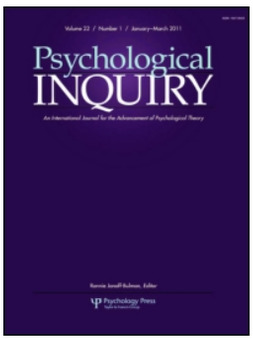

Routledge
Taylor & Francis Group

Psychological Inquiry

An International Journal for the Advancement of Psychological Theory

ISSN: 1047-840X (Print) 1532-7965 (Online) Journal homepage: http://www.tandfonline.com/loi/hpli20

Categories and Their Role in the Science of Emotion

Lisa Feldman Barrett

To cite this article: Lisa Feldman Barrett (2017) Categories and Their Role in the Science of Emotion, Psychological Inquiry, 28:1, 20-26, DOI: 10.1080/1047840X.2017.1261581

To link to this article: http://dx.doi.org/10.1080/1047840X.2017.1261581

Published online: 26 Feb 2017.

PSYCHOLOGICAL INQUIRY
2017, VOL. 28, NO. 1, 20–26
http://dx.doi.org/10.1080/1047840X.2017.1261581

Routledge
Taylor & Francis Group

COMMENTARIES

Categories and Their Role in the Science of Emotion

Lisa Feldman Barrett[a,b]

[a]Department of Psychology, Northeastern University, Boston, Massachusetts; [b]Department of Psychiatry and the Martinos Center for Biomedical Imaging, Massachusetts General Hospital/Harvard Medical School, Cambridge, Massachusetts

A perplexing situation persists in the science of emotion: There are a *multitude* of theories that vary a great deal from one another in almost every way imaginable: on the details of how an emotion should be defined; on where to draw the boundaries for what counts as an emotion and what does not; on which emotions matter; on how emotions are different from related concepts like mood, reward, and motivation; and on how various phenomena such as facial movements, physiological changes, and feelings should be treated. Theories even disagree on what observations and measurements counts as evidence for hypothesis testing. Scientists attempt to bring order to this dizzying cornucopia of theories in the same way that all living creatures deal with variation: We form categories.

Scientists love to sort things into groups or sets and then name them. Aristotle famously catalogued all sorts of stuff—from animals to governments—into strictly ordered "typologies" or "taxonomies." Carl Linnaeus created a taxonomy of plants, animals, and minerals that, to some extent, is still in use by biologists today. Physicist and novelist Alan Lightman eloquently described the lure of categorization:

> To name a thing, one needs to have gathered it, distilled and purified it, attempted to identify it with clarity and precision. One puts a box around the thing, and says what's in the box is the thing and what's not is not. … For scientists, it is a great comfort, a feeling of power, a sense of control, to be able to name things this way. (Lightman, 2005, p. 45–46)

In the science of emotion, we treat the multitude of theories exactly the same way we treat the phenomena themselves (i.e., the emotional events): We create taxonomies (of emotion theories, or of emotional events). And true to form, we cannot agree on which taxonomy is best. In the science of emotion, our colleagues largely do seem to agree on one thing, however: Scientific progress usually means cleaving larger categories into ever more precise groupings as an attempt to tame the huge amounts of variation and find signal in noise.

Categories are, of course, a necessary part of science. Category formation requires focusing on some similarities shared by a group of instances and ignoring their differences. This allows us to draw boundaries between what is the same and what is different. When it comes to organizing the multitude of emotion theories into a few broader groupings, a scientist must focus on some conceptual similarities at the expense of others. By grouping variable theories into a category while separating others into different categories, a scientist digests variation, inviting readers to ignore certain distinctions between the theories (within the same category) and to focus attention on other distinctions (different categories). This highlights an important observation: Most categories are not perceiver-independent, natural kinds. The similarities between members of the same category, and the differences across categories, are not absolute or fixed but are rooted in *human concerns*. A category of emotion theories (like most categories we deal with in science) is a grouping of theories that is *treated as similar for some purpose* (Murphy, 2002), with reference to a scientist's particular *goal* (Barsalou, 1983).

Herein lies the fundamental challenge for the science of emotion: Scientists sometimes fail to see how their own interests and goals influence which similarities they focus on, and which they ignore, both when it comes to categories of emotion theories and when categorizing emotional events, however labeled (as "emotions" or as "emotional episodes"). The human brain is so effective at creating similarities that it fails to recognize its own contributions to category formation. The result is naïve realism.

