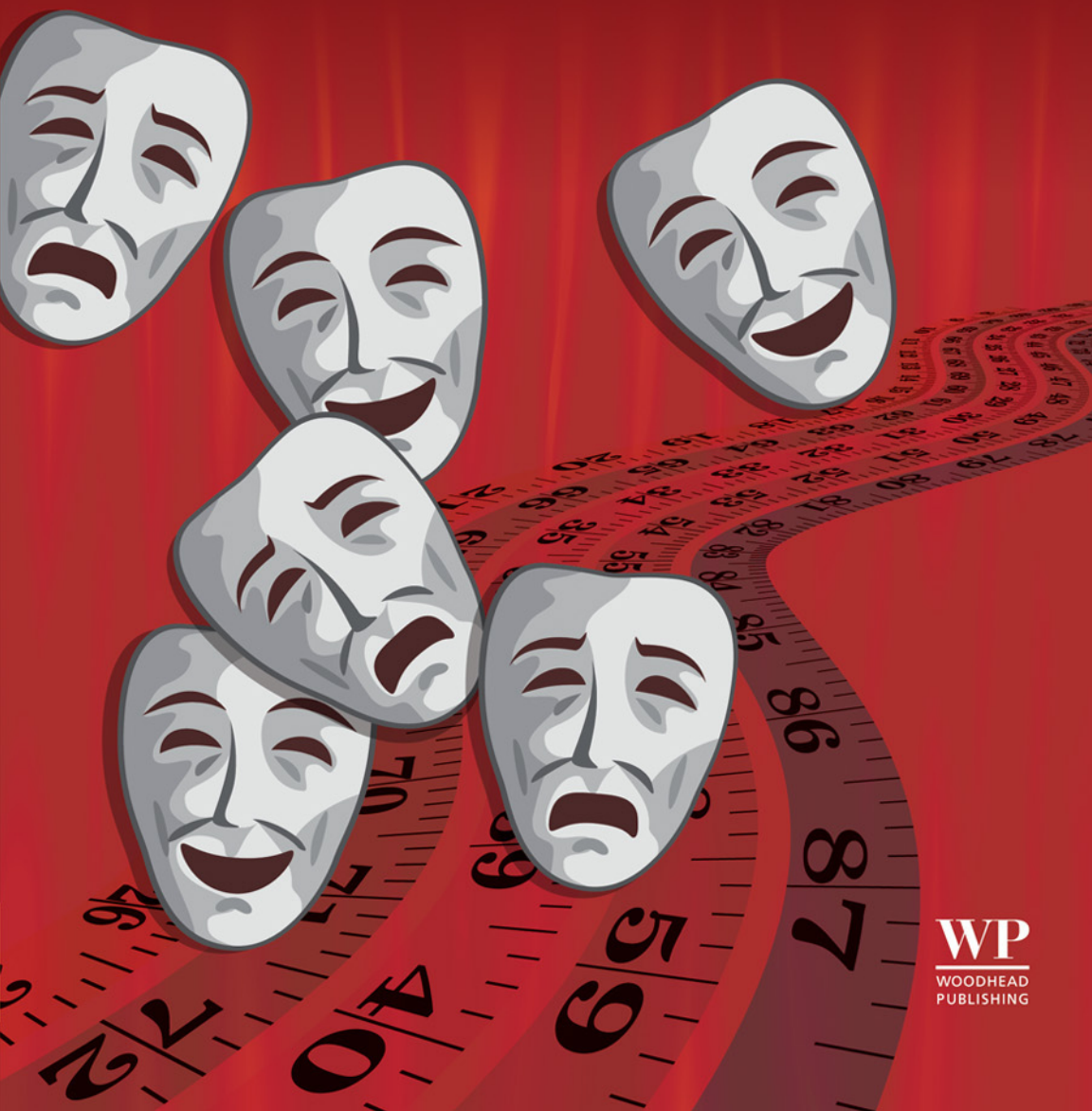
Emotion Measurement

Edited by Herbert L. Meiselman



WP
WOODHEAD PUBLISHING

Navigating the Science of Emotion

2

Lisa Feldman Barrett[1,2]
[1]Department of Psychology, Northeastern University, Boston, MA, United States
[2]Department of Psychiatry, The Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital/Harvard Medical, Charlestown, MA, United States

1 Preliminary observations

Those who are new to the science of emotion have a common experience: when they crack open a textbook or review chapter, they are confronted with a perplexing *multitude* of emotion theories that vary a great deal from one another in almost every way imaginable; theories disagree 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. In an attempt to bring some order to this dizzying cornucopia of theories, writers have created a particular narrative framework where theories are assembled into a few broader groupings, or categories, much like what appears in this volume. In chapter "Theoretical Approaches to Emotion and Its Measurement," for example, we are told that there are three families (ie, categories) of emotion theories (basic emotion, dimensional, and appraisal), and that they agree on very little other than that emotions, as brief reactions with synchronized components (expressions, action tendencies, bodily reactions, feelings, and appraisals), are triggered by "relevant" and "significant" objects in the world. By grouping variable theories together into a category, while separating others into different categories, a writer digests the variation, inviting you (the reader) to ignore certain distinctions between the theories (within the same category) and to focus your attention on others (in different categories). Any category is a grouping of items, events, objects, and even theories that are treated as similar *for some purpose* (Murphy, 2002). So, a category of theories is a grouping of theories that are deemed similar for some purpose. In any science, the organizing principle for grouping theories—the purpose—is determined by the writer's goal.

For the past century or so, the goal of many writers has been to argue over whether certain emotions are "basic" (forged in evolution, present in other animals, and having their own, unique biological mechanism, as well as their own pattern of facial expressions, autonomic reactivity, vocalization, behavior, and so forth). Some writers argue for the existence of "basic" emotions by either selectively reviewing evidence, or attempting to explain why it is reasonable to claim that emotions are "basic" in the

Emotion Measurement. DOI: http://dx.doi.org/10.1016/B978-0-08-100508-8.00002-3

face of growing empirical evidence that does not support the hypotheses of basicness. Other writers argue in favor of abandoning the idea of "basic" emotions, without offering much in the way of alternative theory to replace it. Some writers merely survey the literature, describing disagreements in as neutral a way as possible. All yield a similar sort of framework. Sometimes there are three categories (as in chapter: Theoretical Approaches to Emotion and Its Measurement); sometimes more (check out the Wikipedia page for "emotion" https://en.wikipedia.org/wiki/Emotion).

In my view, the standard narrative framework in the science of emotion constitutes one of the largest barriers to scientific progress, leaving both newcomers and scientists ill-equipped to make informed decisions about how to measure emotions. The framework is muddled, first, by errors. For example, some scientists advocate the utility of measuring "affect," ie, general feelings of valence (pleasure and displeasure) and arousal (calm and still to highly activated). Based on this, other scientists mistakenly describe these theories as reducing all emotions to two dimensions (eg, chapter: Theoretical Approaches to Emotion and Its Measurement). Then, they criticize the theories for their inability to distinguish different emotions from one another. In the standard narrative, for example, anger is defined as a high-arousal, negative state. But so are fear, disgust, guilt, and a variety of other emotions. The problem here does not lie in the theories themselves—it lies in the narrative framework. So-called "dimensional" theories of emotion do not actually exist. Most of these theories hypothesize valence and arousal as necessary, but not sufficient, features of emotion. Categorizing theories as "dimensional" betrays a fundamental misunderstanding of how these theories work. A careful read of the literature reveals that no theory has ever hypothesized that emotions can sufficiently be reduced to or *explained by* valence and arousal. Instead, these theories hypothesize that valence and arousal are important (and perhaps necessary) *descriptive features* of all emotions. Mischaracterizing theories as "dimensional" and then criticizing them for the limitations of a "dimensional approach" is not helpful to scientific clarity. The grouping of "dimensional" theories is actually better understood as "psychological construction" theories, which are described in some detail below (also see Barrett & Russell, 2015a, 2015b).[1]

[1] Psychological construction theories hypothesize that emotions are created as the interpretation of affective changes. They integrate dimensional and categorical perspectives, in the following way: It is hypothesized that all emotional events, at their core, can be *described* as having psychologically primitive affective properties (they feel pleasant/unpleasant, a property called valence, and agitating/quiescent, a property called arousal); this is the "dimensional" aspect of the theories. Psychological construction theories also propose, however, that people automatically and effortlessly use some type of mechanism to these affective changes meaningful in relation to objects and events in the world; this is the "categorical" aspect of the theories. Psychological construction theories are mislabeled as merely "dimensional" because people view them through a reductionistic lens, causing a profound misunderstanding: they attempt to redefine an emotion (or ontologically reduce an emotion) to it's most basic elements, which they mistakenly believe to be valence and arousal. As far as I know, no modern psychological constructionist theories have suggested that emotions are nothing more than valence and arousal (ie, no modern theories posit that affect alone provides a sufficient characterization for emotion). Instead, most theories characterize emotions as phenomena that *emerge* from the interaction of more basic mechanisms. Emergence implies that the product (the emotional instance as a whole) is more than the sum of its parts, and has properties that the core systems (the individual contributing parts) do not, making reductionism impossible. Furthermore, valence and arousal, as descriptive features of affect, are not even basic mechanisms. The affective circumplex (Barrett, 2004; Russell, 1980; Russell & Barrett, 1999; see chapter: Theoretical Approaches to Emotion

A further error in the standard narrative is that it characterizes valence and arousal as causal processes or mechanisms when the theories themselves do not. Valence and arousal are descriptive properties, not mechanisms that cause anything. They are also not unique to emotion (valence and arousal are fundamental features of all thoughts, beliefs, memories, perceptions, and so on; simply put, they are descriptive features of consciousness; for a discussion, see Barrett & Bliss-Moreau, 2009; Wundt, 1897/1998).

The standard framework also obscures some important variations within each theoretical "family." For our purposes, the most important distinction that is muddled involves the "appraisal" theory category. One variety of appraisal theories, which I have called *causal* or *classical appraisal theories* (Barrett, Mesquita, Ochsner & Gross, 2007; Gross & Barrett, 2011), assumes that "appraisals" are literal cognitive mechanisms that cause the subjective evaluations that, in turn, either cause or constitute emotions (Arnold, 1960a, 1960b; Frijda, 1986; Lazarus, 1966; Roseman, 2011; Sander, Grandjean, & Scherer, 2005; Scherer, 2005; see chapter: Theoretical Approaches to Emotion and Its Measurement). But another variety of appraisal theory characterizes "appraisals" as descriptive features, much like valence and arousal; for example, if an emotion is associated with the appraisal of "novelty," then during the experience of that emotion, something is experienced as novel. This descriptive variant of the appraisal category (Clore & Ortony, 2000, 2008; Ortony & Clore, 2015) is agnostic on the mechanisms that produce the appraisals. So, whereas causal appraisal theories would assume that an experience of novelty is caused by a literal novelty-detector in your brain (just as they mistakenly assume that the experience of pleasure and displeasure is caused by a literal valence detector in your brain), descriptive appraisal theories don't assume that there is a single mechanism that causes novelty (or valence). These *constitutive* or *constructive appraisal theories*, like psychological construction theories, do not assume that there is any parallelism between a mechanism and the resulting experience (ie, there is no anger mechanism causing anger, no "goal relevance" mechanism evaluating goal relevance, and so on).

