Book review

In search of emotions

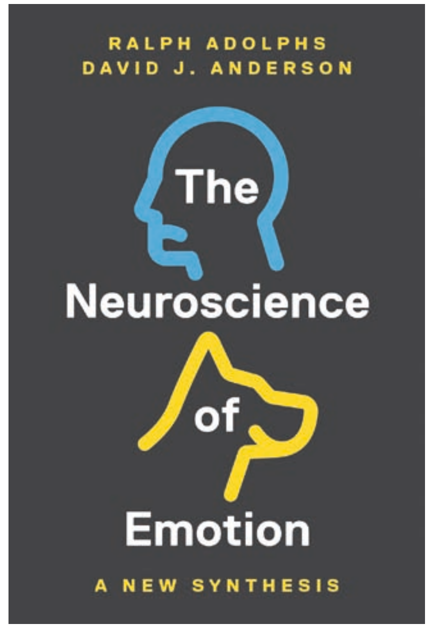
Lisa Feldman Barrett

The Neuroscience of Emotion: A New Synthesis
Ralph Adolphs and David J. Anderson
(Princeton University Press, Princeton, NJ; 2018)
ISBN: 978-0-691-17408-2

Scientists agree that emotions exist but can't seem to agree on what they are and how they work. *The Neuroscience of Emotion: A New Synthesis* by neuroscientist Ralph Adolphs and neurobiologist David Anderson attempts to bring order to this scientific morass. The authors elegantly articulate a particular point of view in the centuries-long debate over the nature of emotion. Written in a clear, straightforward style, *The Neuroscience of Emotion* is primarily a book of questions in which Adolphs and Anderson describe their vision for a mature science of emotion and offer a roadmap for getting to the answers.

To understand the authors' particular neurobiological perspective, it's important to realize that a word like 'fear' names a *category* of events — a group of instances that are similar to one another in some way. Broadly speaking, any given episode of fear may include numerous physical features, such as initial conditions in the body and world, visceral changes in the body, motor movements such as freezing or attacking, sensory information from the various movements, and the brain state that controls the whole process and represents any resulting sensations and associated subjective experiences. The experiences, also known as psychological features, might include conscious feelings of pleasantness or unpleasantness and activation or quiescence (together called 'affect'), a conscious experience of the surrounding world (such as whether the situation feels safe or threatening, novel or familiar, and so on: often called 'appraisals'), an awareness of oneself as being in an emotional state (called an 'experience of emotion'), and an awareness of what one's sensory and motor changes are 'for' (referred to as a 'function' or 'goal'). All of these features extend over some temporal window and occur in a particular context. The 'great emotion debate' boils down to the 'what' and 'how' of similarity. Which features count as part of an episode of fear and which are epiphenomenal? Which are similar across other fear episodes? And are emotions — as biological categories — natural kind categories that are waiting to be discovered by scientists, akin to Darwin's descriptions in *The Expression of the Emotions in Man and Animals*? Or are they conceptual categories, similar to Darwin's discussion of animal categories in *On the Origin of Species*, where similarity is in the eye (well, brain) of the beholder? If instances within the same emotion category are highly variable in their physical features, then the similarities holding the category together are psychological, created in the brain of some perceiving agent who is making sense of these physical fluctuations (i.e., the similarities are perceiver dependent). These distinctions are what philosophers of science call ontological commitments: assumptions about *what* exists, which features are relevant, where the boundaries of phenomena lie, and so on. Ontological commitments are not discovered by scientists: they are *stipulated* by scientists and set boundary conditions on what is possible to discover with scientific inquiry.

One important strength of *The Neuroscience of Emotion* is that Adolphs and Anderson describe their own ontological commitments: for them, emotion categories are natural kind categories with necessary and sufficient features. Based on these commitments, the authors sketch a neuroscience approach that resembles a proposal that was offered almost a century ago by physiologist Walter Cannon and his student Philip Bard: that emotions are states within an animal's central nervous system. These states are thought to be triggered by events in the world and cause survival-related actions associated with defense, foraging, reproduction, thermoregulation, and so on, as well as the related visceromotor changes



within the animal's body that support these actions. Cannon wrote that emotion states are located in the neurons that provide central control of the body. Adolphs and Anderson make pretty much an identical assumption. (In fact, if you substitute the phrase 'bodily control' every time the neurobiology of 'emotion' is discussed, most sentences would continue to retain their meaning.) Our authors remain reasonably agnostic as to the neurons in question, whereas Cannon argued that the neurons were located in the thalamus and hypothalamus.

Like Cannon, Adolphs and Anderson separate conscious, emotional experience from the definition of emotion. Their goal is to craft what they hope will be a more objective scientific approach, free from human phenomenology, to render emotions capable of being studied in species that may not have subjective experiences or in which experience cannot be studied. Here, the authors identify the fundamental ontological commitment in their approach: any worthwhile science must begin with the *assumption* that emotions exist across the animal kingdom rather than the *question* of what is species general and what is species specific. In their view, a solid science of emotion will "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". Given this assumption, one cannot *discover* whether emotions exist in non-human animals or whether emotions emerge from a combination of species-general and species-specific mechanisms. This approach, the authors conjecture, will allow scientists to improve drug discovery for humans by using model systems in non-human animals, understanding how emotions arose in evolution, and building robots with emotions.