In the science of emotion, the conceptual pie is sliced and diced into many different taxonomies (just take a look at the Wikipedia page for "emotion" for some examples). One common taxonomy of emotion theories includes a category called "basic emotion theories," a category called "appraisal theories," and a category called "construction theories" (sometimes distinguishing psychological and social variants; e.g., Gross & Barrett, 2011). In several papers, I have suggested that this standard taxonomy constitutes one of the largest barriers to progress in the science of emotion because it both conceals meaningful variation within any single category of emotion theories and obscures important conceptual similarities across theories (for discussions, see Barrett, 2006a, 2015, 2017, in press-a; Barrett, Mesquita, Ochsner, & Gross, 2007; Gross & Barrett, 2011). Jim Russell has also written about similar concerns (e.g., Russell, 2015). In her recent paper, our colleague Agnes Moors (this issue) reveals herself to be a kindred spirit.

Moors's Categories

Agnes Moors (this issue) offers her own version of an emotion theory taxonomy. She wants us to understand that, from her perspective, there are two superordinate categories of emotion theories: "classic versus skeptical" theories. The category "classic theories" finds similarities between basic emotion theories and what Moors terms "discrete appraisal" theories (which she also calls "Flavor 1" appraisal theories; Moors, 2014), whereas "skeptical theories"

CONTACT Lisa Feldman Barrett l.barrett@neu.edu Department of Psychology, Northeastern University, 125 Nightingale Hall, Boston, MA 02115-5000, USA.

contain Russell's psychological construction theory (Russell, 2003) and "dimensional appraisal theories" (both her own and Scherer's, which she refers to as "Flavor 2" appraisal theories). These two superordinate theory categories are distinguished by whether the emotional phenomena in question form natural kind categories with necessary and sufficient features (even if probabilistic), as well as by the exact sort of causal stimulus-response sequences that are proposed to account for those emotional phenomena.

There is a lot to like about Agnes's approach. She offers a very systematic framework that attempts to bring order to the conceptual disorder of the field, sorting and naming, which can feel immensely gratifying, as Lightman observed. She distinguishes between theories of emotion, which assume that emotions are organized into categories with a classical organization (of necessary and sufficient features), and those that don't assume classical categories (for a discussion of classical vs. nonclassical categories, see Murphy, 2002; for this discussion applied to emotion concepts, see Barrett, 2017; Clore & Ortony, 1991; Russell, 1991). The formalizations in her framework require scientists to become aware of their own assumptions and goals. As a consequence, it becomes harder to believe that similarities are real in nature and that variation within each category is some of kind of error in a more universal sense. This, alone, makes her analysis useful.

Varieties of Appraisal Theory

Perhaps most important, Agnes aptly builds on others' work by fertilizing a well-worn path that others have paved before her. Various colleagues and I have observed that the "appraisal" category of emotion theories is heterogeneous (for discussion, see Barrett, 2015, 2017; Barrett et al., 2007; Clore & Ortony, 2000, 2008, 2013; Gross & Barrett, 2011; Ortony & Clore, 2015). Some appraisal theories stipulate that "appraisals" are literal cognitive mechanisms that produce subjective evaluations, which in turn either cause or constitute emotions (e.g., Arnold, 1960a, 1960b; Frijda, 1986; Lazarus, 1966; Roseman, 2011; Sander, Grandjean, & Scherer, 2005; Scherer, 2009). These theories stipulate that descriptive mental features, such as novelty, are caused by a literal novelty detector in your mind or brain (i.e., a process of the same name). Other theories in the appraisal category, by contrast, define appraisals only as descriptive mental features, without making any causal claims; for example, if an instance of emotion contains some "novelty," then something is experienced as novel during the emotional event, but there is no presumption that this experience is caused by single mechanism of the same name or otherwise (e.g., Clore & Ortony, 2000, 2008; Ortony & Clore, 2015). There is conceptual value to distinguishing these appraisal variants, because in the former case, states and processes amount to the same thing: Separate and qualitatively distinct mechanisms (i.e., appraisals) are assumed to produce correspondingly specific and qualitatively distinct states that are described with mental features of the same name. In the latter case, no such process-content dualism is required; there is no presumed parallelism between a mechanism and the resulting experience (there is no anger mechanism causing anger, no "goal-relevance" mechanism evaluating goal relevance, etc.), making the more descriptive appraisal theories easier to integrate with psychological construction theories (for discussion, see Barrett, 2013, 2015, 2017; Barrett et al., 2007; Gross & Barrett, 2011).

Improving How We Map the Conceptual Space of Emotion Theories

Identify Essentialism Where It Exists

Agnes's taxonomy can be improved by considering a conceptual distinction that separates a constructionist theory like Russell's from what she calls dimensional appraisal theories (or Flavor 2 appraisal theories): *essentialism*. Essentialism is the belief that within a category, instances named by the same word (such as *anger, pride, awe*, etc.) or a phrase (e.g., "skeptical emotion theories") share a deep, immutable causal mechanism that makes them what they are (this is essentialism as described by John Locke). Russell's psychological construction theory of emotion contains no essentialism. By contrast, theories of appraisals-as-mechanisms (including dimensional appraisal theories), like basic emotion theories, indulge in Lockean essentialism; I have referred to these as "classical emotion theories" (Barrett, 2017, in press-a, in press-b). For example, classical theories of emotion assume that a psychological phenomenon is caused by a dedicated mechanism of the same name (e.g., in basic emotion theories, the experience of fear is caused by a "fear" mechanism; in theories of appraisals-as-mechanisms, the experience of novelty is caused by a "novelty" mechanism). Furthermore, classical theories hypothesize a specific, dedicated underlying causal mechanism for each emotion category, either a population of dedicated neurons (for a review, see Tracy & Randles, 2011) or a particular configuration of appraisals (e.g., Scherer, 2009).