The standard narrative framework not only conceals meaningful variation within a category of emotion theories, but it also obfuscates similarities across categories. The most important similarities are related to *essentialism*. Essentialism is the belief that a category of instances named by the same word (such as anger, pride, awe, etc.) or a phrase (eg, "basic emotion theories," "appraisal theories," etc.) share a deep, underlying causal mechanism (this is essentialism as described by John Locke). Basic emotion and causal appraisal theories indulge in Lockean essentialism. For example, both groupings of theories assume that a psychological phenomenon is caused by a dedicated mechanism of the same name (eg, in basic emotion theories, the experience of fear is caused

and Its Measurement, Fig. 2.2, which can also be found in Barrett & Russell, 1999, Fig. 2.1) is not an explanatory theory of emotion. It is a low dimensional, descriptive map that represents two properties or features of emotional experiences. These are properties or features of experience, valence and arousal, themselves cannot be mechanistically reduced, and are emergent properties of more basic processes. One well-known psychological construction theory by Russell (2003) is agnostic on how affect (described as feelings of valence and arousal) is caused. Another psychological construction theory, my own Conceptual Act Theory (Barrett & Bliss-Moreau, 2009; Barrett, Wilson-Mendenhall & Barsalou, 2015), makes very specific hypotheses about how affect arises from more fundamental mechanisms in the brain and body.

by a "fear" mechanism; in causal appraisal theories, the experience of novelty is caused by a "novelty" mechanism). Furthermore, both groupings of 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 (Scherer, 2009; see chapter: Theoretical Approaches to Emotion and Its Measurement).

Basic emotion and causal appraisal theories also share another version of essentialism: the belief that a group of instances either share a "fingerprint" (ie, a pattern of features that are similar across instances of the category) or that there is one best "instance" of the category (like a prototype; this is the sort of essentialism advocated by Plato). Both groupings assume that emotion categories have a "fingerprint," or a Platonic essence that issues from the Lockean essence (ie, 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 supposedly caused by the dedicated emotion circuit (in basic emotion theories) or by the pattern of appraisals (in causal appraisal theories). In principle, both theory groupings allow for variation within an emotion category. 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 account for this variation, 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 person.[2] Appraisal theories, to some extent, were motivated to account for variation in emotional phenomena, and while in principle they acknowledge the likelihood of such variation, their theorizing and research tends to focus on the presumed "basic" categories (cf. Barrett, Ochsner, & Gross, 2007).

Essentialist assumptions persist, despite accumulating evidence that they are false, for many reasons, most notably because of a phenomenon called "psychological essentialism" (Medin & Ortony, 1989), which allows people to posit a hypothetical or unseen essence in the absence of any evidence of what the essence might be (eg, the "affect program" concept) (Ekman & Cordaro, 2011; Tomkins & McCarter, 1964).

[2] Consider the idea of a "fingerprint." The pad of your finger has ridges which do not change from instance to instance. The sweat, dirt, ink, or other substances on the ridges of your finger are transferred to the surface of a brass door handle, a wooden table, a piece of paper, or whatever you touch, leaving a print. Your fingerprint won't look exactly the same each time you touch a surface. Sometimes you might grip the handle with more pressure or less. Some surfaces are rough with strong traction, while others are smoother and allow your fingers to slip a bit. Sometimes your skin might be warmer and more pliable. Even though your fingerprint does not look identical each time, it looks similar enough and is unique to you, and only you, so that it can be used to identify you but not other people.

This inoculates believers against disconfirming evidence, allowing them to continue to theorize about and believe in the existence of emotion essences, and to focus on the small proportion of published studies that support their existence, despite the even larger number of studies that disconfirm them (Barrett, 2006a, 2012, 2013; Barrett et al., 2007; Guillory & Bujarski, 2014; LeDoux, 2015; Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012; Touroutoglou, Lindquist, Dickerson, & Barrett, 2015).

The errors, the important differences between theories of the same category, and the implicit similarities in assumptions across categories, leave the science of emotion with an ineffective organizational framework making clear measurement recommendations almost impossible. Without meaningful theoretical integration, newcomers find it difficult to identify what is known about emotion with any degree of certainty. Scientists are paralyzed in a "disconfirmation dilemma" (Greenwald & Ronis, 1981) that makes accumulating knowledge about emotion almost impossible (except within theory-based silos). Lakoff recently called emotion an *essentially contested concept*: everyone agrees that emotions exist, but a variety of meanings are simultaneously employed for emotion, and scientific inquiry seems unable to settle the matter. Indeed, the science of emotion is struggling today with the very same dilemmas as it was a century ago (Gendron & Barrett, 2009). And the status of the empirical literature today is not that different: there are some experiments that support the hypothesis of emotion essences, but these are the tip of a much larger iceberg of scientific evidence that does not. Despite tremendous investments of research time and money, emotions remain mysterious and deeply perplexing.

Scientists realize that this confusion has a high price tag: measures don't work as well as expected, or sometimes at all. Studies don't replicate. There is a persistent confusion about how to measure emotions. And emotions don't predict behavior as well as they could (or by some accounts, at all). And the costs are rising. In the new "emotion economy," many businesses are investing millions of dollars and tremendous person-hours developing "emotion-aware" technology that they have been led to believe will be able to "read" emotions from perceiver-independent measurements of the face, body, and behavior (ie, they are developing technology and algorithms to measure emotions solely by tracking how facial muscles move, how autonomic signals in the body, how electrical signals across the scalp change, and so on). These companies have unwittingly ignored years of disconfirming evidence and embraced essentialism. But if the scientific literature is to be trusted, and one includes all well-designed and well-executed studies (not just those that support's one deeply held beliefs), then our best conclusion is that emotion essences don't exist. The unfortunate consequence is that so much effort and investment will be wasted. Even worse, people might end up concluding that emotions are not an important part of the equation in predicting behavior, all because they unwittingly used a set of unwarranted guiding assumptions to begin with. The empirical literature indicates the immediate need for a different scientific approach to explain what emotions are and how to measure them.

In this chapter, I offer a different approach—a narrative framework to reorganize the field according to whether or not theories assume the existence of emotion essences. Denying the existence of emotion essences does not mean denying the existence of emotions (Barrett, 2012). Instead, it means acknowledging the importance of

true variation in emotional life, and attempting to capture that variation when measuring an emotion category. It also means explaining the existence of emotion categories in all their glorious variability, without the need for emotion essences. Emotion theories vary in the extent to which they indulge in essentialism and in the type of essentialism they incorporate, if at all, and understanding these distinctions provides a more useful roadmap for the science of emotion, including ways to measure emotion.

For the remainder of the chapter, I introduce and discuss a new narrative framework, one that will allow you to make sense of the rest of this volume's chapters in terms of the underlying assumptions that are so rarely revealed and openly evaluated. Hopefully, this will give you a firmer footing on which to make informed measurement decisions. My goal is not to convince you that essentialism is detrimental. As you might have surmised, I believe it is. Essentialism has been show to interfere with scientific thinking, particularly when it comes to evolution and natural selection (Gelman & Rhodes, 2012) and emotion (Barrett, in press). Over a century ago, William James (1890/1998) pleaded for psychology to abandon essentialism: "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, in fact, that is has, ironically, led the field to completely misinterpret James's ideas as essentialist in nature (cf. Gendron & Barrett, 2009). But my goal in this chapter is not to convince you that William James's original intent was correct. Instead, I have opted for the more modest goal of providing you with a framework for identifying whether a theory is assuming essentialism or not. Hopefully, this will equip you to properly evaluate the remaining chapters in this volume, as well as understand the consequences of the assumptions (whether they are explicitly stated or not). By the end of this chapter, you will also hopefully be more aware of your own assumptions, as well as their consequences of your measurement choices. Within the next section, I not only offer a framework for navigating the science of emotion, but I also consider the general measurement model for each approach. The chapter ends with a discussion of a new analytic technique—pattern classification—which is touted as the way to identify emotion "fingerprints," "signatures," or "biomarkers," and how such claims are mathematically mistaken because the technique is being viewed through the lens of essentialism.

2 To essentialize or not to essentialize? That is the question

Throughout the ages, philosophers, scientists, and other scholars who concern themselves with the nature of the human mind have disagreed on the nature of emotion in similar ways. In general, scholars have aligned themselves with one of two positions. One position incorporates some sort of essentialism, which I refer to as *the classical view of emotion*; the second assumes variability is the norm, which I refer to as *the construction approach to emotion*. A revised framework for modern emotion theories, organized by essentialist assumptions, is presented in Fig. 2.1 (for a brief historical review, see Appendix A).

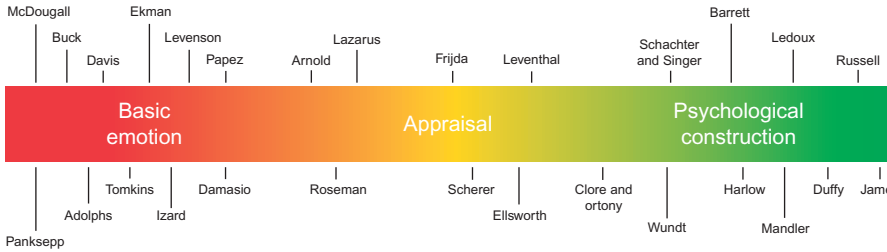


Figure 2.1 Modified from Gross and Barrett (2011). Emotion theories are loosely arranged along a continuum, populated with representative theories drawn from the field of psychology. Three "zones" are distinguished: (1) basic emotion, in red, for example, Anderson and Adolphs (2014), Buck (1999), Damasio (1999), Davis (1992), Ekman (1972), Izard (1993), Levenson (1994), McDougall (1908/1921), Panksepp (1998), and Tomkins (1962, 1963); (2) appraisal, in yellow, for example, Arnold (1960a, 1960b), Clore and Ortony (2008), Frijda (1986), Lazarus (1991), Leventhal (1984), Roseman (1991), Scherer (1984), and Smith and Ellsworth (1985); and (3) psychological construction, in green, for example, Barrett (2009), Duffy (1941), Harlow and Stagner (1933), James (1884), LeDoux (2015), Mandler (1975), Russell (2003), Schachter and Singer (1962), and Wundt (1897/1998). Theories in the red zone and the left-most portion of the yellow zone are much more essentialist than those in the right-most part of the yellow zone and the green zone (which are all non-essentialist theories). Indeed, the greatest heterogeneity in essentialist assumptions can be found in the appraisal zone, where classical appraisal theories (eg, Arnold, Roseman, and Lazarus) share many similar assumptions with basic emotion theories, whereas constituative appraisal theories (eg, Clore and Ortony) have more in common with psychological construction theories.

