Functionalism Cannot Save the Classical View of Emotion
(long version)

Lisa Feldman Barrett

Department of Psychology, Northeastern University, Boston, MA.

Athinoula A. Martinos Center for Biomedical Imaging

Psychiatric Neuroimaging Division, Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Charlestown, MA

Corresponding author: l.barrett@neu.edu

"The past is never dead. It's not even the past." – William Faulkner, *Requiem for a Nun* (Faulkner, 1975, p. 80).

Since the time of the Ancients, people have believed that the human mind is structured as a set of mental abilities, each producing its own unique state. These *faculties* were thought to be innate and caused by an all-powerful force (first by the gods, then by a single god, and eventually by natural selection).[1] This approach, called *faculty psychology*, gave us the classical view of emotion. Ralph's engaging paper belongs to this of views. His theory is broadly integrative, sampling from many of the hypotheses that are found in a variety of classical views. But perhaps more importantly, his theoretical approach embodied the *assumptions* of the classical view, which are generally consistent with a cherished (Western) theory of human nature (Barrett, 2017). These assumptions allow the classical view to "ring true" – to appear obvious and therefore pleasing -- when compared to the seemingly complex and counter-intuitive constructionist approach to understanding emotion. In this commentary, I situate Ralph's ideas within the broad historical landscape of the classical view that has been shaped by attempts to defend those assumptions while at the same time remaining responsive to the accumulating empirical evidence that calls them into doubt.

A brief history of the classical view of emotion during the last century

The 19th century saw psychology transformed from mental philosophy into a scientific discipline. This was a golden age for the classical view of emotion. Neurologists and physiologists searched for the physical basis of mental categories, including emotion categories, attempting to discover their biological fingerprints and essences. Scientists formed grand theories as they battled, like the Ancients before them, over whether emotions were caused by biological changes in body or in brain. A compromise was struck—emotions were assigned to live in ancient parts of the brain that control the body, dubbed the "limbic system", a.k.a., our inner beast, whereas cognitions were assigned to the cortex like a crown designed for us by evolution, setting the stage for the "triune brain" concept. But the scientists of the classical view quickly ran into a problem. They were unable to identify any specific set of facial movements, autonomic nervous system changes, or neural circuits consistently associated with the instances of a single emotion category. Unable to find the physical basis of mental categories (including anger, sadness, fear, and so on), scientists recast emotions as *functional states*. These states were ill-defined (or defiantly not defined at all),[2] which was just fine, because during this era of *functionalism*, instances of emotion were studied solely by manipulating their *causes* and measuring their *effects*. Functionalism did not solve the problem, but instead produced a plethora of highly variable, contextual findings. Amid variation, functionalism quickly gave way to behaviorism, where the mind disappeared as a topic of scientific study, and psychology became the science of behavior. It's just a small step to go from deciding that mental events *cannot* be studied scientifically to assuming that they *should* not be studied. Correspondingly, emotions went from being mental states to states of the nervous system that caused specific behaviors.

Psychology eventually emerged from the darkness of behaviorism in the 1950s and 60s, rescued by cognition (arriving to the cognitive revolution in Plato's chariot).[3] The mind was reinstated as a topic of scientific inquiry, this time as a set of computations. It was faculty psychology all over again, with computers (rather than bodily organs) as the driving metaphor. The mind was once again carved into a set of abilities (e.g., memory, anger, fear, etc.). In doing so, states and processes amounted to the same thing: separate and qualitatively distinct processes (e.g., anger and fear) were assumed to produce correspondingly specific and qualitatively distinct states of the same name. The science of emotion experienced a renaissance as the classical view of emotion re-emerged using similar assumptions (e.g., an "affect program" for anger produced a state of anger (in Ekman and Tompkins theories); a "primary process emotion system" for fear produced a state of fear (in Panksepp's theory); an appraisal process for novelty produces a state of novelty (in Scherer's theory), and so on). Eventually, the cognitive revolution gave birth to the field of cognitive, affective and social neuroscience; hypothetical processes in the mind became blobs in the brain: tiny blobs (in intracranial measurements of neurons) or larger blobs (in brain imaging). Different neural populations of all sizes were thought to contain the circuits for different mental processes, each one responsible for creating a specific mental process in a one-to-one fashion, which passed information like a baton in a relay race (e.g., a perception computed in one part of the brain is supposed to be passed to cognitive or emotional processes implemented in other parts of the brain, which is supposedly passed, in turn, to the motor system for a response). Faculty psychology's search for mental circuits progressed from localized blobs to networks of connected blobs and then to distributed patterns of blobs, all the while maintaining the assumption that memory processes would eventually be localized to memory circuits, sensory processes would eventually be localized to sensory circuits, motor processes eventually would be localized to motor circuits, and emotion processes would eventually be localized to emotion circuits. So far so good.