Classical theories also share another version of essentialism: the assumption either that within a category, instances share a "fingerprint" (i.e., a pattern of features that are similar across instances of the category) or that there is one best "instance" of the category (this is essentialism as advocated by Plato). The Platonic essence is usually assumed to issue from the Lockean essence (i.e., each emotion category supposedly has a specific, synchronized pattern of measurable changes in the face, in the body, in behavior, etc., that can be used to diagnose instances of that category), and the pattern is caused by the dedicated emotion circuit (in basic emotion theories) or by the pattern of appraisals (in theories where appraisals are mechanisms).[1]

[1] Originally (starting with Irons, 1894; Gendron & Barrett, 2009), appraisal theories were offered to account for variation in emotional phenomena, and although in principle they acknowledge the likelihood of such variation, their theorizing and research tends to focus on the presumed "basic" categories for a number of years (Barrett, Ochsner, & Gross, 2007). More recently, dimensional appraisal theories have dropped the assumption that appraisals produce "basic emotion"-type fingerprints; instead, each appraisal dimension or pattern of appraisals is thought to cause a synchronized, diagnostic set of physical changes (see Moors, this issue). In principle, both theory categories (basic emotion and appraisals-as-mechanisms) allow for some variation within an emotion category like anger. Just like a fingerprint can vary from one instance to another because of the oils and substances on your fingertips, the temperature of your skin, and the surfaces you touch, even though the underlying ridges on your skin are constant, so too do basic emotion theories allow for variation in movements of the face, in electrical signals of the autonomic nervous system, in acoustical changes of the voice, in voluntary movements of the body, and so on. Some randomness is expected, and other processes, independent of an emotion itself, are thought to interfere with the fingerprint, such as "display rules" or other regulatory strategies, such as suppression (Ekman & Cordaro, 2011; Gross, 2015; Matsumoto, Keltner, Shiota, Frank, & O'Sullivan, 2008; Roseman, 2011; Tracy & Randles, 2011). Nonetheless, it is assumed that an emotion's fingerprint exists and can be used to uniquely "recognize" the emotion in the same way that a fingerprint uniquely identifies an individual.

Due to their mutual essentialism, classical theories share another characteristic: They are virtually nonfalsifiable. They are bolstered, despite evidence that persistently calls them into doubt, by a phenomenon called "psychological essentialism" (Medin & Ortony, 1989). Psychological essentialism permits scientists to posit a hypothetical or unseen essence in the absence of any evidence whatsoever of what the essence might be. For example, Ekman's hypothetical affect program (Ekman & Cordaro, 2011; Tomkins & McCarter, 1964), Panksepp's hypothetical FEAR system (Panksepp, 1998), and Adolphs's "central emotion state" or "functional emotion state" (Adolphs, in press; Anderson & Adolphs, 2014) are examples of psychological essentialism; even the idea that an appraisal is a causal mechanism that produces an identical, descriptive mental feature of the same name can be understood as an example of psychological essentialism.

Some scientists consider essentialism a useful strategy for scientific inquiry because they believe that it mirrors the structure of the real world (i.e., they believe the world is full of natural kind categories; Bloom, 2000; Kornblith, 1993; Pinker, 1997). Others, however, believe that essentialism is a particularly poor strategy for science to take (e.g., Lewontin, 2000). It is worth pointing out, however, that the history of science can be read as a long, slow march away from essentialist thinking, discovering that universal laws are actually contextual (e.g., in physics, with the discovery of quantum mechanics) and discovering that variation is meaningful and is not in error (e.g., in biology, with Darwin's [1859/1964] *On the Origin of Species*, and then again a century later with the study of epigenetics and genomics). Moreover, essentialism has been shown to interfere with scientific thinking, particularly when it comes to understanding biological categories (Mayr, 2004); the workings of evolution and natural selection (Gelman & Rhodes, 2012); and, yes, the nature of emotion (Barrett, 2017). Over a century ago, James (1890/1950) pleaded for the new science of psychology to abandon the essentialism of faculty psychology when he wrote, "The trouble with the emotions in psychology is that they are regarded too much as … psychic entities, like the old immutable species in natural history" (p. 449). Essentialism is so powerful, however, that it has led the field to completely misinterpret James's constructionist ideas as essentialist, creating an illusory James–Lange theory (Barrett, 2017; Gendron & Barrett, 2009). Denying the existence of emotion essences is not equivalent to denying the existence of emotional phenomena (Barrett, 2012). A really powerful theory of emotion might even explain why people essentialize in the first place (see Barrett, 2017).