3 The classical view of emotion

In the first theoretical approach, which I call the *classical view*, an emotion is understood as a separate and independent ability, or *faculty*, caused by its own separate *processes*. In this approach, emotions are categorically different phenomena from perceptions and cognitions, and each emotion (eg, anger, sadness, fear, and so on) is categorically different from every other emotion, each being caused by a different mechanism. Some phenomena, like feelings of fatigue or love, are not considered to be emotions at all. Emotions, in this approach, are basic elements of the mind and body. In basic emotion theories, you "have" emotions and "recognize" emotions, such as happiness, and certain emotions are assumed to be psychologically and biologically primitive (meaning that the emotion cannot be further reduced at these levels of measurement). In causal appraisal theories, emotions can be caused and described in terms of their components (ie, the appraisal mechanisms).

In the classical view, each emotion faculty is assumed to have its own innate *physical essence* that distinguishes it from all other emotions. This might be a Lockean essence (an underlying causal mechanism that all instances of an emotion, such as happiness, share, making them that kind of emotion and not some other kind of emotion (eg, interest)). Lockean essences might be biological, such as a set of dedicated neurons (exemplified in basic emotion theories, ie, Tracy & Randles, 2011). They

might also be psychological, such as a set of evaluative mechanisms called "appraisals" (exemplified in causal appraisal theories). An emotion is also assumed to have a Platonic essence (a physical fingerprint that instances of that emotion share, but that other emotions do not, such a set of facial movements (an "expression") and a pattern of autonomic nervous system activity). Of course, no one is expecting an emotion's fingerprint to be identical each and every time the emotion is triggered, but it is assumed that instances of a category are similar enough to be easily diagnosed as the same emotion using objective (perceiver-independent) measures alone. Simply put, a handful of emotions (anger, sadness, fear, disgust, surprise, happiness, and perhaps a few others) are said to be natural kinds with firm boundaries in nature (Barrett, 2006a, 2006b). Scientists seem unable to agree on which emotions are "basic" or are natural kinds, however (for a discussion, see Ortony & Turner, 1990).

3.1 Measurement implications

There is a particular measurement theory implied in the classical view of emotions. If emotions are natural kinds, then the best way to measure them is using perceiver-independent tools, such as facial electromyography, measures of autonomic nervous system activation, or brain imaging. All measures (except perhaps, self-reports of subjective experience) should be correlated and therefore interchangeable with one another, because all signals have a common cause (the emotion essence) (for a discussion, see Barrett, 2000, 2006a, 2006b, 2011a, 2011b; Coan, 2010; Russell, 2003). This idea is embedded in classical measurement theory and depicted using the notation of probabilistic graphical models in Fig. 2.2. From an information theory standpoint, individual measures do not carry unique information about the emotion. Self-reports, on their own, are considered a fallible way of measuring emotions, because according to this view, people may not be conscious of their emotional state. So it is necessary to

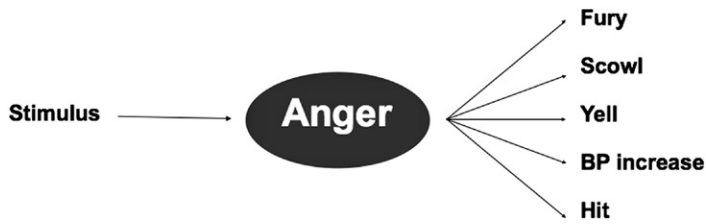


Figure 2.2 The measurement model for the classical view of emotion. *BP*, blood pressure. This is a causal model and a measurement model. The causal mechanism (the emotion essence) is depicted in the black oval. The resulting emotion fingerprint is on the right. In structural equation modeling, which is a type of probabilistic graphical model, the oval is latent (ie, not measurable) but its existence is established by correlations between the output measurements that make up the fingerprint (scowling, yelling, etc.). Various emotion essences have been proposed, including an affect program (eg, Ekman), appraisals (eg, Roseman, Scherer), brain circuits (Tracy & Randles, 2011), a dynamical brain system (Lewis, 2005), and even a pattern of brain activity (Kragel & LaBar, 2015).

scientifically test the accuracy of self-reports by comparing them with the perceiver-independent measurements before they can be used (and they should be correlated when accurate). When self-reports do not correlate with more objective measures (as is often the case), then self-reports are assumed to be inaccurate.

In the classical view, emotions are universal in humans (meaning that measurement tools should reveal the same underlying causes and fingerprints for each emotion across cultures), and are homologous in other animals (meaning that studying emotion in nonhuman animals will sufficiently reveal the secrets of emotions in humans). Therefore, classical conditioning with an aversive cue (like an electric shock) becomes "fear learning" (for a discussion see Barrett, 2012; LeDoux, 2014). Reward processing in rodents and monkeys should be the same as in humans because both are located in the so-called "limbic system," which is supposed to be evolutionarily primitive in the brain and identical in all mammals (or at least in primates) (MacLean, 1949, 1990; Panksepp, 1998); only humans are supposed to have the well-developed neocortex necessary for cognition and therefore for sophisticated emotion regulation. One implication of this formulation is that emotion words and concepts (which are part of cognition) are separate from the emotion themselves (Izard, 1993, 2011); self-reports of emotional experience are assumed to involve consciously accessing an emotional state (with more or less accuracy) compared to and described with adjectives on a questionnaire (such that the experience of emotion is presumed to be independent of the emotional state itself).

3.2 Drawbacks to using the classical view's measurement model

The major stumbling block in using the classical view to guide emotion measurement is that it does not fit the majority of data that have been collected to study it (cf. Barrett, 2006a, 2006b; Barrett et al., 2007). Some studies do support the classical view (many of which are reviewed in chapter: Theoretical Approaches to Emotion and Its Measurement), but these exist in a context of many more studies that disconfirm it. Disconfirming studies are rarely discussed in scientific reviews that are written in support of the classical view, which ultimately confuses newcomers to the field who, because they are reading broadly in an effort to familiarize themselves, quickly realize that there is substantially more to digest and integrate than those selectively written reviews provide.

Broader summaries of the literature do exist, however, and they paint a very different picture of the empirical landscape. Narrative summaries clearly show that perceiver-independent measures from the face, the body, and the brain do not correlate with one another to reveal an emotion's essence (ie, no Lockean essences have been identified for any emotion; Barrett, 2006a, 2006b; Barrett et al., 2007; Mauss & Robinson, 2009; Russell, 2003). Nor does neuroscience reveal evidence of emotion essences. For example, meta-analyses of neuroimaging studies have been unable to locate a specific region of the brain dedicated to a specific emotion (Lindquist et al., 2012). The brain does not contain an intrinsic network that is specifically dedicated to any emotion (Touroutoglou et al., 2015). Even individual neurons are not dedicated to specific emotions (Clark-Polner, Wager, Satpute, & Barrett, in press-b; Guillory &

Bujarski, 2014; Quiroga, Reddy, Kreiman, Koch, & Fried, 2005; Viskontas, Quiroga, & Fried, 2009).[3] For a discussion of some reasons why experiments on nonhuman animals do not reveal emotion essences, see Barrett (2012) and LeDoux (2014, 2015).

Nor do research findings give evidence of a consistent, specific physical fingerprint for each emotion (ie, no Platonic essences have been identified for any emotion; Barrett, 2006a, 2006b; Barrett et al., 2007). The facial expressions that are assumed to be specific to each emotion were not discovered but were actually stipulated, first by Darwin (1872/1965), and then later by Tomkins and McCarter (1964) (for a brief history, see Widen & Russell, 2013); scientists routinely repeat the mantra that these particular facial movements are universal displays of emotion, but studies that measure facial movements using facial electromyography do not support this claim (for reviews, see Barrett, 2006a, 2006b, 2011a, 2011b; Russell, Bachorowski, & Fernández-Dols, 2003). Even experiments using perceiver-based coding methods (called the Facial Action Coding System; FACS; Ekman & Friesen, 1978) do not find consistent evidence of these expressions during emotion (eg, see Table 13.2 in Matsuomoto et al., 2008).[4] Furthermore, engineers have shown that different combinations of action units can produce similar-looking expressions, violating a basic assumption of the classical view (Tian, Kanade, & Cohn, 2001). The point is not that people never scowl during anger or pout during sadness, but that they do not routinely do so. No one is claiming that facial movements are random or meaningless. The evidence clearly shows that people move their faces in various ways during happiness, during anger, or during any emotion. But variation is the norm within each emotion category. Because there is no set of facial movements that is consistent and specific to each emotion category (ie, no essential "expression"), emotion "recognition" studies are really "perception" studies where "agreement" (among perceivers, or between

[3] For example, amygdala neurons respond too slowly to be the brain essence of fear (usually responding about 250 ms after the image is shown). If you consider that it takes about another 500+ ms to mobilize a physical movement, then this is too slow to sound a fear "alarm" when a fearful object or event occurs. Cells in the medial temporal lobe (including the amygdala) appear to act as a memory cache for important things (eg, photos of friends, family, famous people, the patients themselves, landscapes, directions; some cells don't respond to anything for a few days, and then begin to respond when the experimenters walk into the room); at some other point, the cells might adopt and code for something entirely different that becomes important (Cerf, personal communication, 7/30/15).

[4] FACS coders isolate specific facial movements and indicate whether they are present or absent. Another version of FACS, called "Emotion FACS" or EMFACS, has coders decide whether an expression is present or absence by indicating whether an entire group of muscle movements (the expression) occurs en-mass, rather than detecting each movement one at a time; EMFACS produces better results supporting the Classical View, but it is less reliable (www.erikarosenberg.com), less precise and less objective. For example, of the hundreds of published studies using FACS and EMFACS to code facial muscle movements during emotion, a recent scholarly review listed only 25 that reported test subjects making spontaneous facial movements matching the configurations in the posed photos (Matsumoto et al., 2008; Table 13.2). Of the 15 studies using FACS coding, only 7 actually found evidence that spontaneous facial movements matched the expected configurations; 5 studies found that facial movements distinguished pleasant and unpleasant feelings, and 3 studies were not designed to test the question in the first place because the test subjects were only presented with pleasant stimuli like jokes and cartoons. Of the nine studies using EMFACS coding, all found evidence supporting the claim that people make spontaneous facial movements during emotion matching the expected facial expressions.

perceivers and the experimenter's expectations) serves as an estimate "accuracy" (instead of comparing a person's perceptions to some objective index of whether or not the emotion is present in the target).