If you share the authors' ontological commitments, then you will likely find the book tightly argued and principled in its approach. However, what seems obvious to some scientists camouflages deeper questions for others, so they continue to disagree about these ontological commitments. As a consequence, *The Neuroscience of Emotion* cannot resolve the great emotion debate. Rather, Adolphs and Anderson clarify the affirmative position for a science that begins with the assumption of emotion categories as natural kinds. It is the clarity of their arguments, rather than the details of the content, that may help to advance the debate and perhaps contribute to its resolution.

For example, the book helpfully explains a catch-22 in the natural kinds approach. If scientists are to study emotions as species-general states, then "emotions should not be defined, fundamentally, in terms of neurobiology … [but] at a more abstract, functional level". That is, we must focus scientific inquiry on the conditions that trigger the state and the motor consequences of the state. (What does the state *do*, and what sort of evolutionary problem does it solve?) Functional approaches to emotion emerged in the early 20th century when scientists were attempting and failing to find the physical basis of emotion categories using the scientific methods of neurology and physiology. But the function of any emotion category is rooted in *human inference* — a psychological description for the causes of behavior — rather than in physical mechanisms. As the authors explain, "if emotions were defined literally as brain states, we would have to provide a different definition for humans, for flies, for octopuses — because these species have completely different kinds of nervous systems". Hence the catch-22: to have an objective, cross-species science of emotion, free from the foibles of human experience, one must define emotions in a way that depends on human phenomenology, i.e., the experience of the humans who are doing the scientific observing.

What a scientist can learn (or fail to learn) about emotion in any experiment is determined by that scientist's own definition of emotions. You can't shock a fruit fly, measure the degree to which it avoids something associated with the shock, and discover something about the neural circuit for fear in the fly's brain unless you stipulate at the outset that a state of fear is elicited by shock and causes withdrawal. Generalizing empirical findings about fear across species, even when the observables differ, rests on the *assumption* that the underlying fear state remains the same, and this assumption is an inference on the part of the scientist. According to Adolphs and Anderson, the stimulus evoking fear can differ in flies and worms and fish and birds and dogs and monkeys and people. The resulting fear behaviors can also differ. And the intervening brain states, in physical terms, also differ. Heck, the stimulus, brain state, and response can differ in two fearful humans in the same situation or in the same person who is fearful on different occasions. By necessity then, any induction about fear upon which scientific generalization depends is rooted in the scientist's mental inference about the abstract, functional meaning of a central fear state.

Many notable scientists including Darwin and Einstein have recognized that concepts and beliefs are as much part of a scientist's toolbox as the technologies and methods that they employ in experiments. 'Space' and 'time' are conceptual tools that physicists use to describe and study what we refer to as 'reality' (meaning the stuff we experience either directly or indirectly with our measurement tools). When physicists conduct an experiment, they measure variations in energy that they describe as electrons or quarks or a Higgs boson. Electrons, quarks and the Higgs boson do not exist in physical reality in a way that is separate from the human scientists who study them. *Something* exists that is nicely described by our concepts for electrons, quarks and the Higgs boson when physicists measure things in the ways that they typically do. Conceptual tools are not inert — they are necessary for the proper workings of science. Whether or not Pluto is categorized as a planet or a 'planetoid' determines whether what is known about Pluto becomes part of the accumulated body of scientific knowledge about planets that is then inductively generalized to form hypotheses about other planets.

So what's the problem when it comes to the science of emotion? For a start, scientists often use common-sense concepts to guide their definition of a functional state, so a mature science of emotion based on natural kind assumptions is at risk of being a sort of fancy folk psychology approach to understanding emotions that won't bring us any closer to improving drug discovery or building emotive robots. More importantly, research shows that the inferred functions for fear (or anger or any other emotion category named in English) vary by context and person. Folk concepts of emotion also differ substantially across cultures. These observations imply that a functional approach may be ill-suited to build a universal science of emotion that applies to all humans, let alone to all creatures on the planet.

Whether or not you agree with the ontological commitments that are offered by Adolphs and Anderson, *The Neuroscience of Emotion* is definitely worth reading. It's the best articulation that I've seen of this point of view in the science of emotion. Just realize that what you take away from the book depends on the assumptions that you bring to it. For me, the book was a thought-provoking journey. (My copy is filled with marginalia.) It is also a reminder that, as scientists, we always view our subject matter through the somewhat foggy lenses of our own, very human experiences, whether we realize it or not. And so, we are never quite as objective as we hope.

Department of Psychology, Northeastern University, 360 Huntington Avenue, Boston, MA 02115, USA. Department of Psychiatry and the Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA 02129, USA.
E-mail: L.Barrett@Northeastern.edu