CellPress

My Word

What is An Emotion?

Ralph Adolphs[1], Leonard Mlodinow[2], and Lisa Feldman Barrett[3,4,5]

In 1884, William James, the American psychologist, famously posed the question: what is an emotion? After more than a century of scientific inquiry, however, emotions remain essentially contested concepts: scientists disagree on how they should be defined, on where to draw the boundaries for what counts as an emotion and what does not, on whether conscious experiences are central or epiphenomenal, and so on. Such disputes have sown great discord among scientists, leaving the field in perpetual upheaval, and without a unified framework for guiding scientific inquiry and accumulating knowledge.

What follows is a dialogue between two neuroscientists who study emotions in humans: Ralph Adolphs (The Neuroscience of Emotion: A New Synthesis) *and Lisa Feldman Barrett* (How Emotions Are Made: The Secret Life of the Brain)*. Adolphs and Barrett agree that commonsense notions about emotion do not provide a solid ground on which to build a mature science of emotion. But they take very different scientific approaches, disagreeing on the most fundamental assumptions of what emotions are and how they work. We hope that this dialogue, moderated and edited by the theoretical physicist and noted science writer Leonard Mlodinow* (Subliminal: How Your Unconscious Mind Rules Your Behavior; Elastic: Flexible Thinking in a Constantly Changing World)*, clarifies some of the key points of debate that may help with an eventual solution.*

LM: It is a reflection of the current state of emotion science that you are both leading researchers in the field, and yet your differences are apparent at the most fundamental level of discourse. So, let's start at the beginning. What is an emotion?

RA: In my view, emotions are the functional states of the brain that provide causal explanations of certain complex behaviors — like evading a predator or attacking prey [1,2]. I think that there is a single functionally-described state for instances of the same emotion category — fear, anger and so on. People in these different functional states will process somatic and environmental information differently.

A functional state is defined by what it does, not by how it is actually constituted in the brain, or by the feelings it may be associated with. As a result, functional states provide objective, public, criteria that make it possible to objectively infer them. In contrast, if emotions are classified according to how you feel, this is not a public criterion (indeed, it's not clear there is any reliable criterion, as Wittgenstein famously argued [3]).

LFB: So objective means that a bunch of people observe inputs and outputs and publicly agree on the function they infer as a consequence of those observations? You are defining objectivity by consensus?

RA: If by 'people', you mean scientists, then for present purposes, yes.

LFB: But consensus is not scientific validity — it is reliability.

A perceiver's brain makes sense data meaningful by categorizing it, and in doing so imposes a socially agreed-upon emotional function. In a given moment, my brain might categorize my sense data as an instance of sadness and your brain might categorize the sense data coming from me — my movements, my vocal acoustics, in the same context — as an instance of anger. We can compare our inferences to one another, or what is normative in that particular situation, but there are no objective criteria to say who is right. Research indicates that changes in heart rate or blood pressure have physical functions for maintaining the body and planning movements, but they have no inherent emotional meaning. The same changes in peripheral physiology are observed across a wide variety of emotional states [4]. Unfortunately, science, so far, has not identified truly objective criteria (perceiver-independent criteria) for when animals are in a particular emotional state and when they are not, despite over a century of trying.

RA: If this were true, we could never communicate about emotions. We use functional definitions all the time. For example, here's a simple functional definition: the function of the heart is to pump blood. That's why we call internal blood pumps artificial hearts; they are still hearts because they have the same function. What's inferred or not objective about this?

LFB: Sure, a heart pumps blood. But a pulmonologist might say the function of the heart is to bring oxygen to the rest of the body's tissues. An endocrinologist might say that the function of the heart is to bring glucose and hormones to the body's tissues. What is the function of a coffee-maker? To make coffee? To be a stand that props up your cookbooks?

RA: Of course, functions are not unique. But neither are they arbitrary. I can describe any physical object in many different ways, but this non-uniqueness in how we can individuate objects and events doesn't detract from the objectivity of the description. If I induce fear in a rat, the functions of that fear include defense from threat, and so on. They are evolutionary in origin so there is a fact of the matter about which functions, out of all possible ones I could conjure up, are relevant. That makes the functional description objective. Describing the coffee maker as the object whose function is to hold up books is of course a joke: that is not the function it was designed for, and not a function anybody else would agree upon.

LFB: I agree that functions are not arbitrary or unconstrained. But that does not mean functions are evolutionary in origin. To assume so is teleology, which we know can reflect erroneous causal reasoning (for example [5]). The important point here is that the evolved function that you assign to an emotion category is the result of human inference.