But history was to repeat itself. Soon after experiments started accumulating, scientific reviews began to offer up the same bad news as a century earlier: studies that were purposely designed to isolate the specific physical basis of mental categories (including emotion categories) and distinguish them from one another were consistently unable to do so (for an overview, see (Barrett & Satpute, 2013). This time around, it was possible to add a broader array of neuroscience investigations of the human brain to the list of failed attempts. Functionalism, once again, is being pressed into service to rescue the classical view of emotion in the face of mounting disconfirming evidence, just as it was a century ago (for a list of disconfirming evidence, see Table 1 in Barrett, this issue).

As an approach to understanding emotions, or any set of mental categories, functionalism employs a specific set of philosophical assumptions. When scientists like Ralph and I debate the nature of emotion, we are deliberating over the validity of these assumptions. My goal here is to highlight them, and offer my point of view, in hopes of moving the dialogue forward. Without calling attention to them, they wreak havoc in scientific communication in stealth, even infiltrating points of apparent agreements (see Table 1).

Classical view assumptions

1. Functionalism is teleology. There are two popular flavors of teleology in the science of emotion. Input-output functionalism defines an emotion by its causes and effects. This is called *teleomatic functionalism*: a process, once initiated by a stimulus, persists towards a specific, unchanging end-point under a variety of circumstances. Teleomatic functionalism is a form of Platonic essentialism (i.e., implying that all instances of a given emotion category share a biobehavioral fingerprint, signature, or biomarker). As I described above, teleomatic functionalism was popular in the early part of the 20th century. Adaptational functionalism defines an emotion as a state that supposedly evolved to serve a particular utility or purpose (it is sometimes called the "intentional design stance", e.g., Keleman, Rottman, & Seston, 2013).[4] This is called *teleonomic* *functionalism.* The state is thought to be caused by an evolved program that is responsible for creating the evolved state. Teleonomic functionalism is a form of Lockean essentialism (i.e., implying that each emotion category has instances that share some kind of unchanging essence) as well as Aristotelian essentialism (i.e., a process has a very specific goal that is anticipated when the process is initiated). Ralph's ideas about emotion rely on teleonomic functionalism (and he is joined by other scientists of the classical view, including Cosmides and Tooby).[5]

Teleology makes good common sense, but scientists who study how people reason about the natural world, as well as some evolutionary biologists, believe that teleology reflects erroneous causal reasoning that interferes with a correct understanding of evolution and natural selection (e.g., (Gould & Lewontin, 1979; Keleman et al., 2013). I'll add a related concern: The evolved function of any biological category is a human *inference*, especially where emotions are concerned. Teleological approaches offer *a mental inference to explain the causes behavior*, focusing on the needs or goals of an animal (e.g., neurons in the amygdala contain the circuit for fear that evolved as a protection from threat and danger), instead of offering a physical explanation (e.g., neurons in the amygdala help to control the pattern generator that coordinates actions to produce freezing behavior that is sometimes, but not always, present during fear). The evolutionary biologist Ernst Mayr made a cogent argument for avoiding teleology and functionalism when studying the features that contribute to the adaptedness of an organism (e.g., Mayr, 2004, p. 48) because they encourage metaphorical language that cannot be verified in physical terms.[6] For example, causal ascriptions like "eyes widen in fear to increase vigilance and detect possible threats" and "a heart races in fear to facilitate escape from a predator when necessary" are *mental inferences* (my term) or *attributions* (Ralph's term) of psychological function (i.e., inferences about the adaptive role that the feature plays in the life of the animal). By contrast, inferences of physiological function, such as "eyes widen to expand peripheral vision" and "a heart pumps blood and increasing during running" are examples of *action identification* (Vallacher & Wegner, 1987). Action identification is the level of explanation that Mayr advocated when studying adaptive features; mental inferences are to be avoided.