Incorporate the Hypothesis of Emotion Populations

In creating categories for emotion theories, Agnes's (this issue) taxonomy anchors on the classical view of emotion and adjusts away from it, defining "skeptical" theories as those that do not make classical assumptions. But our understanding of the scientific landscape would be significantly improved by actively embracing what we gain when we eject essentialism: an awareness of the vast *variation* in emotional life. The fundamental importance of variation as a phenomenon to be explained (rather than error to be ignored) is nicely illustrated by Russell's theoretical approach. Attempting to explain the existence of emotional phenomena, in all their glorious variability, without the need for emotion essences, gives us a new category of emotion theories: those that appreciate the idea of population thinking (a useful concept from biology; Barrett, 2013, 2017, in press-b).

Implicit in Russell's constructionist theory of emotion (and explicit in my theory of constructed emotion) is the hypothesis that an emotion category refers to a *population* of highly variable, situated instances (also see Barrett, 2006b, 2013, 2015, 2017, in press-b). So an emotion is not an entity with firm boundaries in nature—it is a category of instances that vary because each one is tailored to the constraints of the immediate environment. For example, an instance of happiness can be pleasant and arousing (e.g., you are finishing a challenging task with no errors and hear applause), pleasant and quiescent (e.g., you feel comfortable and rested after a good night's sleep), and even unpleasant (e.g., you want to call your friend to share your recent success, but he is unreachable; Wilson-Mendenhall, Barrett, & Barsalou, 2013). The actions you make in happiness will depend on the situation (e.g., you might laugh, smile, cry, jump, sigh, shout, slam your fist against a table, etc.)—whatever has been most functional for you in past, similar situations. And your autonomic nervous system changes will be similarly variable across instances of happiness, because those responses support action (Obrist, 1981; Obrist, Webb, Sutterer, & Howard, 1970); therefore, there is no one-to-one correspondence between a behavior, a physiological pattern, and an emotion word (cite meta-analyses). Thus, in a construction approach, an emotion category (and, in fact, any mental category) is assumed to have no Platonic essence.

Recognize the Hypothesis of Domain General Mechanisms

Russell's theory also exemplifies a second key feature of all constructionist theories: instances within the same emotion category (e.g., happiness); instances across emotion categories (e.g., happiness vs. fear); and even instances of nonemotion categories such as thoughts, beliefs, perceptions, and so on, emerge from a more fundamental set of *common or domain general processes* (Barrett, 2013, 2015, 2017, in press-b; Barrett & Satpute, 2013). As a result, an emotion category (and, in fact, any mental category) is assumed to have no Lockean essence. No mental category (emotion or otherwise) is presumed to be any more biologically basic than another. The validity of any constructionist theory depends on specifying the mechanisms of causation, not on finding a stable pattern of observable consequences stemming from those mechanisms (whereas in the classical view, there is more emphasis on the latter than the former, because fingerprints should always be measurable, whereas essences can be hidden).

Dimensional appraisal theories, like other theories of appraisals-as-mechanisms, argue against the existence of domain general mechanisms. Agnes, for example, argues that emotional and nonemotional episodes are caused by different appraisals (e.g., the stimuli in the former are evaluated as more goal relevant than the latter). As a result she argues, following Frijda (1986), that action tendencies have more control precedence. But this distinction is easily pushed aside, both logically

and empirically (nervous systems are wired in such a way that every action ever performed is goal relevant; anything else would be metabolically frivolous).

Degeneracy

Constructionist theories incorporate another important concept from biology, called *degeneracy* (Edelman & Gally, 2001; Marder & Taylor, 2011; Tononi, Sporns, & Edelman, 1999). Degeneracy, which refers to the capacity for structurally dissimilar systems or processes to give rise to identical outcomes (Edelman & Gally, 2001), is a property of virtually every level of analysis in biological systems, from the systems inside cells to the entire organism. For example, different proteins can catalyze the same reaction of enzymes (Edelman & Gally, 2001; Tononi et al., 1999), different antibodies can bind to the same antigen (Edelman & Gally, 2001), different genotypes can produce the same phenotype (Edelman & Gally, 2001; Tononi et al., 1999), different neurons can give rise to the same network (Marder & Taylor, 2011; Tononi, Edelman, & Sporns, 1998; Tononi et al., 1999), and different patterns of network interaction can give rise to the same behavior (Price & Friston, 2002). Emotion categories have degenerate instances (varying in their associated facial configurations, autonomic configurations, and appraisals), and any emotional instance (whether you call it an emotion or an emotional episode) can be caused by a different pattern of different domain general system interactions implemented as different brain states (see Clark-Polner, Johnson, & Barrett, in press).