Variation is also the norm for autonomic measurements taking during emotion. Although some writers have made a very persistent case for the existence of autonomic signatures (Friedman, 2010; Kreibig, 2010), actual meta-analytic summaries do not support such claims (Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 2000; Lindquist, Siegel, Quigley, & Barrett, 2013; Siegel et al., under review; for a discussion, see Quigley & Barrett, 2014). A recent meta-analysis from our lab, for example, summarized findings from over 200 experiments measuring autonomic reactivity during instances of emotion categories and failed to find distinct autonomic fingerprints for any emotion category. Instead, we documented tremendous variation both within and across categories; experimental context explained some of the variation, but even studies using the same methods, stimuli, and sampling from the same population of participants display such variation (eg, compare findings from Kragel & LaBar, 2013, with Stephens, Christie, & Friedman, 2010).

"Variation is the norm" is a fair summary of the experimental literature on emotion, to date. While some experiments do give evidence of emotion fingerprints, many more do not. Studies are either unable to distinguish one emotion from another over and above their basic affective differences, or the patterns that distinguish one emotion from another do not replicate across studies (for a recent review, see Barrett, 2013).

Probably the most robust and replicable finding in the science of emotion comes from classic emotion perception studies, but only when a particular experimental set up is used: perceivers are shown a posed, caricatured face or body, or listen to a posed, caricatured vocalization, and then are provided with a small set of emotion words from which they choose the correct label. This is called force-choice responding (with samples from remote cultures, sometimes perceivers hear a brief story about an emotion, and then are presented with two or three stimuli from which they must choose the correct match). Hundreds of experiments using this method have been performed on perceivers sampled from a range of cultures and they generally find that perceivers choose the correct response more often than chance, leading to claims that emotions are universally "recognized" (for a meta-analytic review of emotion perception in faces, see Elfenbein & Ambady, 2002). But remove the conceptual context (ie, the words and stories) and ask participants to freely label the emotion cue, or interfere with a perceiver's ability to access emotion concept knowledge during the task, and agreement rates drop precipitously (for a review, see Barrett, 2011a, 2011b). In fact, upon close inspection, it becomes obvious that emotion perception is highly influenced by context (Barrett, Mesquita, & Gendron, 2011; Gendron, Mesquita, & Barrett, 2013). Most importantly, of those participants who have little familiarity with Western cultural practices and norms, only those who are inadvertently taught US emotion concepts during the course of the perception task and who are provided with a small selection of emotion words to choose from appear to perceive emotions in a universal (ie, American) way; without conceptual support within the task, there is little, if any, evidence that emotion perceptions are universal (see Gendron, Roberson, & Barrett, 2015; Gendron et al., 2014a, 2014b).

Here is the takeaway point: if you adopt the classical view to guide study design and emotion measurement, you are accepting a set of assumptions that are thus far unvalidated. Much of the confusion in the science of emotion could be avoided if scientists properly evaluated their assumptions (eg, discover, rather than stipulate, how people move their faces and bodies, how vocal acoustics change, and how autonomic reactivity fluctuates during naturalistic instances of an emotion concept in a range of contexts). In the meantime, it is crucial to be cautious about studying and measuring emotions as if they have essences when it remains unclear if they do. I find it ironic that most studies in the science of emotion are designed to induce and observe only the most stereotypical examples of emotion, yet those studies routinely produce evidence of substantial variation in facial movements, autonomic patterns, and brain activity, well beyond what would be expected by error. So imagine what the science of emotion would be like if our starting assumption was that variation is the norm, and we attempted to measure and capture that variation, both within a person across contexts as well as across individuals both within and across cultures. And, in fact, an increasing number of reports are, in fact, doing just that (Ceulemans, Kuppens, & Van Mechelen, 2012; Hortensius, Schutter, & Harmon-Jones, 2011; Kuppens, Van Mechelen, & Rijmen, 2008; Kuppens, Van Mechelen, Smits, De Boeck, & Ceulemans, 2007; Nezlek, Vansteelandt, Van Mechelen, & Kuppens, 2008; Stemmler, Aue, & Wacker, 2007), but more are desperately needed.

Those who adhere to the classical view have a standard approach to solving the ever-widening chasm between their assumptions and the scientific evidence. The response to variation is usually to create more fine-grained typologies, in an attempt to bring nature under control and make it easier to identify emotion essences. Maybe there are "primary" and "secondary" emotions? Maybe there are "basic" and "non-basic" emotions? Maybe there are "non-social" and "social" emotions? Scientists sometimes try other typological divisions, like distinguishing emotion "elicitation" from "emotion regulation," and when that does not do the trick, they distinguish "steps" and "sequences" in their emotion elicitation phase of things. They distinguish "affect" from "value," and then cleave "value" into different types of "valuation." And so on. But there is another approach to emotion, one that accounts for all the empirical evidence (both the evidence that supports the classical view and that which does not). This is an essence-free view that considers typological thinking as a lot of misplaced creativity and effort.

4 The construction approach to emotion

Throughout the ages, a second approach to understanding the nature of emotion has been called by many different names, but can be united and collectively referred to as the *construction approach* to emotion (cf. Gendron & Barrett, 2009; Lindquist & Barrett, 2012). Constructionist approaches come in three flavors: social construction (how emotions are influenced by social roles and values), psychological construction (how emotions emerge from more basic psychological processes related to making meaning of affective feelings; see Fig. 2.1), and neuroconstruction (experience wires a brain for emotion during brain development). Only recently have these three varieties

of construction been integrated into one *theory of constructed emotion* (Barrett, 2013, in press; Barrett et al., 2015); in prior research, this approach was called the "conceptual act theory" (Barrett, 2006a, 2006b).

In the construction approach, an emotion is not a distinct "faculty" with its own distinct mechanism. Instead, one key hypothesis that unites all constructionist theories is that an emotion word, such as happiness, refers to a *population* of highly variable instances, each of which is tailored to a specific situation or context (Barrett et al., 2015). So an emotion is not an entity with firm boundaries in nature—it is a category of instances. Instances within a category vary because each one is tailored to the environment, that is, there are no Platonic emotion essences or fingerprints. For example, an instance of happiness can be pleasant and arousing (eg, you are finishing a challenging task with no errors and hear applause), pleasant and quiescent (eg, you feel comfortable and rested after a good night's sleep), and even unpleasant (eg, 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 (eg, you might laugh, smile, cry, jump, sigh, shout, slam your fist against a table, and so on)—whatever has been most functional for you in past, similar situations. And your cardiovascular response 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 (eg, Lindquist et al., 2013). Thus, in the construction approach, an emotion does not have an essence. Emotion categories are not natural kinds (Barrett, 2006a, 2006b). Without essences, it is not meaningful to ask what is, and what is not, an emotion. Nor do you "have" 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, often including other creatures who each have a nervous system. In a sense, you can think about emotions as tools, born of the social reality we create, to influence and regulate one another's nervous systems (Barrett, 2012).

A second key hypothesis uniting different constructionist theories is that instances within the same emotion category (eg, happiness), instances across emotion categories (eg, happiness vs fear), and even instances of nonemotion categories such as thoughts, beliefs, perceptions, and so on, all emerge from the dynamic interactions of more fundamental *common or domain general processes* within the nervous system (Barrett & Satpute, 2013; Lindquist & Barrett, 2012). No emotion category (or, in fact, any mental category) is assumed to have a Lockean essence. No emotion category is presumed to be any more biologically basic than any other. The validity of constructionist approaches depends on specifying the shared causal mechanisms, not on finding a single mechanism for each emotion, or a stable pattern of observable consequences stemming from those mechanisms (in the classical view, there is more emphasis on finding fingerprints than on identifying a single causal mechanism, because fingerprints should always be measurable, whereas essences can be hidden). Because constructionist theories rely on population thinking and domain-general mechanisms rather than essentialism, variability is *assumed* to be the norm, rather than a nuisance to be explained after the fact.

Because constructionist theories are much less intuitive, and because they are highly heterogeneous, it usually helps to focus in more closely on one theory to more fully understand its implications for the measurement of emotion. Here, I will focus my own theory of constructed emotion.

4.1 The theory of constructed emotion

According to the theory of constructed emotion, an instance of an emotion category (and any other mental category) emerges as the brain makes meaning of incoming sensory inputs from the body and the world. Every waking moment of your life, your brain is taking in constantly changing, noisy 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 inputs from your blood rushing, your muscles stretching, your lungs expanding, and so on; the autonomic nervous system, endocrine system, and immune system all create sensory changes within what scientists call the internal milieu of your body, and you 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. This means using past experience, organized as concepts, to explain what caused the sensations and what to do about them (ie, how to act). Here is a succinct summary of the theory:

> *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, have a look at Fig. 2.3.

If you are like most people who have never seen the image in Fig. 2.3 before, 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 (Fig. A.1), and then return to this page.

After viewing the appendix (Fig. A.1), most people now see an object in Fig. 2.3. So what does this exercise demonstrate? Your brain 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 Fig. 2.3. 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 Fig. 2.3 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 actually aren't present on the page, linking

Figure 2.3 An example of categorizing to construct an experience.

the blobs into the shape of a cow whose image isn't physically there on the page. Scientists call this "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 experienced 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 this works, not because the images "trigger" reactions, but because people are simulating the action that they would make in that situation, as well as the interoceptive sensations (from the core of the body) that they would receive. 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 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 chapter, 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 not only the neurons for vision, audition, touch, taste, and smell, but also for interoception, because from the brain's perspective, the body is part of the brain's world (since the body holds the brain); it's also likely true for affect, which is the low dimensional experience of interoceptive sensations. 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 not only making sensory predictions—it is also making 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 (ie, predicted motor changes produce sensory predictions) so that, in a sense, sensation follows (and is dependent on) action (Barrett, in press; Barrett & Simmons, 2015; Chanes & Barrett, 2016; Clark, 2013; Hohwy, 2013; Friston, 2010).

When your brain creates a prediction from past experience, it does not issue one neural pattern, but an entire population of potential predictions, each one having some probability (computed with Bayesian priors) of being the best fit to the current circumstances (Barrett, in press). 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 of predictions is a *concept*, constructed as you need it, on the fly (what Barsalou and colleagues call an "ad hoc" concept; 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 (ie, they categorize your sensory inputs). So, constructing meaning by correctly anticipating (predicting and adjusting to) incoming sensations is what I mean when I say that the brain is using emotion concepts to categorizing sensations to construct an instance of emotion. Sensations are conceptualized (ie, categorized) so that they are (1) actionable in a situated way and therefore (2) 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 the outside world. When past experiences of emotion (eg, happiness) are used to categorize the predicted sensory array and guide action, then an experience 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 the neuroscientist Edelman (1998) referred to experience as "the remembered present."