Given the role of human inference in functionalism's causal accounts of behavior, we can understand functionalism, scientifically, as an act of meaning making. Luckily, decades of insightful research by social psychologists has examined how humans observe continual streams on ongoing physical movements and construct discrete, goal directed behaviors. Consider, for example, the hidden teleology embedded in the classical view's discussion of *behaviors* as caused by hypothetical emotion programs or states. Animals, including humans, do not emit punctate "behaviors." Decades of research clearly shows that humans, as perceivers, *see* behaviors (coordinated sequences of actions that have a beginning and an end) because we categorize a continuous stream of movements into an event and make them meaningful by inferring a goal (for a discussion and list of relevant references, see (Barrett, 2006b); see also the literature on event segmentation, e.g., Zacks et al., 2001).[7] When we don our lab coats as scientists, we continue this trend. We observe the movements of an animal (human or non-human) and then automatically, with no effort or sense of agency, infer that the animal is emitting discrete behaviors, in line with their own goals. There is nothing wrong with this, in principle. It is, after all, how normal perception works. The problems arise when as scientists, we fail to appreciate our own hand (i.e., our concepts) in what we observe. The distinction between *actions* vs *behaviors* is lost on many scientists (including neuroscientists and biologists) who seem not to appreciate the power of their own meaning making capacities and the role it plays in their own scientific activities.

In science, as in life, it is never a good idea to have too much confidence in your own perceptions to reveal reality. That is the path to *naïve realism*. Naïve realism mistakes experience for physical reality. In scientific inquiry, naïve realism is usually a path to illusion, our very own scientific yellow brick road to the Wizard of Oz. (For three examples, see Table 2). Still, naïve realism is hard to resist. Our brains become wired to notice (or create) some similarities and ignore others, creating boundaries between groups of instances (i.e., categories), and then believe that those boundaries exist in nature (Barrett, 2017). The results is essentialism, which I discuss in greater detail below.

2. Definitions of emotion are stipulated, not discovered. In philosophy, a stipulation is definition by fiat. The history of science reveals that emotions are treated as discoveries, but in fact they are prescriptions, which leads to problems in scientific reasoning. Scientists began with a framework for understanding emotion (using categories from mental philosophy), which then dictated the sorts of questions asked, experiments designed, and interpretations offered (and even what counted as data in the first place). For example, some theories of the classical view define emotions, as we have seen, as functional or goal-based adaptations. It is only this stipulation -- a mental inference linking observed actions (e.g., freezing) to certain function or goals (e.g., for fear) -- that allows scientists to claim that the circuitry for the actions is evidence for emotion circuits. You can't shock a rat, measure freezing and use the data to understand "fear" unless at the outset you stipulate that fear is elicited by a shock and causes freezing. What you learn (or fail to learn) about emotion in any experiment is determined by how you define emotions in the first place.

Stipulating that an emotion state is defined by the goals or functions that it serves does not solve this conundrum. Most emotion categories are associated with more than one goal or function, depending on the context. For example, anger has been associated with (but not limited to) overcoming an obstacle that someone blameworthy has put in your path, winning a competition or enhance performance in some way, protecting against a threat, dealing with an offense or with someone who acted unfairly, desiring aggression, appearing powerful or signaling dominance, lashing out in frustration, and even enhancing self-insight (see Barrett, 2017). So how do scientists know what the correct goal is for each emotion category? The answer is simple: the function of fear, or anger, or any emotion is, in a very fundamental way, part of its *stipulated* definition in a given context (meaning, it can be any one of a number of goals, depending on the interests and proclivities of the researcher in a given experimental context).

Can a function be inductively discovered with data? In principle, yes, if enough data are collected on enough people for enough time across enough contexts. And the possibility of applying big approaches to the study of emotion has the potential to provide us the inductive foundation we currently lack. Huge amounts of data will never give us an objective (perceiver-independent) measure of a goal, however, because goals are inferred by a human perceiver, and therefore the validity of our inferences, scientific or otherwise, rest on consensus (i.e., collective intentionality; see (Barrett, 2012).[8] And the implicit execution of this consensus, using a common storehouse of cultural knowledge, is the foundation of human civilization and a key feature of civil society. It is also the basis of much scientific research (particularly when it focuses on mental processes). But when this collective intentionality goes unrecognized, and is mistaken for carving nature at its joints, it is a bug in the scientific process. It is the reason that the stipulations of the classical view of emotion deeply resonate with many scientists and non-scientists alike. Anything else seems needlessly provocative and wrong-headed. Why deny the reality of what is right in front of your eyes? When scientists debate about the nature of emotion, we are usually wrestling with our a priori stipulations.