LM: Lisa, rather than starting with emotion categories and seeking to define them through their functions, your approach starts with the

architecture and functions of brains and asks how such brains make emotions. Tell us about your theory.

LFB: All brains are faced with an inverse inference problem: ambiguous, noisy sense data continually arrives from inside the animal's body (the result of allostasis) and from the surrounding environment (the animal's niche). The brain does not have access to the causes of the sense data so it must *infer* them. So, a brain constructs inferences — hypotheses about the causes of sensations — by remembering past events that are similar to present conditions. This is the basic hypothesis of predictive coding approaches to understanding brain function [6]. So here is my hypothesis: a brain solves its inference problem by continually constructing *ad hoc* concepts to make sense of the cacophony arriving from its sensory organs [7,8]. By 'concept' I mean a representation of a category, a group of events or objects that are similar in some way.

LM: Where does emotion enter the picture?

LFB: Emotions, in my view, are those embodied representations that shape the animal's action and become the animal's experience of the world in that upcoming moment. In humans, the representations may (or may not) include an explicit awareness of the emotion state, which we colloquially refer to as the 'emotional experience'. The hypothesis about ad hoc concepts is not specific to emotion, by the way; it holds for every action you take and every experience you have, whether it is classified as an emotion or not, including thirst, hunger, pain, boredom, sexual desire, and so on.

RA: But this doesn't tell me what makes a particular state an emotion rather than something else.

LFB: I hypothesize that the same processes are at work when the brain creates an instance of emotion and when it creates non-emotion states. When a brain constructs an *ad hoc* category by reimplimenting prior instances where it escaped from a predator, for example, then it is constructing a situation-specific concept for evading a predator, which we call 'fear'. That concept is a set of hypotheses, or a plan, that includes action preparation. It anticipates the needs of the body and attempts to meet those needs before they arise (a plan for allostasis). And it includes predictions of the sensory consequences that will result from those movements. In effect, the *ad hoc* fear concept is a functional state of fear that is preparing the animal to act and experience the world in a specific way.

LM: Where do the prior instances of fear come from?

LFB: An infant brain learns them. As an infant brain develops, it is learning the sensory-motor patterns that become available for later predictions. In humans, the patterns include words used by parents and caregivers. Words invite infants and young children to form abstract categories — this is how children learn that instances of a category are functionally similar even when their physical features are different. If a parent or caregiver tells a child that she is fearful in some contexts when she is crying, when she is clingy, when she hides, and when she hits, then she will learn that these instances are *functionally* similar even though they may be physically different.

LM: Are there unconscious and conscious components of emotion? Do scientists who study emotions have to be careful to distinguish which form they are considering?

RA: My view is completely silent about conscious experience. We can come up with a functional story about the actual emotion (preventing harm and so on), but the functional role of their conscious experience is unknown. Emotions have a function, whereas our conscious experience of them has no clear function.

By my reading, Lisa's theory is a theory only about the *conscious experience* of emotion (aka 'feelings'), not a theory of emotion *per se*.

LFB: This is a common misperception. My scientific approach contains hypotheses about how a brain constructs momentary states which may or may not include emotional experiences — that is, an awareness of being in a particular emotional state and being capable of labeling it and communicating about it — but it is not correct to say that this is the only thing that my approach targets for explanation.

Any instance of emotion includes numerous features, including experiences of pleasure or displeasure, comfort or discomfort, arousal or sleepiness. In my work, we refer to these features of experience as *affect*, and they arise from the brain's continual allostatic control of the body, so they are not specific to instances of emotion. Instances of emotion also includes numerous unconscious physical features, such as changes within the internal milieu of the body, as well as the unconscious inferences that result in a conscious experience of the surrounding world during the emotional state (sights, smells, tastes, and so on); this latter experience is not an emotional experience.

LM: I imagine that your differences regarding conscious and unconscious human emotion are reflected in your views on animal emotion. Do animals have emotions?

RA: Yes. In fact, I think the evidence argues for homologous emotion circuits that are shared by humans and many other mammals.

LFB: I disagree—there may be homologous circuits for actions, but for these to become circuits for emotions, you have to stipulate a mental inference that observed actions (for example, freezing) are caused by certain function or goals (for example, protecting against threat in fear). Only then can scientists claim that the circuitry for actions is evidence for emotion circuits. What we learn (or fail to learn) about emotion in any experiment is determined by how we define emotions in the first place.

RA: So, Lisa, do animals consciously experience their surroundings? Are there affective features of that?