Abandon Stimulus → Response Hypotheses

Agnes's analysis relies on several varieties of stimulus–response (or S–R) mechanisms, and in this regard her approach is in need of a significant tune-up using recent discoveries in modern neuroscience. Her S–R approach, as she describes it, is inconsistent with the anatomic, signal processing, and metabolic properties of the brain (for a review, see Barrett, 2017, in press-b; Barrett & Simmons, 2015; Chanes & Barrett, 2016; Sterling, 2012; Sterling & Laughlin, 2015).

An instance of an emotion category (and any other mental category) emerges as the brain makes meaning of incoming sensory input from the body and the world. Every waking moment of your life, your brain is taking in constantly changing, noisy, incomplete sensory information from the world and transforming it into sights, sounds, smells, and so on. From your brain's perspective, your body is another domain that is sending constantly changing, noisy, incomplete sensory input from your blood rushing, your muscles stretching, your lungs expanding, and so forth; the autonomic nervous system, endocrine system, and immune system all create sensory changes within what scientists call the internal milieu of your body, and your brain makes sense of these as affective feelings that belong to physical symptoms, emotions, thoughts, perceptions, and so on. How does your brain make sensations meaningful? By categorizing them—using past experience, organized as concepts, to explain what caused the sensations and what to do about them (i.e., how to act). This is the basis of my own theory of constructed emotion:

> In every waking moment, your brain uses past experience that function as concepts to guide action and give sensations meaning. In this manner, your brain models your body in the world. When the concepts involved are emotion concepts, your brain constructs instances of emotion.

To demonstrate categorization using concepts from past experience, see Figure 1. If you are like most people who have never seen the image in this figure, then you are in a state of "experiential blindness." Your brain cannot categorize the visual input—it cannot make sense of it—so all you see are black and white blobs. To cure your experiential blindness, please turn to the appendix, and then return to Figure 1.

After viewing the appendix, most people now see an object in Figure 1. So what does this exercise demonstrate? Your brain

Figure 1. An example of categorizing to construct an experience.

added information, stored from your (very recent) past, to make sense of the incoming sensory input (visual) to *construct* your experience of the object in Figure 1. This example is instructive in several ways. First, the construction process is ongoing, obligatory, and automatic; notice that you had no sense of agency or effort in the construction process. No matter how hard you try, you cannot introspect about how your brain accomplished this feat of making incoming sensations from Figure 1 into a meaningful visual experience. Also, it is virtually impossible to "unsee" the object—to deconstruct the experience by the sheer force of will.

To the best of our current knowledge, here is what went on in your brain. Neurons in certain parts of your brain changed the firing of neurons in your visual cortex to construct your experience of lines that aren't present, linking the blobs into the shape of a cow, the image of which isn't physically there on the page. Scientists call it "simulation" (Barsalou, 2008). Simulation is when the neurons in some parts of your brain changed the firing of sensory neurons in other parts of the brain so that you can, for example, see lines and other visual features without sensory input. Simulation can be visual, as in this example, but it also involves your other senses. If you've ever had a song stuck in your head, or put food into your mouth, expecting to taste one thing but then experiencing the shock of tasting something entirely different, then you have experienced simulation in other sensory modalities. Memories, daydreams, mind wandering—these are also examples of simulation. In the science of emotion, we measure this kind of simulation all the time without realizing it. We hook people up to blood pressure monitors, electrocardiograms, and so on; show them evocative images; and then measure changes in autonomic nervous system activity, even though people are sitting perfectly still. Whenever you ask respondents to report on an experience that has happened in the past, this is also tapping simulation. Simulation during brain scanning produces activity in somatosensory and motor cortices when subjects are completely still, in primary visual cortex when eyes are closed, and even in primary interoceptive cortex (for sensing changes in the core of the body) when there is no real threat or reward immediately present (Wilson-Mendenhall et al., 2013). Understanding the brain dynamics of simulation—how people apply knowledge wired into the brain to create experiences and perceptions in a particular context—will reveal how the brain constructs experiences and perceptions of emotion.