3. The classical view of emotion is non-falsifiable. We have encountered essentialism before in this discussion: essentialism is the belief that a category of instances of the same type (e.g., instances of fear) share a deep, immutable causal mechanism that makes them what they are. *Psychological essentialism* (Medin & Ortony, 1989) allows people to posit a hypothetical or unseen essence when the causal mechanism cannot be identified or in the absence of any evidence whatsoever of what the essence might be. For example, Ekman's hypothetical affect program, Panksepp's hypothetical FEAR system, and Ralph's "central emotion state" or "functional emotion state" are all examples of psychological essentialism. Psychological essentialism is a natural phenomenon. Research in developmental psychology shows that young infants learn to essentialize with words and human brains become wired to essentialize (more or less) which is evident by examining the anatomical structure of the cortical sheet (for a review, see Barrett, 2017). But in scientific inquiry, psychological essentialism is problematic, because it inoculates believers against disconfirming evidence, allowing them to continue to believe in the existence of emotion essences, despite accumulating evidence that disconfirms them (Barrett, 2017). And as we know, any collection of beliefs that cannot be disconfirmed is more of an ideology than a scientific theory.

The standard approach to solving the ever-widening chasm between the essentialist assumptions of the classical view and the scientific evidence is to (a) create more fine-grained typologies in an attempt to bring nature under control and make it easier to identify those illusive emotion essences, (b) invoke phenomena that are presumed to be epiphenomenal to emotion, such as display rules, emotion regulation, or error, or (c) claim that more data is needed, preferably from better experiments with more potent inductions, more powerful measurement techniques, and more sophisticated computational approaches.

When it comes to data, more is always better. But let's face it: scientists have been searching for the physical essences of emotions, in one form or another, for over a century, and the classical view remains, as always, in doubt. If emotion states are so crucial to our survival, they should not be that difficult to see when placed under the lens of scientific scrutiny.

Conclusion

In all areas of science: physics, chemistry, and biology, progress has involved a shift - a paradigm shift — away from essentialism and naïve realism towards a more dynamic, contextual and constructionist approach to the physical world. Neuroscience is achieving astounding discoveries – some of which transform our basic understanding of our own minds in a similar paradigm shift – but the classical view remains trapped in a prison of its own assumptions. These assumptions make the classical view seem eminently reasonable, but they do not make it useful or correct.

The scientists of the classical view believe they offer a theory of emotion that is free of concepts, when in reality they use received concepts so automatically and fluidly that they seem no longer truly aware of doing so. All science relies on human concepts, and contrary to the example Ralph offers in his paper, this is a true for the astronomy as it is for the science of emotion (e.g., celestial bodies are perceiver-independent, but planets are not; Pluto's recent demotion from "planet" to "dwarf planet" is a case in point). Science is not a body of facts that pop up, like little lightbulbs, to illuminate a golden path to universal truth. Science is a process of transforming numbers into meaningful information through the use of concepts. A scientist's concepts are her flashlight, determining what variation she observes as signal and what she ignores as error. So it is a mistake to assume that astronomy merely involves observing the sky through the lens of a telescope. There is also the invisible lens of a scientist's own concepts (or some other scientist's concepts), whether he realizes it or not.

Ralph's summary of the classical view (this issue) illustrates that as scientists, we are never quite as objective as we think we are. We always see our subject matter through the somewhat foggy lenses of our own experiences, whether we realize it or not. And so our scientific findings are never quite as value neutral as we hope they will be.

Emotions, as they appear to you, are not the fundamental reality of the brain's architecture. They are mental features that are a product of that reality. This is a hard-won realization. It requires giving up certain assumptions and embracing others, and above all, it requires learning a new set of concepts. Until then, the classical view of emotion will remain seductive, so obviously true and beyond doubt that no amount of disconfirming evidence will shake the foundations of your confidence that your experience reveals the truth (see Table 3).

Without concepts, people are experientially blind. Changes in air pressure do not become sounds; wavelengths of light do not become sights; chemicals do not become smells. If the classical view of emotion seems obvious to you, then a new set of concepts might be just the thing: they might allow you to discover something that has been right in front of you all along, but that you have thus far been unable to see.

---

[1] In modern terms, faculties are the *categories* that describe what the mind *is* (as a formal ontology of states) and how the mind is *caused* (i.e., functional architecture of processes)

[2] "I do not propose to attempt any description of the emotional qualities nor of the bodily expressions of "the emotions". If the reader does not know what it is to be afraid, or angry, or disgusted .... no amount of description, however, eloquent, will enlighten him." (McDougall, 1923, p. 328-9)

[3] Plato wrote that the human psyche consists of three parts: rational thoughts, passions (which today we would call emotions), and appetites like hunger and sex drive. Rational thought was in charge, controlling the passions and appetites, an arrangement that Plato described as a charioteer wrangling two winged horses.

