computers do not have emotions. So whatever account a science of emotion ends up with, it should respect these starting points—if it does not, something probably went wrong with the science. This second broad desideratum, then, asks of a science of emotion that it gives us some objective way to interpret data, such that we could decide whether an organism is in an emotion state or not.

The other entries under “scientific features” provide related properties that help to translate the objective approach into a useful science in practice. If a science of emotion were so encapsulated that it remained within psychology, or even within philosophy, that would be a pity. If psychology and neurobiology used different terms, or terms with different meanings, that would be a bad sign. So we want to aim for a science of emotion that is integrative in the sense of connecting across disciplines. It should provide something meaningful and useful to the philosopher, psychologist, and neurobiologist alike.

The third and fourth broad entries, experimental approach and generalizability, sketch further features that are more specific to our particular framework. One could certainly imagine a science of emotion that does not exhibit all, or even any, of these features—but such a science would not look modern, would have a narrow domain of application, and would fail to capitalize on the most exciting current opportunities in techniques, analyses, and objects of study. This would also be a pity. We would like a science of emotion to use cutting-edge methods (for instance, optogenetics, the ability to experimentally cause an emotion in an animal by using light to manipulate neurons in its brain; see chapter 5). We would like to have computational models and mathematical equations that can allow us to better explain and predict emotions. And we would like to be able to study emotions not only across human subjects but also in apes and monkeys and dogs and cats and rats and even...
flies and octopuses. In short, the entries under these third and fourth broad desiderata aim to make a science of emotion a modern, vibrant, and frankly exciting and fun science.

Finally, the last entry in our list of desiderata is about applications. These are extensions: what you could eventually do with a science of emotion, once you have it figured out. This book only applies to the very first subentry: we can begin to use a science of emotion to discover facts about emotion. Which animals have emotions? Which brain mechanisms produce emotions? How can we measure emotions? Once we have accrued more data from such a science of emotion, we will be able to apply it to assessment (for instance, diagnosing whether a patient is depressed or not), intervention (treating the depression), and even construction (building robots that have emotions). We think all three are possible to do, but none of them are easy to do yet, and so we do not treat them within the scope of this book.