A full explanation of simulation is beyond the scope of this commentary, so a brief summary will have to suffice. The first thing you must realize is that your brain is not merely responding to stimuli in the world. Your simulations function like predictions that continuously *anticipate*, rather than react to, sensory inputs from the world. Your brain is wired to be a generative model of your world by using past experience to actively create simulations that best fit the situation you are in. The second insight is that predictions, as simulations, are then corrected by sensory input from the world, so information from the world is feedback on how good the simulations are. This includes the neurons not only for vision, audition, touch, taste, and smell but also for interoception, because from a brain's-perspective, the body is part of the brain's world (because the body holds the brain). Your brain is constantly generating predictions of upcoming sensations and then adjusting these predictions (more or less) by computing error signals that track the difference between the predicted sensations and those that are incoming from the sensory world. And the brain is making not only sensory predictions but also motor predictions; it is anticipating the motor changes that will be required in a moment from now by changing the firing of motor neurons before they are needed. In fact, your brain generates visceromotor predictions (to control your autonomic nervous system, your neuroendocrine system, and your immune system) and voluntary motor predictions first, and then anticipates the sensory consequences of those visceromotor/motor predictions (i.e., predicted motor changes produce sensory predictions) so that, in a sense, sensation follows (and is dependent on) action (see Barrett, 2017, in press-b; Barrett & Simmons, 2015; Chanes & Barrett, 2016; Clark, 2013; Friston, 2010; Howhy, 2013).

When your brain creates a prediction from past experience, it issues not one neural pattern but an entire population of potential predictions, each having some probability (computed with Bayesian priors) of being the best fit to the current circumstances (Barrett, 2017). This population of neural patterns is, for all intents and purposes, being treated by your brain as similar for some purpose—to make meaning of and dealing with the impending sensory array. Another insight of the theory of constructed emotion, then, is that this population is a *concept*, constructed as you need it, on the fly (e.g., Barsalou, 1983, 2003; Barsalou, Simmons, Barbey, & Wilson, 2003). Certain predictions will provide a better fit to the incoming sensory input, and these become your perception and guide your action. So, constructing meaning by correctly anticipating (predicting and adjusting to) incoming sensations is what I mean when I say that the brain is categorizing sensations to construct an instance of emotion. Sensations are conceptualized (i.e., categorized) so that they are (a) actionable in a situated way and therefore (b) meaningful, based on past experience. The sensory array in need of prediction and action contains both interoceptive inputs from the body representing the allostatic changes in the body's various systems (the internal world) and exteroceptive inputs representing sensory changes in outside world. When past experiences of emotion (e.g., happiness) are used to categorize the predicted sensory array and guide action, then an experence of that emotion (happiness) is experienced or perceived. An emotional instance is constructed the way that all other perceptions are constructed, using the same neural systems (and, correspondingly, the same domain-general psychological processes). This is why neuroscientist Gerald Edelman (1998) referred to experience as the remembered present.

Conclusions

Categories are necessary: We can't live or do science in a world where everything is different from everything else, where learning in one situation does not allow us to predict in the next. For categories to be maximally useful in science, we have to choose our scientific (and philosophical goals) very carefully. We must also realize that definitions of emotion are stipulated, not discovered, and that our goals as scientists are often rooted in a

(sometimes implicit) theory of human nature (Barrett, 2017, in press-b).

Although many emotion theories are rife with essentialism, there are theoretical approaches that formulate emotion categories, and measure them, without reifying them with essences. In my view, this is what constructionist theories of emotion do (whether they are theories of psychological construction, social construction, neuroconstruction, or an integration of all three, plus a little rational constructionism thrown in for good measure, i.e., the theory of constructed emotion; Barrett, 2017). Constructionist theories and descriptive appraisal theories (but not theories of appraisals as mechanisms) incorporate population thinking and domain-general mechanisms rather than essentialism. In those theories, variability is *assumed* to be the norm rather than a nuisance to be explained after the fact.

When we augment Agnes's taxonomy as I have suggested, we can see that some of the most intractable questions are rooted in the classical views of emotion but completely dissolve away in constructionist approaches. In constructionist theories, for example, it is no longer meaningful to ask what is, and what is not, an emotion. You don't "have" emotions, "display" emotions, or "recognize" them. You construct emotions as experiences or perceptions—they emerge from complex dynamics within your nervous system, which is constantly in dynamic interaction with the surrounding context that usually includes other creatures, each of whom has a dynamically fluctuating nervous system. And the need to create a false dichotomy between "emotions" and "emotional episodes" all but falls away. Whatever you call them, emotional phenomena can be understood as events within a nervous system that continuously transitions from one state to another, described by low dimensional features (such as valence, arousal, and various appraisal dimensions).

Funding

This article received funding from the National Cancer Institute (U01 CA193632), the National Institute on Aging (R01 AG030311), and the U.S. Army Research Institute (W911NF-15–1-0647; W911NF-16–1-0191).

References

Adolphs, R. (in press). How should neuroscience study emotions? By distinguishing emotion states, concepts, and experiences. *Social Cognitive and Affective Neuroscience*.

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

Arnold, M. B. (1960a). *Emotion and personality: Vol. 1. Psychological aspects*. New York, NY: Columbia University Press.

Arnold, M. B. (1960b). *Emotion and personality: Vol. 2. Physiological aspects*. New York, NY: Columbia University Press.