[4] There is a version of teleology that does not take this stance, and that privileges social and relational contexts (see Waxman & Medin, 2013) but the classical view does employ this particular teleological approach.

[5] In both versions of functionalism, the triune brain, thoroughly disconfirmed as a theory of brain organization (e.g., Striedter, 2005), miraculously remains more or less intact: studies are continually motivated and interpreted according to the assumption that some type of emotion lives in the amygdala, down regulated by some sort of cognition that lives in the prefrontal cortex.

[6] Mayr defines a program as "coded or prearranged information that controls a process (or behavior) leading it towards a goal. The program contains not only the blueprint of the goal but also the instructions for how to use the information in the blueprint" (Mayr, 2007, p. 53). Notice that in Mayr's hands, the classical view of emotion would be focused on finding the biological blueprint (presumably genetic material, anatomical structure, and so on) and the instructions for using the information in the blueprint in physical terms (e.g., genomics, the dynamics of neural firing, neurochemistry, etc.). A program is not a metaphorical process or state defined as a latent construct (which is how scientists of the classical view use the term "program"). Nor is there anything in Mayr's definition that insists there must be one essential packet in the blueprint or one single mechanism in the blueprint (variation and degeneracy can be situated nicely in his description). Any feature can be adaptive even if the processes that cause it did not evolve for that specific purpose. By failing to distinguish processes and states, and using metaphorical language, teleology can lead scientists to confuse description and cause.

[7] Behaviors are perceived as discrete acts, but the brain encodes continuous physical movements (much like

colors are perceived categorically but the brain actually encodes a continuous spectrum of wavelengths) (for a discussion, see Barrett, 2006a). Research on person perception has produced a highly developed psychological model that is empirically well grounded and explains how continuous movements are categorized into discrete, meaningful behaviors (for a list of review papers, see Barrett, 2006a). Conceptual knowledge about people drives the categorization of movements into behaviors in much the same way as conceptual knowledge about color drives color perception. In this way, scientists see "behaviors" rather than simply movements. They automatically and effortlessly partition continuous movements into recognizable, meaningful, discrete acts, using concept knowledge.

[8] Some scientists implicitly use authority in place of objective validity criteria. This is the oft-heard "Darwin said so" defense. But when it comes to emotions, Darwin's writings were not free from teleology or essentialism (Barrett, 2017). Darwin's writings on emotion contains no privileged information. *The Expression of the Emotions in Man and Animals* is a book written by a very thoughtful scholar who was sometimes sloppy in his language, who sometimes ignores his own conceptual innovations, and who wrote a book on emotion that should be studied like any other book, complete with a thoughtful consideration of the motives of its author.

References

Anderson, D. J., & Adolphs, R. (2014). A framework for studying emotion across species. *Cell, 157*, 187-200.

Anderson, M. L. (2014). *After phrenology: Neural reuse and the interactive brain*: MIT Press.

Barrett, L. F. (2006a). Emotions as natural kinds? *Perspect Psychol Sci, 1*, 28-58.

Barrett, L. F. (2006b). Solving the emotion paradox: categorization and the experience of emotion. *Personality and social psychology review, 10*(1), 20-46. doi:10.1207/s15327957pspr1001_2

Barrett, L. F. (2012). Emotions are real. *Emotion, 12*(3), 413-429. doi:10.1037/a0027555

Barrett, L. F. (2015). Construction as an integrative framework for the science of emotion. In L. F. Barrett & J. A. Russell (Eds.), *The psychological construction of emotion.* (pp. 448-458). New York: Guilford.

Barrett, L. F. (2017). *How emotions are made: The secret life the brain.* New York, NY: Houghton-Mifflin-Harcourt.

Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annu Rev Psychol, 58*, 373-403. doi:10.1146/annurev.psych.58.110405.085709

Barrett, L. F., & Satpute, A. B. (2013). Large-scale brain networks in affective and social neuroscience: towards an integrative functional architecture of the brain. *Current Opinion in Neurobiology, 23*(3), 361-372. doi:10.1016/j.conb.2012.12.012

Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin, 130*(4), 553-573. doi:10.1037/0033-2909.130.4.553, 2004-15935-002 [pii]

Buckner, R. L., & Krienen, F. M. (2013). The evolution of distributed association networks in the human brain. *Trends in Cognitive Sciences, 17*(12), 648-665.