To see how this works, lets do a thought experiment. For example, in the past, you might have experienced the comfort of dozing on a hammock on a sunny day, the comfort of hugging a long-lost friend, the comfort of eating a piece of chocolate cake, the comfort of a warm bath, the comfort of flying on an airplane, and the comfort of reclining on a sofa in front of a fire. Each instance of comfort is not identical to every other, and when the brain creates an ad hoc concept for comfort to predict incoming sensory inputs, it constructs simulations (as potential actions and perceptions) of those instances that are most similar to the current situation (each prediction having some probability of being correct, given past experience). So the brain simulates an on-line concept of comfort, not in absolute terms, but with reference to your particular goal in the moment (eg, to relax and minimize stress, to feel close to others through shared comfort, and so on). This means that an emotion word like "comfort" or "happiness" has a specific meaning, but its specific meaning can change from one instance to the next (Barrett, in press).

The theory of constructed emotion is an evolutionary theory, but it does not hypothesize that emotions are universal. Unlike the classical view, which takes its inspiration from the more essentialist *The expression of the emotions in man and animals* (Darwin, 1872/1965), the theory of constructed emotion uses conceptual innovations found in Darwin's *On the origin of species* (cf. Barrett, 2013). In fact, Darwin is credited with vanquishing essentialism in biology in *Origin* (Mayr, 2004), so it is ironic that he went on to write a highly essentialized treatment of emotion slightly more than a decade later (for a hypothesis of why Darwin did this, see Barrett, in press; also see Fridlund, 1992). The theory of constructed emotion's use of population thinking comes from Darwin's *Origin* (eg, a species is not a natural kind category with one, fixed Platonic form serving as its essence, where variation around this form is error; instead, a species is a conceptual category populated with unique individuals who have differing degrees of fit to the environment). The theory of constructed emotion also incorporates Darwin's focus on holism, or the need to study a part in the context of the whole that influences it. Intrinsic to holism is the importance of studying a phenomenon in context, rather than attempting to find general, context-free laws (as is typical in certain forms of reductionism; Mayr, 2004).

The theory also incorporates several other concepts from biology, the most important being that there is more than one cause to produce the same phenomenon, called degeneracy (Edelman & Gally, 2001; Marder & Taylor, 2011; Tononi, Sporns, & Edelman, 1999). Degeneracy 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 intrinsic 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). Degeneracy refers to the capacity for structurally dissimilar systems or processes to give rise to identical outcomes (Edelman & Gally, 2001), such as many different facial configurations, autonomic configurations, or brain states mapping to the same emotion category.

Figure 2.4 The measurement model for theory of constructed emotion. This is a measurement model, but not a causal model, of emotion (ie, this figure does not depict the mechanisms that cause emotion, but only how measures might configure to assess an instance of emotion) (for a fuller explanation, see Barrett, 2011a, 2011b).

4.2 Measurement implications

The measurement model implied by the theory of constructed emotion is difficult to draw, because it involves tracking a high-dimensional brain and body state dynamically over time in an iterative way. A poor approximation is depicted in Fig. 2.4, although strictly speaking, this figure has limitations for modeling the theory of constructed emotion (for a discussion, see Barrett, 2011b).

In this measurement model, assessments of facial movements, autonomic reactivity, and vocal acoustics will not necessarily be correlated, and therefore carry unique information about an emotional episode (ie, they are not necessarily interchangeable from an information theory standpoint). (This measurement model is useful, because in reality, measures from different modalities are rarely correlated with one another.) One measure cannot stand in for another, so that *optimal* measurement of emotion requires a multimodal approach. The experience of an emotion can change from one instance to the next, and therefore cannot be stipulated in advance; instead, the relationships between measures must be inductively discovered within a person across contexts. From this standpoint, it will never be possible to properly measure emotion by merely measuring the face, or skin conductance, or any single or set of physical measures alone.

In the theory of constructed emotion, without a multimodal measurement approach, it is perhaps best to measure emotion via self-report, because there is no "objective" way of determining when someone is, or is not, in a particular emotional state (Barrett, 2006b). Furthermore, to date, we have no ability to model emergence mathematically when it comes to emotion (where the product of dynamically interacting systems has properties that the systems themselves do not). Self-reports, on their own, have limitations, of course, because they only capture some instances of emotion (those of which the respondent is aware); in this view, as is the classical view, people may not be conscious of the emotion they have constructed (although for an entirely different set of theoretical hypotheses). But when self-reports do not correlate with more objective measures, the self-reports are *not* necessarily assumed to be inaccurate (more on this below). Moreover, because emotion concepts are integral to the construction of emotional experiences and perceptions, words and other symbols that prime emotion concepts (ie, that launch predictions and simulations) will influence what is experienced and felt. As a consequence, self-reports of emotional experience

are influenced by the words that we give respondents to communicate their experiences or perceptions. It is possible to change a person's feeling merely by the type of measurement instrument you give them. And respondents will use the measure you give them to report what they want to tell you, which may not necessarily match what you are asking (eg, if a respondent feels excited, but you ask if he/she is happy, the respondent will use the item to tell you how excited he/she is).

In the theory of constructed emotion, emotions are not universal (meaning studying all aspects of emotion, including emotion concepts, is crucial across cultures). Nor are emotions assumed to be homologous in other animals (meaning that studying emotion in nonhuman animals will *not* reveal all the secrets of emotions in humans, although such study is undoubtably crucial to learn about some ingredients of emotion.). Humans and other animals are hypothesized to share some species-general core systems, and some species-specific core systems. Therefore, classical conditioning with an aversive cue (like an electric shock) is not "fear learning" but "threat learning" (for a discussion see Barrett, 2012; LeDoux, 2014). Reward processing in rodents and monkeys might be the same as in humans in some ways, but it might also be different because the brain did not evolve according to a phylogenetic scale like sedimentary rock; the cortex did not evolve on top of preserved subcortical regions like icing an already-baked cake; as brains grow, they expand and reorganize (Barrett et al., 2007; Striedter, 2005).

Another implication is that questions about "recognition accuracy" are not scientific, because emotions have no perceiver-dependent essences with which to compare to a perceiver-based judgment (whether a perception of someone else's emotion or a self-report of experience). Instead, what we actually measure is consensus (Do you and I agree on the emotion you are feeling? Do you and I agree on the emotion that some other person is feeling? Do our perceptions agree with the cultural norms for emotion in this specific situation?). The reliance on consensus (or agreement) is not a bug—it is a feature that reflects the status of emotions as social (not biological) kinds (Barrett, 2012).

Emotions are not assumed to be perceiver-independent phenomena, waiting to be discovered in nature by a human mind. They are instances that are created within a human mind, in concert with other human minds. Emotions depend on the human mind for existence—they are perceiver-dependent phenomena. Thus, emotions are made, not found. They are perceived, not detected. And measuring emotion *requires measuring human experience and perception*, as well as "objective" measures, such as facial muscle movements, cardiovascular reactivity, and so on. Measuring emotion means capturing when and how mere physical changes are categorized so as to serve the psychological functions of an emotion (as opposed to other times when the same physical changes are not understood as an emotion and therefore serve some other psychological function; Barrett, 2012).

4.3 Drawbacks to using the construction approach's measurement model

Complexity and cost, both in terms of time and money, are the primary drawbacks to using the measurement model of the theory of constructed emotion. The more

observable aspects of emotion (facial movements, autonomic responses, etc.) cannot be used to stand in for or validate a person's own experience, at least with the methods that we have currently available. Validation requires measuring the underlying mechanisms that create an instance of emotion, and optimally this requires developing a more suitable epistemological approach (ie, following individuals in context over time) rather than anchoring and adjusting way from an approach that was developed to evaluate the classical view.

In the meantime, lack of correspondence between verbal reports and behavior does not necessarily indicate that the verbal reports are invalid. In fact, because instances of emotion are thought to emerge from more basic processes, the instances will have features that are not reflected in measurements of the individual processes themselves. Thus, for now, verbal report, even with all of its failings, may be the best means of assessing the experience of emotion in a quick and efficient way. If you want to know whether people are experiencing an emotion, you have to ask them (cf. Barrett, 2006a). This may sound easy, but there are a few landmines here as well. You can't assume that a feeling of happiness in one context is similar to the feeling of happiness in another (Wilson-Mendenhall, Barrett, Simmons & Barsalou, 2011; Wilson-Mendenhall et al., 2013). You also can't necessarily assume that two people mean the same thing by happiness, as people differ in emotional granularity; for some people, the word "happiness" refers to a specific feeling state, whereas for others, it refers to a general, pleasant feeling (eg, Barrett, 2004; Barrett & Bliss-Moreau, 2009). It is also not possible to assume that an emotion word means the same thing in different cultures (eg, Barrett, in press; Pavlenko, 2014). In fact, a construction mindset helps explain why certain emotion categories exist in some cultures, but not in others, and that what counts as an emotion in some cultures is not an emotion in others (for a discussion, see Barrett, in press; Pavlenko, 2014; Russell, 1991). The implication is that it is always a good idea to include a measure of emotion concepts whenever measuring the experience of emotion.

5 A cautionary note: Beware of lurking essentialism

Whether you rely on the assumption of emotion essences or not, it is important to be vigilant for the use of an essentialist mindset where it does not belong, lest you misinterpret your own (or someone else's) findings. A good example of such misinterpretation can be found in the increasingly frequent use of pattern classification approaches to distinguish the instances of one emotion category from another. Pattern classification techniques refer to a family of statistical methods designed to categorize data by learning from existing categories or grouping variables to make predictions about and assign membership to new instances. These techniques are being used with measures of autonomic physiology, facial movements, and changes in neural response measured as blood oxygenation level-dependent (BOLD) signal within

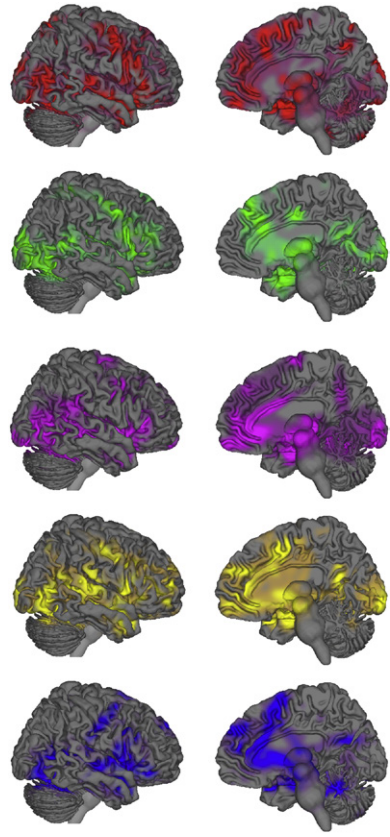


Figure 2.5 Intensity maps for each of the five emotion categories examined by Wager et al. (2015). Classification rates were Anger (red) = 43%, Disgust (green) = 76%, Fear (pink) = 86%, Happiness (yellow) = 58%, and Sadness (blue) = 65%.

the brain (that is divided up into three-dimensional cubes called voxels) (Kassam, Markey, Cherkassky, Loewenstein, & Just, 2013; Kragel & Labar, 2013, 2015; Park, Jang, Chung, & Kim, 2013; Rainville, Bechara, Naqvi, & Damasio, 2006; Saarimäki et al., 2015; Yuen et al., 2012). Researchers train a classifier using some set of measurements for known instances of emotion categories, and then use the classifier to diagnose new instances of those categories using similar measurements. For example, Fig. 2.5 presents the multivoxel patterns that successfully classified five categories of emotion above chance in our recent meta-analytic paper (Wager et al., 2015). We trained classifiers on brain maps from existing studies of anger, sadness, fear, disgust, and happiness, and then used the classifiers to diagnose the emotion being represented in brain maps from new studies.