LFB: As an animal brain maintains allostasis, it automatically predicts the sensory consequences of the

predicted movements (exteroceptions) and of the internal changes in the body (interoceptions). So I hypothesize that animals consciously experience their surroundings (they see, hear, taste, and so on). In vertebrates and perhaps some invertebrates, the associated interoceptions may be experienced as affective features of those percepts.

But that is different from suggesting that non-human animals have emotional experiences — that they are aware of being threatened, fearful, or have other emotional experiences, as Jaak Panksepp was hypothesizing. For an animal to experience emotions like a human, the animal's brain must be able to make *ad hoc* emotion concepts like a neurotypical human does.

LM: Lisa, Ralph believes that emotions are definite functional states of the brain, while your theory emphasizes that everybody's experience of emotion may be different. Don't emotions have more-or-less well-defined functions?

LFB: I agree with Ralph that an instance of an emotion category, such as anger, is a functional state. But the available data lead me to hypothesize that the function of anger will vary situationally, depending on the context. For example, studies show that human anger can accomplish purposes as diverse as removing an obstacle that is blocking an important goal, helping to enhance performance to win a competition or a negotiation, allowing a person to appear powerful or signal dominance, or protecting against a hostile threat [7]. Ad hoc concepts for anger or any category can vary in their function across situations (e.g., a bird can function as a pet or as food).

RA: On what basis would you then call all of these 'anger' according to your view?

LFB: Whenever a brain reimplements past instances of anger to construct an *ad hoc* concept of anger for the purposes of guiding action, anger is constructed. Any instance of anger has motor features, sensory features, affective features, functional features, and so on; in a given situation, only some of these features are similar to past instances of anger, and the set of similar features will vary on a situation-by-situation basis. So, two instances become anger by virtue of human inferences of similarity.

One difference between our perspectives, I think, is that I propose human inference is necessary to construct instances of emotion, whereas your view is that human inference identifies instances of emotion but plays no role in constructing them.

RA: That's right. Your approach is very problematic because it lacks public, objective criteria. I think we should individuate emotions by the stimuli and circumstances that cause them, together with the behaviors and other psychological states they influence. We can reliably assign emotion states to animals (and to people) using criteria that are based on a theory about what emotions do, on their causal relations.

LFB: Your approach lacks objective criteria too. Emotions are individuated by how sensations are categorized and actions are planned. Scientists who "reliably assign emotion states to animals (and to people) using functional criteria" are simply promoting their inferences to the status of 'objective'. What if I look at a freezing animal and infer she is confused and vigilant, and you look at the same animal and infer she is fearful? Who is right?

RA: Stated this way, absolutely everything involves inference. Astronomers also infer stars from what their telescopes (and visual systems tell them). Do you agree? So you think stars are not objective?

LFB: Something exists that is described well by the concept 'a star' given our current measurement abilities. But many scientists have made the point that we cannot separate the observer from the observed. As Einstein said, "Physical concepts are free creations of the human mind, and are not, however it may seem, uniquely determined by the external world".

RA: So, you would say that emotions are like stars and galaxies with respect to how objective they are. If this is your standard of objectivity, I'm fine with that. I also think emotions are as objective as stars and galaxies.

LFB: I disagree. Scientists may objectively measure when an animal attacks a conspecific, but we cannot objectively claim that this was fear, anger, or some non-emotional state like being territorial. We can objectively measure when an animal is not moving in a novel cage, but we cannot objectively claim that the animal is fearful vs. metabolically frugal.

LM: Despite your differences, I think one thing you both agree on is that current taxonomies for emotion categories (happy, sad, angry, and so on) have a lot of problems. Do you think that any future taxonomy has exactly the same problems as the current one (that is, it is made up by people on the basis of nonscientific, non-objective criteria)?

LFB: Yes. All existing taxonomies that are anchored in folk psychology categories will have the same problems and therefore are not a useful guide for scientific investigation.

RA: What about one that is not folk psychological, like my functionalist view of emotions?

LFB: I think your functionalist view runs into the same problems as the folk psychology taxonomies.

RA: How about your own theory then? You use emotion words. What makes your use of the terms like 'fear' valid?

LFB: I have never claimed that my use of the word 'fear' is 'right' in any objective sense. My approach doesn't require it. My hypothesis is that emotions are categories of social reality [9]. Categories of social reality are rooted in *human consensus* —shells, rocks, or little pieces of paper become real as money when a group of people agree to impose a certain function on physical objects (i.e., people agree to trade them for material goods), thereby acting on them or with them in certain ways. Similarly, physical changes in the body, facial movements, actions, etc. have no intrinsic *emotional* meaning. You and I are imposing an emotional function on behavior when we perceive

a freezing rat as fearful. The inability of scientists to discover objective functional criteria for emotional states is not a bug — it is a feature of what emotions are and how they work.