Clore, G. L., & Ortony, A. (2008). Appraisal theories: How cognition shapes affect into emotion. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 628-642). New York, NY: Guilford Press.

Faulkner, W. (1975). *Requiem for a nun*. New York: Random House.

Finlay, B. L., & Uchiyama, R. (2015). Developmental mechanisms channeling cortical evolution. *Trends in neurosciences, 38*(2), 69-76.

Gilbert, C. D., & Li, W. (2013). *Top-down influences on visual processing* (Vol. 14): Nature Publishing Group.

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London. Series B. Biological Sciences, 205*(1161), 581-598.

Gross, C. T., & Canteras, N. S. (2012). The many paths to fear. *Nature Reviews Neuroscience, 13*(9), 651-658.

Iwata, J., & LeDoux, J. E. (1988). Dissociation of associative and nonassociative concomitants of classical fear conditioning in the freely behaving rat. *Behavioral Neuroscience, 102*(1), 66-76.

Keleman, D., Rottman, J., & Seston, R. (2013). Professional physical scientists display tenacious teleological tendencies: purpose-based reasoning as a cognitive default. *Journal of experimental psychology: General, 142*, 1074-1083.

Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychol Bull, 139*(4), 917-940. doi:10.1037/a0030811

Lambie, J. A., & Marcel, A. J. (2002). Consciousness and the varieties of emotion experience: a theoretical framework. *Psychological Review, 109*(2), 219.

Le Pelley, M. E., Mitchell, C. J., Beesley, T., George, D. N., & Wills, A. J. (in press). Attention and associative learning in humans: An integrative review. *Psychological Bulletin*.

Li, S. S. Y., & McNally, G. P. (2014). The conditions that promote fear learning: prediction error and Pavlovian fear conditioning. *Neurobiology of Learning and Memory, 108*, 14-21.

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion, 23*(2), 209-237.

Mayr, E. (2004). *What makes biology unique?: considerations on the autonomy of a scientific discipline*: Cambridge University Press.

McDougall, W. (1923). *An outline of psychology*. London: Methuen.

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. *Similarity and analogical reasoning, 179*, 195.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning* (Vol. 47): Urbana: University of Illinois Press.

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philos Trans R Soc Lond B Biol Sci, 362*(1481), 773-786. doi:10.1098/rstb.2007.2087

Skerry, A. E., & Saxe, R. (2015). Neural representations of emotion are organized around abstract event features. *Current biology, 25*(15), 1945-1954.

Sterling, P. (2012). Allostasis: a model of predictive regulation. *Physiology & behavior, 106*(1), 5-15.

Striedter, G. (2005). Principles of Brain Evolution (Sinauer, Sunderland, MA).

Vallacher, R. R., & Wegner, D. M. (1987). What Do People Think Theyre Doing - Action Identification and Human-Behavior. *Psychol Rev, 94*(1), 3-15. doi:Doi 10.1037//0033-295x.94.1.3

Waxman, S. R., & Medin, D. L. (2013). Teleological reasoning about nature: intentional design or relational perspectives? *Trends in Cognitive Sciences, 17*(4), 166-171.

Wilson-Mendenhall, C. D., Barrett, L. F., & Barsalou, L. W. (2015). Variety in emotional life: within-category typicality of emotional experiences is associated with neural activity in large-scale brain networks. *Social Cognitive and Affective Neuroscience, 10*(1), 62-71. doi:10.1093/scan/nsu037

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., . . . Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nat Neurosci, 4*(6), 651-655.

Acknowledgements

Many thanks to Ajay Satpute for his helpful comments on an earlier draft of this commentary. This paper was prepared with support from the National Institutes on Aging (R01 AG030311), the National Cancer Institute (U01 CA193632), the National Science Foundation (1638234) and the US Army Research Institute for the Behavioral and Social Sciences (W911NF-15-1-0647 and W911NF- 16-1-0191). The views, opinions, and/or findings contained in this paper are those of the authors and shall not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documents