The lure of essentialism leads scientists to claim that the patterns are something like neural essence for each emotion category. For example, Saarimäki et al. (2015) claimed to find "fingerprints" and "signatures" for certain emotion categories, while Kragel and LaBar (2015) claimed to find their "biomarkers," where a biomarker is a measurable indicator of some category, such that its presence in an instance indicates that the instance belongs to a particular category (Strimbu & Tavel, 2010).[5] In medicine, a biomarker is a measurable substance that is present in all members of a category. Biomarkers must be sensitive and unique to a specific category to work properly. Both Saarimäki et al. (2015) and Kragel and LaBar (2015) interpreted their findings as support for the classical view of emotion, when, in fact, they actually found evidence in support of the theory of constructed emotion (Clark-Polner et al., in press-b).

Patterns that successfully distinguish one emotion category from another are not emotion essences. They are not biomarkers, fingerprints, or signatures of an emotion category. The implication in using terms like "biomarker," "fingerprint," and "signature" is that the pattern for an emotion category is it's brain state—the elements of the pattern (eg, the pattern of voxels) are assumed to be both unique to a single emotion category and unchanging across its instances. Yet, a pattern derived from pattern classification techniques does not appear in *every instance* (or in fact, in *any instance*) of a category, even when the pattern can classify a category's instances with 100% accuracy. We have demonstrated this with simulations (Clark-Polner, Johnson, & Barrett, in press-a). Patterns should be understood from the standpoint of population thinking—the pattern is an abstract, statistical summary of a category's instances; the pattern does not (and need not) exist in nature to work well. Although as a group, the instances of any emotion category can be diagnosed with a pattern, the pattern itself is an abstraction. Similarly, the average middle-class US family has 3.13 children, but this is an abstract representation, because no family actually has 3.13 children. To assume that a pattern is the fingerprint or biomarker for an emotion category is to mistake a statistical summary for the norm. The take-away point is that successful pattern classification using *any* type of measures provides evidence that emotion categories, similar to biological categories, are conceptual categories populated by unique and highly variable instances that do not share any necessary features.

Furthermore, a quick review of some recent neuroimaging findings (eg, Kassam et al., 2013; Kragel & Labar, 2015; Saarimäki et al., 2015; Wager et al., 2015) indicates that patterns distinguishing one emotion category from another in one study do not replicate in another study; the same is true for studies that successfully created patterns of autonomic physiology, even when two studies use the same stimuli, the same experimental method, and sample participants from the same population (Kragel & LaBar, 2013; Stephens et al., 2010). Once again, this is evidence that variation is the norm.

[5] Kragel and LaBar (2015) are careful to point out "it is unlikely that the patterns we identified perfectly capture the essence of an emotion, but are better characterized as an amalgamation of the components which make emotions unique" (p. 1446).

6 Conclusions

Essentialism is not necessarily a bad thing. Utilizing it unknowingly is, however. The goal of this chapter is to allow you to view the current science of emotion as a continuation of the long-standing debate over whether mental categories are carved into nature by essences, or whether they are more flexible groupings of highly variable and situated instances, created from more basic mechanisms (for a discussion, see Lindquist et al., 2012). Some scientists consider essentialism a useful strategy for scientific inquiry because they believe that it mirrors the structure of the real world (ie, they believe the world is full of natural kind categories) (eg, Bloom, 2000; Kornblith, 1993; Pinker, 1997). Others, however, believe that essentialism is a particularly poor strategy for scientific inquiry (Lewontin, 2000). You can make your decision. Just do it explicitly, and with an appreciation of the consequences.

If you are going to use the classical view to guide your measurement of emotion, you should do so with caution, realizing that the bulk of the scientific evidence does not yet support it. This leaves you vulnerable to investing a lot of time and money in an enterprise that might seem deeply intuitive, but that might deliver very little in return. Alternatively, for those of us who relinquish essentialism as a guiding assumption for the nature and measurement of emotion, we should refrain from feeling smug or superior. Beware the lesson of Charles Darwin, who became famous in one field (biology) by vanquishing essentialism, while becoming famous (psychology) in another by relying on it.

It is worth pointing out that the history of science can be read as a long, slow march away from essentialist thinking, discovering that universal laws are actually contextual (eg, in physics, with the discover of relativity theory and then quantum mechanics) and discovering that variation is meaningful and is not error (eg, in biology, with Darwin's *On the origin of species*, and then again a century later with the study of epigenetics and genomics). Construction approaches have also emerged in psychology, precisely when it is discovered that instances grouped together as the same phenomenon do not share sufficient organizational coherence to be explained by a common mechanism (eg, recognizing variation is not always error and is, in fact, meaningful, such that instances of the same category are caused by different mechanisms). More recently, it has been discovered that instances which have been designated as different phenomena (eg, "stress," "emotion," and "memory") in fact arise from the same mechanisms, revealing shared regularities across these instances that had been thus far ignored (ie, recognizing similarities across categories).

Essentialism is a habit of the human mind that is difficult to vanquish, particularly when it comes to thinking about emotion. The very enterprise of measuring emotion tempts us with essentialism. As William James put it, "Whenever we have made a word… to denote a certain group of phenomena, we are prone to suppose a substantive entity existing beyond the phenomena, of which the word shall be the name" (James, 1890, p. 195). Essentialism is also difficult to vanquish because when people group instances together and treat them as similar, they are often unaware of their own goals in the process. As a consequence, they come to mistakenly believe that the similarity they perceive is real in nature, and the variation is error in a more universal

sense, rather than merely not useful for meeting a particular goal in some particular context (but may very well be useful for some other goal in some other context).

Categorization is 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. But it's possible to use categories, and measure them, without reifying them with essences.

Appendix A

The last several thousand years of scholarly writing on the nature of emotion can be understood as an ongoing debate between essentialism and construction (of one sort or another). Although a comprehensive history of this literature is beyond the scope of this chapter, a few touch points are instructive for the uninitiated. In the Western scholarly tradition, essentialist approaches to emotion include Plato and Aristotle in Ancient Greece; Descartes (1649/1989) during the Enlightenment, with a theory that foreshadowed Carl Lange (1885/1922) and Damasio (Damasio & Carvalho, 2013); Gall, of phrenology fame (Zola-Morgan, 1995); Irons (1897a, 1897b) and Dewey (1895) who were the first modern classical appraisal theorists; Darwin (who vanquished essentialism in biology with *On the origin of species* but then went on to write a highly essentialized book on emotions a decade later); Carl Lange (1885/1922), who crafted a modern basic emotion theory which Dewey (1895) then tattooed on to William James's (1890) constructionist theory (mangling James's theory to create the James–Lange theory and creating a misunderstanding that has survived to this day;[6] McDougall (1923) with a theory very similar to Panksepp's basic emotion theory; Panksepp, 1998); Allport (1922, 1924) who invented the facial feedback hypothesis; Cannon (1927) who localized emotion to a specific brain region; Papez (1937) who expanded the region into a circuit; and MacLean (1949, 1990) who created an elixir of Plato's tripartite mind (rational thoughts, passions (which

[6] Notice that William James was a constructionist. He wrote, for example, that "Surely there is no definite affection of 'anger' in an 'entitative' sense" (1894/1994, p. 206), believing instead that each instance of emotion had its own associated physical state. James believed that believing in emotion essences was the psychologist's fallacy. The "trouble with emotions in psychology" he wrote, is that they are regarded too much as absolutely individual things. … But if we regard them as products of more general causes (as "species" are now regarded as products of heredity and variation), the mere distinguishing and cataloguing becomes of subsidiary importance (James, 1890/1998, p. 449). Instead, James prescribed an entirely constructionist approach for the study of psychology, including emotion. "A science of the relations of mind and brain" James wrote, "must show how the elementary ingredients of the former correspond to the elementary functions of the latter" (1890/1998, p. 28). Although it is beyond the scope of this chapter, it is fascinating how James's constructionist theory of emotion was transformed (by ignoring the constructionist features) to integrate it with Lange's highly essentialized vasomotor theory of emotion was strongly essentialistic (each emotion had its own bodily essence), creating the James-Lange theory (first named by Dewey, 1895). So the James–Lange theory is based on a profound misunderstanding of William James.

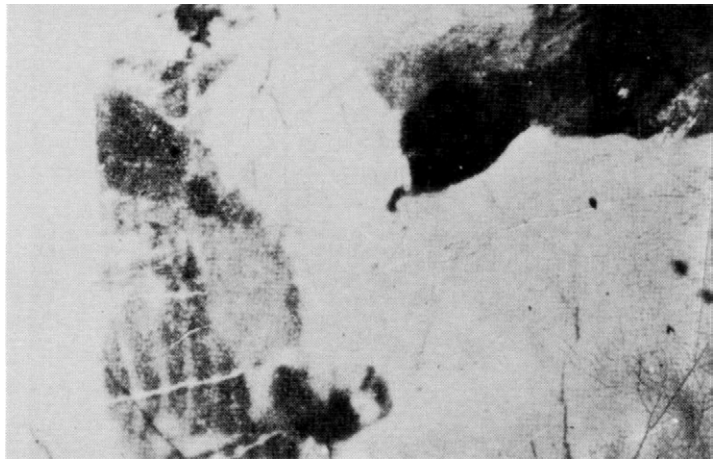


Figure A.1 An example of categorizing to construct and experience.

today we would call emotions), and appetites like hunger and sex drive), Aristotle's phylogenetic scale (scala naturae) idea, and Darwin's ideas from *The descent of man* (1871), wrapped them in neuroanatomy to create the triune brain concept which is still popular today (Panksepp, 1998) (for a review, see Gendron & Barrett, 2009). In the traditional Buddhist view (the Abhidharma school) dating back to the 2nd century BCE, a mind is created from a set of universal, discrete 82 psychophysical elements called *dharmas*, 52 of which are mental; several of the mental dharmas bear a striking resemblance to certain emotions that are considered to be biologically basic (anger, pride, joy, shame, regret, jealousy).