LM: Do you think animal research might supply important evidence regarding which of your viewpoints is closer to the mark?

RA: Yes, in my view the clearest starting point for understanding emotions will come from the ethological study of simpler animals, not from human psychology.

LFB: I agree that some clear information will come from the ethological study of non-human animals. Unfortunately, current ethological studies narrowly focus on the neurobiology of physical actions and the physiology that supports them. We must also study how an animal's brain builds and modifies ad hoc concepts to categorize and impose functions in its ecological niche.

As for the use of 'simpler' animals, they will allow important but incomplete insights into the nature of emotion in humans. Emotional episodes in humans emerge from the interplay of species-general and species-specific computational ingredients. For example, the degree of abstraction in the similarities that characterize a species' *ad hoc* concepts is related to certain aspects of brain architecture which varies by species.

LM: Okay, let's consider a concrete example. A person spots a threatening figure nearby on a dark street. The person increases her walking speed and continues to walk quickly for several minutes after the encounter is over. A fruit fly encounters a threatening stimulus and increases its walking speed, also persisting for several minutes after the threat is removed. In each case, what is happening with regard to emotion?

RA: Both exhibit fear, although in somewhat different behavioral ways. The main difference is that the human's fear state incorporates a lot more information, and has much richer interactions with other psychological variables. Also, the human would of course verbally report that they feel afraid, whereas the fruit fly cannot.

LFB: Fear is not being exhibited. It is being inferred. The human brain and the fly brain constructed *ad hoc* concepts to plan and control fast walking. The human's concept is likely rooted in similarities that are more abstract. Her brain can reimplement past episodes whose similarities are functional rather than physical, such as other occasions of threat that are not on a dark street, or that she herself may only have read about. In contrast, the fly's brain may be constructing the concept from instances with similar physical features because, as far as I know, a fly's brain is not equipped to do the degree of abstraction accomplished by humans. Referring to these two events by the same name — fear — gives us false confidence that what we learn from studying fear in the fly will explain fear in the human. But fear in humans has both species-general and species-specific features. Ignoring this may be one reason why we've had difficulty discovering effective treatments for fear-related disorders.

LM: Fear is currently a very active area of research. What do you think is the most exciting recent finding in emotion research?

RA: For me, the most exciting findings are coming from specific circuits for threat in animals. There is very elegant work on how specific cell populations in the hypothalamus process aggression [10], or how specific circuits in the superior colliculus and periaqueductal gray process fear [11]. We know a lot about the stimuli (the smell of another animal, or an overhead looming shadow, respectively), about the behaviors (aggression, freezing), and even about the circuit computations. These systems are giving us by far the most detailed mechanistic account of how emotion processing happens in the brain. They are a very small and narrow part of the story in relation to all of emotion in humans but the hope is that they will provide us with fundamental principles that give us insight into how emotions are processed in general.

LFB: Agreed, although I would interpret those findings differently. Here are two recent papers I like. It was recently discovered that the same neurons in the basolateral amygdala that increased their firing when a rat began foraging for rewards or aborted foraging (in the face of explicit threat) also increased their firing in control tasks devoid of explicit threats or rewards [12]. These neurons were coding for movement velocity, suggesting that they are involved in regulating physical actions more generally. In addition, recently, scientists designed a study to allow rats to behave differently in response to the same stimulus, and observed evidence that contradicts the hypothesis that there is a one-to-one correspondence between conditioned appetitive or aversive responses and the activation of specific subsets of valence-coding basolateral amygdala neurons [13].

LM: How about the flip side of that question—what do you think is the biggest deficiency in emotion research today?

LFB: There are four big deficiencies.

First: within-category variability is under-appreciated. An emotion word, such as fear, refers to a collection of instances that are variable in their features. Scientists reason about fear, not about instances of fear, or they reason about predatory fear, rather than instances of predatory fear. They reason about the category rather than the highly variable instances of the category. But the categories, no matter how small, have variable instances, which must be studied and modeled.

Second: context is under-appreciated. Many scientists who study emotions in non-human animals focus on the circuitry for survival-related behaviors. The degree of flexibility and context-dependence in naturally occurring survival-related behaviors is vastly underestimated by the majority of current laboratory research. Typical laboratory settings have intentionally removed the variation that is inherently present in normal ecological contexts. The consequence of stripping away this multidimensional context is that survival-related behaviors will appear to be more immediate and stereotyped than those that are studied within their natural ecological contexts. These

artificially constrained, laboratory-evoked responses are then mapped on to neural circuits without opportunity to observe the graded, contingent and goal-directed nature of natural survival-related behaviors.