Table 1

Apparent Agreements

We agree……	But…..
Developmental and comparative data are crucial to solve the scientific mystery of emotions.	The classical and constructionist views of emotion understand the value of developmental and comparative data in very different ways. The classical view assumes that emotions are species-general whereas the theory of constructed emotion assumes that emotions emerge from the complex dynamics of species-general and species-specific processes.
Emotions do not emit behaviors in any kind of pre-conceptual way. Human perceivers make sense of the animals actions. Emotions are attributed (Ralph's word) or inferred (my word) by the scientist to explain and predict behavior, but these perceptions reveal the physical reality of emotions. Emotions are causal explanations for why a behavior occurred.	The classical and constructionist views agree that emotions are causal explanations for why a behavior occurred, but not in a mechanistic stimulus-response sort of way. We disagree on whether attributions (which are human experiences) are a magnifying glass that reveals what is really out there in the natural world, or whether human inferences result from meaning making activities, which themselves are part of the natural world, and that are ingredients that create emotions out of mere sensations and movements (just like meaning making creates vision from wavelengths of light, sounds from changing air pressure, and smells from chemical compounds, and so on). The classical view assumes that human attributions (i.e., the human experience when watching the animal) reveals what is going on in the animal's brain.
In scientific endeavors, the word "emotion" should be used exactly like the word "vision" or "memory."	Classical and constructionist views use very different understandings of how vision and memory work (for a discussion, see (Anderson, 2014; e.g., Gilbert & Li, 2013; Schacter & Addis, 2007).
Commonsense (folk psychology or faculty psychology) concepts that come from human experience do not provide a solid guide for scientific study.	I suspect Ralph and I disagree about who, of the two of us, is reifying experience and dipping their toes into the murky abyss of folk psychology.
Emotions can exist without any awareness of them.	The theory of constructed emotion utilizes the philosophical distinction between consciousness and awareness, whereas the classical view appears to conflate the two. People can be emotional without awareness, but

	not without consciousness. There are gradations of consciousness, of course, but if you are unconscious, you are asleep or in a coma. The experience of emotion that is not in awareness is called "world-focused" emotion (Lambie & Marcel, 2002).

Table 2

Examples of Naïve Realism in the Classical View of Emotion

A role for human experience in the definition of emotions.	The classical view claims to be a theory of emotions that is free from experience, when in fact the very definition of emotion and interpretation of data is heavily laden with very human perceptions (i.e., causal inferences of function). Those who argue against defining emotions as experiences are themselves using experience as the basis of their scientific categories, with little awareness of doing so.
Control vs. deliberate will.	The classical view uses the experience of control and effort to define controlled processing, thereby confusing concepts like control, will, and volition, when in fact, controlled processing is better defined computationally, and is orthogonal to feelings of control (Barrett, 2017; Barrett, Tugade, & Engle, 2004).
Experience vs. awareness.	Experience is not necessarily only what you can verbally declare. When a rat hears a tone or detects a shock, the animal is consciously experiencing those events. The rat is probably not aware of them in the sense that it cannot talk about them or reflect on them, but conscious experience requires neither of these things. Experience and awareness are related but distinct phenomena, and it is possible to have one (experience) without the other (awareness). Emotional episodes that include awareness (in a human) may be different in some important way when compared to instances of emotion that do not, but this is separate from the claim that experience as epiphenomenal to emotion (which comes very close to resurrecting behaviorism).
Using a concept vs. the ability to verbally describe a concept.	A concept is a collection of representations that are similar for some goal in some context. Constructing and using a concept to guide perception and action is distinct from the ability to verbally describe the concept and list its instances, or their properties, declaratively.