Historically, construction is more varied and difficult to summarize, but key ideas can be found in the writings of Heraclitus in Ancient Greece, Ibn al Haythan (in the middle ages), Locke and Kant during the Enlightenment, Spencer (1855), James (1890, 1894), and Wundt (1897/1998) during the 19th century (all of whom offered arguments against faculty psychology), as well as Duffy (1934a, 1934b, 1941), Dunlap (1932), Hunt (1941), and Harlow and Stanger (1932) in the first half of the 20th century (all of whom observed that emotions had no essences, and therefore must be constructed as a person makes sense of autonomic changes, although no specific mechanisms or processes were offered). In the past, constructionist ideas on the nature of emotion were often nascent, embedded in broader critiques of classical view. More recently, a new generation of psychological construction theories have emerged, articulating a more detailed and nuanced scientific agenda for the study of emotion (Barrett & Russell, 2015a, 2015b; also see Lane & Schwartz, 1987; LeDoux, 2012, 2015; Olsson & Ochsner, 2008; Roy, Shohamy & Wager, 2012; Seth, 2013). Buddhist philosophy also has a more constructionist version of the mind. A range of influential thinkers associated with the *Sautrāntika, Madhyamika*, and *Yogācāra*

schools (cf. 3rd century CE through the 7th century CE), the most well-known of which was named Dharmakirti in the 7th century CE who suggested that the dharmas are not basic elements of the human mind, but are themselves creations of that mind, emerging as a function of human concepts (Dreyfus & Thompson, 2007).

References

Allport, F. H. (1922). A physiological-genetic theory of feeling and emotion. *Psychological Review*, *29*, 132–139.

Allport, F. H. (1924). *Social psychology*. New York: Houghton Mifflin.

Anderson, D. J., & Adolphs, R. (2014). A framework for studying emotions across species. *Cell*, *157*(1), 187–200.

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

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

Barrett, L. F. (2000). *Modeling emotion as an emergent phenomenon: A causal indicator analysis*. Paper presented at the annual meeting of the Society for Personality and Social Psychology, Nashville, TN. Retrieved from: <http://www.affective-science.org/publications.shtml>.

Barrett, L. F. (2004). Feelings or words? Understanding the content in self-report ratings of emotional experience. *Journal of Personality and Social Psychology*, *87*, 266–281.

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. (2009). The future of psychology: Connecting mind to brain. *Perspectives in Psychological Science*, *4*, 326–339.

Barrett, L. F. (2011a). Was Darwin wrong about emotional expressions? *Current Directions in Psychological Science*, *20*, 400–406.

Barrett, L. F. (2011b). Bridging token identity theory and supervenience theory through psychological construction. *Psychological Inquiry*, *22*, 115–127.

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. (in press). *How emotions are made: The new science of the mind and brain*. New York: Houghton Mifflin Harcourt.

Barrett, L. F., & Bliss-Moreau, E. (2009). Affect as a psychological primitive. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 41, pp. 167–218). Burlington, VA: Academic Press.

Barrett, L. F., & Russell, J. A. (1999). The structure of current affect - controversies and emerging consensus. *Current Directions in Psychological Science*, *8*, 10–14.

Barrett, L. F., & Russell, J. A. (2015a). An introduction to psychological construction. In L. F. Barrett & J. A. Russell (Eds.), *The psychological construction of emotion* (pp. 1–17). New York: Guilford.

Barrett, L. F., & Russell, J. A. (Eds.). (2015b). *The psychological construction of emotion*. New York: Guilford.

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.

Barrett, L. F., Lindquist, K., Bliss-Moreau, E., Duncan, S., Gendron, M., Mize, J., et al. (2007). Of mice and men: Natural kinds of emotion in the mammalian brain? *Perspectives on Psychological Science*, *2*, 297–312.

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, *20*, 286–290.

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., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). The conceptual act theory: A road map. In L. F. Barrett & J. A. Russell (Eds.), *The psychological construction of emotion* (pp. 83–110). New York: Guilford.

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

Barsalou, L. W. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes*, *18*, 513–562.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *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.

Buck, R. (1999). The biological affects: A typology. *Psychological Review*, *106*, 301–336.

Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., & Ito, T. A. (2000). The psychophysiology of emotion. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (Vol. 2, pp. 173–191). New York: Guilford.

Cannon, W. B. (1927). The James-Lange theory of emotions: A critical examination and an alternative theory. *American Journal of Psychology*, *39*, 106–124.

Ceulemans, E., Kuppens, P., & Mechelen, I. V. (2012). Capturing the structure of distinct types of individual differences in the situation-specific experience of emotions: The case of anger. *European Journal of Personality*, *26*(5), 484–495.

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-a). Multivoxel pattern analysis does not provide evidence to support the existence of basic emotions. *Cerebral Cortex*.

Clark-Polner, E., Wager, T. D., Satpute, A. B., & Barrett, L. F. (in press-b). "Brain-based fingerprints?" Variation, and the search for neural essences in the science of emotion. In L. F. Barrett, M. Lewis, & J. M. Haviland-Jones (Eds.), *The handbook of emotion* (4th ed.). New York: Guilford.

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: 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* (pp. 628–642) (3rd ed.). New York: Guilford Press.

Coan, J. A. (2010). Emergent ghosts of the emotion machine. *Emotion Review*, *2*, 274–285.

Damasio, A., & Carvalho, G. B. (2013). The nature of feelings: Evolutionary and neurobiological origins. *Nature Review Neuroscience*, *14*, 143–152.

Damasio, A. R. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. New York: Harcourt Brace.

Darwin, C. (1965). *The expression of the emotions in man and animals*. Chicago, IL: University of Chicago Press. (Original work published 1872).

Davis, M. (1992). The role of the amygdala in fear and anxiety. *Annual Review of Neuroscience*, *15*(1), 353–375.

Descartes, R. (1649/1989). *The passions of the soul* (Stephen H. Voss, trans). Indianapolis: Hackett.

Dewey, J. (1895). The theory of emotion. *Psychological Review*, *2*(1), 13.

Dreyfus, G., & Thompson, E. (2007). Asian perspectives: Indian theories of mind. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *Cambridge handbook of consciousness* (pp. 88–114). Cambridge: Cambridge University Press.

Duffy, E. (1934a). Is emotion a mere term of convenience? *Psychological Review*, *41*, 103–104.

Duffy, E. (1934b). Emotion: An example of the need for reorientation in psychology. *Psychological Review*, *41*, 184–198.

Duffy, E. (1941). An explanation of "emotional" phenomena without use of the concept "emotion". *The General Journal of Psychology*, *25*, 283–293.

Dunlap, K. (1932). Are emotions teleological constructs? *The American Journal of Psychology*, *44*, 572–576.

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

Edelman, G. M., & Gally, J. A. (2001). Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences*, *98*(24), 13763–13768.

Ekman, P. (1972). Universal and cultural differences in facial expressions of emotions. In J. K. Cole (Ed.), *Nebraska symposium on motivation, 1971* (pp. 207–283). Lincoln, NE: University of Nebraska Press.

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

Ekman, P., & Friesen, W. V. (1978). *Manual for the facial action coding system*. Consulting Psychologists Press.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological Bulletin*, *128*(2), 203–235.

Fridlund, A. J. (1992). Darwin's anti-darwinism and the expression of the emotions in man and animals. In K. T. Strongman (Ed.), *International review of emotion* (Vol. 2, pp. 117–137). New York: Wiley.

Friedman, B. H. (2010). Feelings and the body: The Jamesian perspective on autonomic specificity of emotion. *Biological Psychology*, *84*(3), 383–393. http://dx.doi.org/10.1016/j.biopsycho.2009.10.006.

Frijda, N. H. (1986). *The emotions*. New York: 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. Sinartra (Eds.), *Evolution challenges integrating research and practice in teaching and learning about evolution* (pp. 3–21). New York: 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.
Gendron, M., Mesquita, B., & Barrett, L. F. (2013). Emotion perception. In D. Reisberg (Ed.), *Oxford handbook of cognitive psychology* (pp. 539–556). New York: Oxford University Press.
Gendron, M., Roberson, D., & Barrett, L. F. (2015). Cultural variation in emotion perception is real: A response to Sauter et al. *Psychological Science*, *26*, 357–359.
Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014a). Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. *Emotion*, *14*(2), 251.
Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014b). Cultural relativity in perceiving emotion from vocalizations. *Psychological Science*, *25*(4), 911–920.
Greenwald, A. G., & Ronis, D. L. (1981). On the conceptual disconfirmation of theories. *Personality and Social Psychology Bulletin*, *7*(1), 131–137.
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*(1), 8–16.
Guillory, S. A., & Bujarski, K. A. (2014). Exploring emotion using invasive methods: Review of 60 years of human intracranial electrophysiology. *Social Cognitive and Affective Neuroscience*, *9*, 1880–1889.
Harlow, H. F., & Stagner, R. (1932). Psychology of feelings and emotions. I. Theory of feelings. *Psychological Review*, *39*(6), 570–589.
Harlow, H. F., & Stagner, R. (1933). Psychology of feelings and emotions. II. Theory of emotions. *Psychological Review*, *40*, 184–195.
Hohwy, J. (2013). *The predictive mind*. Oxford: Oxford University Press.
Hortensius, R., Schutter, D. J., & Harmon-Jones, E. (2011). When anger leads to aggression: Induction of relative left frontal cortical activity with transcranial direct current stimulation increases the anger–aggression relationship. *Social Cognitive and Affective Neuroscience*, *7*(3), 342–347.
Hunt, W. A. (1941). Recent developments in the field of emotion. *Psychological Bulletin*, *38*(5), 249–276.
Irons, D. (1897a). The nature of emotion. *The Philosophical Review*, *6*, 242–256.
Irons, D. (1897b). The primary emotions. *The Philosophical Review*, *6*, 626–645.
Izard, C. E. (1993). Four systems for emotion activation: Cognitive and noncognitive processes. *Psychological Review*, *100*, 68–90.
Izard, C. E. (2011). Forms and functions of emotions: Matters of emotion–cognition interactions. *Emotion Review*, *3*, 371–378.
James, W. (1884). What is an emotion? *Mind*, *9*, 188–205.
James, W. (1890). *The principles of psychology* (Vol. 1). New York: Holt.
James, W. (1994). The physical basis of emotion. *Psychological Review*, *101*, 205–210. (Original work published 1894).
James, W. (1998). *The principals of psychology*. Bristol: Thoemmes Press. (Original work published 1890).
Kassam, K. S., Markey, A. R., Cherkassky, V. L., Loewenstein, G., & Just, M. A. (2013). Identifying emotions on the basis of neural activation. *PloS One*, *8*, e66032.
Kornblith, H. (1993). *Inductive inference and its natural ground: An essay in naturalistic epistemology*. Cambridge, MA: The MIT Press.