Barrett, L. F. (2006a). Emotions as natural kinds? *Perspectives on Psychological Science*, *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.

Barrett, L. F. (2012). Emotions are real. *Emotion*, *12*, 413–429.

Barrett, L. F. (2013). Psychological construction: The Darwinian approach to the science of emotion. *Emotion Review*, *5*(4), 379–389.

Barrett, L. F. (2015). Ten common misconceptions about the psychological construction of emotion. In L. F. Barrett, & J. A. Russell (Eds.), *The psychological construction of emotion* (pp. 45–79). New York, NY: Guilford.

Barrett, L. F. (2017). *How emotions are made: The new science of the mind and brain*. New York, NY: Houghton Mifflin Harcourt.

Barrett, L. F. (in press-a). Functionalism cannot save the classical view of emotion. *Social Cognitive and Affective Neuroscience*.

Barrett, L. F. (in press-b). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*.

Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annual Review of Psychology*, *58*, 373–403.

Barrett, L. F., Ochsner, K. N., & Gross, J. J. (2007). On the automaticity of emotion. In J. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 173–218). New York: Psychology Press.

Barrett, L. F., & Satpute, A. (2013). Large-scale brain networks in affective and social neuroscience: Towards an integrative architecture of the human brain. *Current Opinion in Neurobiology*, *23*, 361–372.

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, *16*, 419–429.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*, 211–227.

Barsalou, L. W. (2003). Situated simulation in the human conceptual system. *Lang Cogn Process*, *18*, 513–562.

Barsalou, L. W. (2008). Grounded cognition. *Annu Rev Psychol*, *59*, 617–645.

Barsalou, L. W., Simmons, W. K., Barbey, A K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*, 84–91.

Bloom, P. (2000). *How children learn the meanings of words: Learning, development and conceptual change*. Cambridge, MA: MIT Press.

Chanes, L., & Barrett, L. F. (2016). Redefining the role of limbic areas in cortical processing. *Trends in Cognitive Sciences*, *20*, 96–106.

Clark, A. (2013). Whatever next? Predictive brains, situated agents and the future of cognitive science. *Behavioral and Brain Sciences*, *36*, 181–253.

Clark-Polner, E., Johnson, T., & Barrett, L. F. (in press). Multivoxel pattern analysis does not provide evidence to support the existence of basic emotions. *Cerebral Cortex*.

Clore, G. L., & Ortony, A. (1991). What more is there to emotion concepts than prototypes. *Journal of Personality and Social Psychology*, *60*, 48–50.

Clore, G. L., & Ortony, A. (2000). Cognition in emotion: Never, sometimes, or always? In R. D. Lane, & L. Nadel (Eds.), *The cognitive neuroscience of emotion* (pp. 24–61). New York, NY: Oxford University Press.

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.

Clore, G. L., & Ortony, A. (2013). Psychological construction in the OCC model of emotion. *Emotion Review*, *5*, 335–343.

Dallenbach, K. M. (1951). A puzzle-picture with a new principle of concealment. *The American Journal of Psychology*, *64*, 431–433.

Darwin, C. (1964). *On the origin of species*. Cambridge, MA: Harvard University Press. (Original work published 1859).

Edelman, G. M. (1998). *The remembered present: A biological theory of consciousness*. New York, NY: Basic Books.

Edelman, G. M., & Gally, J. A. (2001). Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences*, *98*, 13763–13768.

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review*, *3*, 364–370.

Frijda, N. H. (1986). *The emotions*. New York, NY: Cambridge University Press.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*, 127–138.

Gelman, S. A., & Rhodes, M. (2012). Two-thousand years of stasis how psychological essentialism impedes evolutionary understanding. In K. R. Rosengren, S. Brem, E. M. Evans, & G. Sinatra (Eds.), *Evolution challenges: Integrating research and practice in teaching and learning about evolution* (pp. 3–21). New York, NY: Oxford University Press.