Third: the influence of scientists' own inferences is under-appreciated, leading them to confuse what is measured with what is inferred.

Fourth: the distinction between affect and emotion is underappreciated. Affective properties, like valence and arousal, are important features of emotional states, but they are not specific to those states. Affect derives from interoception, and so any brain state that involves interoception will necessarily have affective properties (at least in vertebrates).

RA: I agree with the first two. I would also add the lack of clarity in what people mean by 'emotion'. They use the word typically without explanation, and they often use it in very different senses. The problems are most apparent when trying to relate emotion research in animals (for example, studies of anxiety in rodents) with psychological studies in humans (for example, asking people how anxious they feel with a questionnaire). The dependent measures are vastly different, and so are the concepts of 'emotion' that the researchers are using. We need a common vocabulary.

LM: Any last words?

RA: I thank both of you for this discussion: Len for his help in bringing our viewpoints together and extensive editing, and Lisa for repeated and patient debates with me. I have the greatest respect for Lisa's view and scholarship, and do think she has identified a problem with my functionalist view that I don't know how to answer — how to pick the 'correct' function(s).

LFB: Len, I enthusiastically second Ralph's thanks for guiding us in this discussion. Ralph, my friend, it's always a pleasure. You are a generous colleague, a trusted critic, and I deeply admire your open-mindedness and willingness to consider a range of scientific views. I look forward to our next discussion!

LM: Thanks to you both. I think your insightful dialog illuminated many important issues in the field.

REFERENCES

1. Adolphs, R., and Anderson, D.J. (2018). The Neuroscience of Emotion: A New Synthesis. (Princeton, NJ: Princeton University Press).
2. Adolphs, R. (2019). Emotions are functional states that cause feelings and behavior. In The Nature of Emotion, Second Edition, A.S. Fox, R.C. Lapate, A.J. Shackman, and R.J. Davidson, eds. (New York: Oxford University Press), pp. 6–10.
3. Wittgenstein, L. (1953). Philosophical Investigations. (New York: Macmillan).
4. Siegel, E.H., Sands, M.K., Van den Noortgate, W., Condon, P., Chang, Y., Dy, J., Quigley, K.S., and Barrett, L.F. (2018). Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories. Psychol. Bull. *144*, 343–393.
5. Keleman, D., Rottman, J., and Seston, R. (2013). Professional physical scientists display tenacious teleological tendencies: Purpose based reasoning as a cognitive default. J. Exp. Psychol: Gen. *142*, 1074–1083.
6. Hutchinson, J.B., and Barrett, L.F. (2019). The power of predictions: An emerging paradigm for psychological research. Curr. Dir. Psychol. Sci. *28*, 280–291.
7. Barrett, L.F. (2017). How emotions are made: The secret life of the brain. (New York: Houghton-Mifflin-Harcourt).
8. Barrett, L.F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. Soc. Cog. Affect. Neuro. *12*, 1–23.
9. Barrett, L.F. (2012). Emotions are real. Emotion *12*, 413–429.
10. Lin, D., Boyle, M.P., Dollar, P., Lee, H., Lein, E.S., Person, P., and Anderson, D.J. (2011). Functional identification of an aggression locus in the mouse hypothalamus. Nature *470*, 221–226.
11. Evans, D.A., Stempel, A.V., Vale, R., Ruehle, S., Lefler, Y., and Branco, T. (2018). A synaptic threshold mechanism for computing escape decisions. Nature *558*, 590–594.
12. Amir, A., Lee, S-C., Headley, D.B., Herzallah, M.M., and Pare, D. (2015). Amygdala signaling during foraging in a hazardous environment. J. Neuro. *35*, 12994–13005.
13. Kyriazi, P., Headley, D.B., and Pare, D. (2018). Multi-dimensional coding by basolateral amygdala neurons. Neuron *99*, 1315–1328.

[1]California Institute of Technology, Division of Humanities and Social Sciences, Pasadena, CA 91125, USA. [2]Breckenridge Institute, 62 Emmett Lode Road, Breckenridge, CO 80424, USA. [3]Northeastern University, Department of Psychology, Boston, MA 02115, USA. [4]Massachusetts General Hospital and Harvard Medical School, Department of Psychiatry, Boston MA 02114, USA. [5]Athinoula A. Martinos Center for Biomedical Imaging, Charlestown, MA 02129, USA.
E-mail: radolphs@caltech.edu (R.A.); lmlodinow@gmail.com (L.M.); l.barrett@northeastern.edu (L.F.B.)