Kragel, P. A., & LaBar, K. S. (2013). Multivariate pattern classification reveals autonomic and experiential representations of discrete emotions. *Emotion*, *13*(4), 681–690.
Kragel, P. A., & LaBar, K. S. (2015). Multivariate neural biomarkers of emotional states are categorically distinct. *Social Cognitive and Affective Neuroscience*, *10*(11), 1437–1448.
Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, *84*(3), 394.
Kuppens, P., Van Mechelen, I., & Rijmen, F. (2008). Towards disentangling sources of individual differences in appraisal and anger. *Journal of Personality*, *76*, 1–32.
Kuppens, P., Van Mechelen, I., Smits, D. J. M., De Boeck, P., & Ceulemans, E. (2007). Individual differences in patterns of appraisal and anger experience. *Cognition and Emotion*, *21*, 689–713.
Lane, R., & Schwartz, G. E. (1987). Levels of emotional awareness: A cognitive-developmental theory and its application to psychopathology. *American Journal of Psychiatry*, *144*, 133–143.
Lange, C. G. (1922). The emotions: A psychophysiological study (I. A. Haupt, trans.). In K. Dunlap (Ed.), *The emotions* (pp. 33–90). Baltimore, MD: Williams & Wilkins Company. (Original work published 1885).
Lazarus, R. S. (1966). *Psychological stress and the coping process*. New York: McGraw.
Lazarus, R. S. (1991). *Emotion and adaptation*. New York: Oxford University Press.
LeDoux, J. (2012). Rethinking the emotional brain. *Neuron*, *73*, 653–676.
LeDoux, J. (2014). Coming to terms with fear. *Proceedings of the National Academy of Sciences*, *111*(8), 2871–2878.
LeDoux, J. (2015). *Anxious: Using the brain to understand and treat fear and anxiety*. New York: Viking.
Levenson, R. W. (1994). Human emotions: A functional view. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 123–126). New York: Oxford University Press.
Leventhal, H. (1984). A perceptual-motor theory of emotion. *Advances in Experimental Social Psychology*, *17*, 117–182.
Lewis, M. D. (2005). Bridging emotion theory and neurobiology through dynamic systems modeling (target article). *Behavioral and Brain Sciences*, *28*, 169–194.
Lewontin, R. C. (2000). *The triple helix: Gene, organism, and environment*. Cambridge, MA: Harvard University Press.
Lindquist, K. A., & Barrett, L. F. (2012). A functional architecture of the human brain: Insights from emotion. *Trends in Cognitive Sciences*, *16*, 533–540.
Lindquist, K. A., Siegel, E. H., Quigley, K., & Barrett, L. F. (2013). The hundred years emotion war: Are emotions natural kinds or psychological constructions? Comment on Lench, Flores, & Bench (2011). *Psychological Bulletin*, *139*, 255–263.
Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences*, *35*, 121–143.
MacLean, P. D. (1949). Psychosomatic disease and the "visceral brain": Recent developments bearing on the Papez theory of emotion. *Psychosomatic Medicine*, *11*, 338–353.
MacLean, P. D. (1990). *The triune brain in evolution: Role in paleocerebral functions*. New York: Plenum Press.
Mandler, G. (1975). *Mind and emotion*. New York: Wiley.
Marder, E., & Taylor, A. L. (2011). Multiple models to capture the variability in biological neurons and networks. *Nature Neuroscience*, *14*(2), 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.), (3rd ed.). New York: Guilford.

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion*, *23*(2), 209–237. http://dx.doi.org/10.1080/02699930802204677.

Mayr, E. (2004). *What makes biology unique?* New York, NY: Cambridge University Press.

McDougall, W. (1921). An introduction to social psychology. Boston, MA: John W. Luce. (Original work published 1908).

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

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). New York: Cambridge University Press.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Nezlek, J. B., Vansteelandt, K., Van Mechelen, I., & Kuppens, P. (2008). Appraisal-emotion relationships in daily life. *Emotion*, *8*(1), 145–150.

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*(5), 569–587.

Olsson, A., & Ochsner, K. N. (2008). The role of social cognition in emotion. *Trends in Cognitive Sciences*, *12*, 65–71.

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: Guilford.

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, *97*(3), 315–331.

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. New York: Oxford University Press.

Papez, J. W. (1937). A proposed mechanism of emotion. *Archives of Neurology & Psychiatry*, *38*, 725–743.

Park, B. J., Jang, E. H., Chung, M. A., & Kim, S. H. (2013). Design of prototype-based emotion recognizer using physiological signals. *ETRI Journal*, *35*(5), 869–879.

Pavlenko, A. (2014). *The bilingual mind: And what it tells us about language and thought*. Cambridge: Cambridge University Press.

Pinker, S. (1997). *How the mind works*. New York, NY: W. W. Norton & Company.

Price, C. J., & Friston, K. J. (2002). Degeneracy and cognitive anatomy. *Trends in Cognitive Sciences*, *6*(10), 416–421.

Quigley, K. S., & Barrett, L. F. (2014). Is there consistency and specificity of autonomic changes during emotional episodes? Guidance from the conceptual act theory and psychophysiology. *Biological Psychology*, *98*, 82–94. http://dx.doi.org/10.1016/j.biopsycho.2013.12.013.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, *435*(7045), 1102–1107.

Rainville, P., Bechara, A., Naqvi, N., & Damasio, A. R. (2006). Basic emotions are associated with distinct patterns of cardiorespiratory activity. *International Journal of Psychophysiology*, *61*(1), 5–18.

Roseman, I. J. (1991). Appraisal determinants of discrete emotions. *Cognition and Emotion*, *5*, 161–200.

Roseman, I. J. (2011). Emotional behaviors, emotivational goals, emotion strategies: Multiple levels of organization integrate variable and consistent responses. *Emotion Review*, *3*(4), 434–443.

Roy, M., Shohamy, D., & Wager, T. D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends in Cognitive Sciences*, *16*(3), 147–156.

Russell, J. A., Bachorowski, J.-A., & Fernández-Dols, J.-M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology*, *54*, 329–349.
Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178.
Russell, J. A. (1991). Culture and the categorization of emotions. *Psychological bulletin*, *110*(3), 426–450.
Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*, 145–172.
Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*(5), 805–819.
Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I. P., Lampinen, J., Vuilleumier, P., Hari, R., et al. (2015). Discrete neural signatures of basic emotions. *Cerebral Cortex*.
Sander, D., Grandjean, D., & Scherer, K. R. (2005). A systems approach to appraisal mechanisms in emotion. *Neural networks*, *18*(4), 317–352.
Schachter, S., & Singer, J. E. (1962). Cognitive, social and physiological determinants of emotional state. *Psychological Review*, *69*, 379–399.
Scherer, K. R. (1984). Emotion as a multicomponent process: A model and some cross-cultural data. In P. Shaver (Ed.), *Review of personality and social psychology* (Vol. 5, pp. 37–63). Beverly Hills, CA: Sage.
Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 695–729.
Scherer, K. R. (2009). Emotions are emergent processes: they require a dynamic computational architecture. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1535), 3459–3474.
Siegel, E. H., Sands, M. K., Condon, P., Chang, Y., Dy, J., Quigley, K. S., et al. (under review). Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories. (Manuscript under review.)
Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17*, 565–573.
Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, *48*, 813–838.
Spencer, H. (1855). *Principals of psychology*. London: Longman, Brown, Green & Longmans.
Stemmler, G., Aue, T., & Wacker, J. (2007). Anger and fear: Separable effects of emotion and motivational direction on somatovisceral responses. *International Journal of Psychophysiology*, *66*(2), 141–153.
Stephens, C. L., Christie, I. C., & Friedman, B. H. (2010). Autonomic specificity of basic emotions: Evidence from pattern classification and cluster analysis. *Biological psychology*, *84*(3), 463–473.
Striedter, G. F. (2005). *Principles of brain evolution*. Sutherland, MA: Sinauer Associates.
Strimbu, K., & Tavel, J. A. (2010). What are biomarkers? *Current opinion in HIV and AIDS*, *5*(6), 463.
Tian, Y.-I., Kanade, T., & Cohn, J. F. (2001). Recognizing action units for facial expression analysis. *Transactions on Pattern Analysis and Machine Intelligence*, *23*, 1–19.
Tomkins, S. S. (1962). *Affect, imagery, and consciousness: Vol. 1. The positive affects*. New York: Springer.
Tomkins, S. S. (1963). *Affect, imagery, and consciousness: Vol. 2. The negative affects*. New York: Springer.

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, O. (1998). Complexity and coherency: integrating information in the brain. *Trends in Cognitive Sciences*, *2*(12), 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*(6), 3257–3262.

Touroutoglou, A., Lindquist, K. A., Dickerson, B. C., & Barrett, L. F. (2015). Intrinsic connectivity in the human brain does not reveal networks for "basic" emotions. *Social Cognitive and Affective Neuroscience*. http://dx.doi.org/10.1093/scan/nsv013/.

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.

Viskontas, I. V., Quiroga, R. Q., & Fried, I. (2009). Human medial temporal lobe neurons respond preferentially to personally relevant images. *Proceedings of the National Academy of Sciences*, *106*, 21329–21334.

Wager, T. D., Kang, J., Johnson, T. D., Nichols, T. E., Satpute, A. B., & Barrett, L. F. (2015). A Bayesian model of category-specific emotional brain responses. *PLOS Computational Biology*, *11*(4), e1004066. http://dx.doi.org/10.1371/. journal.pcbi.1004066.

Widen, S. C., & Russell, J. A. (2013). Children's recognition of disgust in others. *Psychological Bulletin*, *139*(2), 271–299.

Wilson-Mendenhall, C. D., Barrett, L. F., & Barsalou, L. W. (2013). Neural evidence that human emotions share core affective properties. *Psychological Science*, *24*(6), 947–956.

Wilson-Mendenhall, C. D., Barrett, L. F., Simmons, W. K., & Barsalou, L. W. (2011). Grounding emotion in situated conceptualization. *Neuropsychologia*, *49*(5), 1105–1127.

Wundt, W. (1897). *Outlines of psychology (C. H. Judd, Trans.)*. Leipzig: Wilhelm Engelmann.

Wundt, W. (1998). *Outlines of psychology* (C. H. Judd, trans.). Bristol, UK: Thoemmes Press. (Original work published 1897).

Yuen, K. S. L., Johnston, S. J., De Martino, F., Sorger, B., Formisano, E., Linden, D. E. J., et al. (2012). Pattern classification predicts individuals responses to affective stimuli. *Translational Neuroscience*, *3*, 278–287.

Zola-Morgan, S. (1995). Localization of brain function: The legacy of Franz Joseph Gall (1758–1828). *Annual Review of Neuroscience*, *18*(1), 359–383.