Gendron, M., & Barrett, L. F. (2009). Reconstructing the past: A century of ideas about emotion in psychology. *Emotion Review*, *1*, 316–339.
Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry*, *26*, 1–26.
Gross, J. J., & Barrett, L. F. (2011). Emotion generation and emotion regulation: One or two depends on your point of view. *Emotion Review*, *3*, 8–16.
Hohwy, J. (2013). *The predictive mind*. Oxford, UK: Oxford University Press.
Irons, D. (1894). Prof. James' theory of emotion. *Mind*, *3*, 77–97.
James, W. (1890). *The principles of psychology* (Vol. 1). New York, NY: Holt.
Kornblith, H. (1993). *Inductive inference and its natural ground: An essay in naturalistic epistemology*. Cambridge, MA: MIT Press.
Lazarus, R. S. (1966). *Psychological stress and the coping process*. New York, NY: McGraw-Hill.
Lewontin, R. C. (2000). *The triple helix: Gene, organism, and environment*. Cambridge, MA: Harvard University Press.
Lightman, A. (2005). *A sense of the mysterious: Science and the human spirit*. New York: Vintage.
Marder, E., & Taylor, A. L. (2011). Multiple models to capture the variability in biological neurons and networks. *Nature Neuroscience*, *14*, 133–138.
Matsumoto, D., Keltner, D., Shiota, M., Frank, M., & O'Sullivan, M. (2008). Facial expressions of emotion. In M. Lewis, J. Haviland, & L. F. Barrett (Eds.), *Handbook of emotion* (3rd ed., pp. 211–234). New York, NY: Guilford.
Mayr, E. (2004). *What makes biology unique*? New York, NY: Cambridge University Press.
Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou, & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). New York, NY: Cambridge University Press.
Moors, A. (2014). Flavors of appraisal theories of emotion. *Emotion Review, 6*, 303–307.
Murphy, G. L. (2002). *The big book of concepts.* Cambridge, MA:: MIT Press.
Obrist, P. A. (1981). *Cardiovascular psychophysiology*. New York, NY: Plenum Press.
Obrist, P. A., Webb, R. A., Sutterer, J. R., & Howard, J. L. (1970). The cardiac–somatic relationship: some reformulations. *Psychophysiology*, *6*, 569–587.
Ortony, A., & Clore, G. (2015). Can an appraisal model be compatible with psychological constructionism? In L. F. Barrett, & J. A. Russell (Eds.), *The psychological construction of emotion* (pp. 305–333). New York, NY: Guilford.
Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. New York, NY: Oxford University Press.
Pinker, S. (1997). *How the mind works*. New York: Norton.
Price, C. J., & Friston, K. J. (2002). Degeneracy and cognitive anatomy. *Trends in Cognitive Sciences*, *6*, 416–421.
Roseman, I. J. (2011). Emotional behaviors, emotivational goals, emotion strategies: Multiple levels of organization integrate variable and consistent responses. *Emotion Review*, *3*, 434–443.
Russell, J. A. (1991). In defense of a prototype approach to emotion concepts. *Journal of Personality and Social Psychology*, *60*,37–47.
Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*, 145–172.
Russell, J. A. (2015). My psychological constructionist perspective. In L. F. Barrett, & J. A. Russell (Eds.), *The psychological construction of emotion* (pp. 183–208). New York, NY: Guilford Press.
Sander, D., Grandjean, D., & Scherer, K. R. (2005). A systems approach to appraisal mechanismsin emotion. *Neural Networks*, *18*(4), 317–352.
Scherer, K. R. (2009). Emotions are emergent processes: they require a dynamic computational architecture. *Philosophical Transactions of the Royal Society of London, Series, B, Biological Sciences*, *364*(1535), 3459–3474.
Sterling, P. (2012). Allostasis: A model of predictive regulation. *Physiology & Behavior*, *106*, 5–15.
Sterling, P., & Laughlin, S. (2015). *Principles of neural design.* Cambridge, MA: MIT Press.
Tomkins, S. S., & McCarter, R. (1964). What and where are the primary affects? Some evidence for a theory. *Perceptual and Motor Skills*, *18*, 119–158.
Tononi, G., Edelman, G. M., & Sporns, S. (1998). Complexity and coherency: Integrating information in the brain. *Trends in Cognitive Sciences*, *2*, 474–484.
Tononi, G., Sporns, O., & Edelman, G. M. (1999). Measures of degeneracy and redundancy in biological networks. *Proceedings of the National Academy of Sciences*, *96*, 3257–3262.
Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review*, *3*, 397–405.
Wilson-Mendenhall, C. D., Barrett, L. F., & Barsalou, L. W. (2013). Situating emotional experience. *Frontiers in Human Neuroscience*, *7*, 1–16.

Appendix

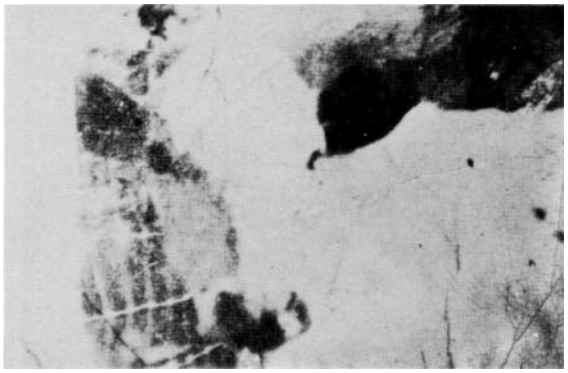

Figure A1. A photograph of a cow that Karl Dallenbach turned into an illusion (Dallenbach, 1951). From *American Journal of Psychology*. Copyright 1951 by the Board of Trustees of the University of Illinois. Used with permission of the University of Illinois Press.