Table 3

Empirical Status of Ralph's Classical View of Emotion

Hypotheses	Comments
Scalability: arousal is the intensity of an emotion state. Changes in intensity recruit discrete behaviors.	Arousal and intensity are independent properties (Kuppens, Tuerlinckx, Russell, & Barrett, 2013). For example, a person can be intensely sad or intensely calm (both considered low arousal). Whether non-reflexive behaviors are discrete, or merely a set of movements that are perceived as discrete, is an empirical question (with an empirical answer; see pg. 3 of text).
Valence: this is a fundamental aspect of behavior (whether to approach or withdraw).	Valence is usually considered a property of consciousness, like brightness or loudness, and is independent from approach-withdrawal. For example, creatures can either approach or withdraw in anger (see work by Harmon-Jones) and in fear (e.g., escape vs. defensive treading). Valence is a mental feature of brain states that are not, strictly speaking, emotional, such as hunger, perceptions, attitudes, and even the meanings of words (for the latter point, see Osgood, Suci, & Tannenbaum, 1957).
Persistence: An emotion state can outlast it's eliciting stimulus.	Any brain state can persist. This is not unique to brain states corresponding to emotions. In fact, the technical definition of a "brain state" is dynamic over time: a brain state is the complete representation of the brain's physical features, as well as the environment that influences the brain's trajectory, including how the brain anticipates and responds to the "stimuli" it perceives in its environment.
Learning: The emotional significance of most stimuli is learned.	Learning (Pavlovian, statistical or otherwise) is domain general and not specific to emotion. A brain is always learning the perceptual regularities (in the case of perceptual categories) and functional regularities (in the case of conceptual categories), particularly when they have relevance for allostasis. Learning can be understood as the encoding and consolidation of prediction error (particularly "fear learning"; e.g., Li & McNally, 2014), particularly when it is of value to energy regulation (Le Pelley, Mitchell, Beesley, George, & Wills, in press).
Hierarchical behavioral control: emotions implement their effects on packages of behavior.	Actions are indeed controlled hierarchically, but not mechanistically. They are more likely controlled probabilistically by predictions, corrected by prediction errors, as I describe in *The theory of constructed emotion* (Barrett, this issue). A human can tremble in fear, jump in fear, freeze in fear, scream in fear, gasp in fear, hide in fear, attack in fear, and even laugh in the face of fear. But this in no way implies that there is one central state of fear that is controlling all these behaviors in an attempt to protect the self from harm. In fact, there are different neural circuits for different actions, all of which are perceived as fear (Gross & Canteras, 2012). When a human freezes as

	part of an instance of the category "fear," this in no way implies that the brain state is identical with a rat who freezes in fear. The hierarchically organized portions of the brain that control the pattern generator for freezing in rats and humans are different in important ways (Barrett, 2017; Buckner & Krienen, 2013; Finlay & Uchiyama, 2015).
Multi-component effects: A central emotion state coordinates physiology, behavior, and other outputs of emotion.	This hypothesis has been proposed and tested many times and was largely disconfirmed (e.g., Barrett, 2006b; Mauss & Robinson, 2009) before it was resurrected by (D. J. Anderson, & Adolphs, R., 2014). For a good example of a disconfirming study related to "fear learning", see (Iwata & LeDoux, 1988). At every moment, from birth till death, various systems in the body are coordinated at a given moment in time to produce efficient energy regulation (i.e., allostasis; Sterling, 2012); this is not specific to the events that we refer to as emotions. The classical view, including Ralph's theory, proposes something much more specific: a central state of anger is hypothesized to have a specific coordinated profile of responses (like a biobehavioral fingerprint). Thus far, this unsubstantiated hypothesis that has a lot of evidence to disconfirm it.
Priority over behavioral control: A central emotion state has "prepotent" control over behavior, interrupting whatever an animal is doing and terminating ongoing goal-directed behavior when an environmental challenge is encountered. They require additional regulatory mechanisms to override their behavioral expression.	It is an open question whether any strong sensory input, particularly if it is unexpected (i.e., prediction error), will interrupt whatever an animal is doing. This is not specific to the events that we refer to as emotions. Some kind of attentional control (whether it is experienced or not) is always required to shift behavioral sets; again, this is not special to emotion (see Barrett, 2017; Barrett et al., 2004 for a discussion of automaticity vs. controlled processing and the experience of control).
Similarity structure: Emotions relate to one another in a similarity space but fairly discrete clusters of instances correspond to specific emotion categories.	The hypothesis of "discrete clusters" has never been empirically tested because fairly stereotypic emotion instances are studied in the lab (e.g., see Wilson-Mendenhall, Barrett, & Barsalou, 2015)). This is a crucial point. Typically, scientists choose stimuli to induce "discrete" emotion states (designed to maximize categorical clustering) and then test for whether they find evidence of discreteness. To properly test the cluster hypothesis, however, it is necessary to study a broader sample of instances than those handful of stereotypes stipulated by Darwin. Otherwise you are finding what you expect to see. Even when scientists sample emotion stereotypes, however, they still observe much more variation within emotion categories than the classical view permits. The similarity relations among conceptual representations of emotion is related to the similarity of their neural representation (Skerry & Saxe, 2015). Ralph writes "the best match between neural and psychological similarity structures held for ratings on appraisal dimensions." This is intriguing because appraisal dimensions are best thought of as descriptions of how we experience the world when we are experiencing emotion (e.g., Barrett, Mesquita, Ochsner, & Gross, 2007; Clore & Ortony, 2008).