Solving the Emotion Paradox: Categorization and the Experience of Emotion

Lisa Feldman Barrett
Department of Psychology
Boston College

In this article, I introduce an emotion paradox: People believe that they know an emotion when they see it, and as a consequence assume that emotions are discrete events that can be recognized with some degree of accuracy, but scientists have yet to produce a set of clear and consistent criteria for indicating when an emotion is present and when it is not. I propose one solution to this paradox: People experience an emotion when they conceptualize an instance of affective feeling. In this view, the experience of emotion is an act of categorization, guided by embodied knowledge about emotion. The result is a model of emotion experience that has much in common with the social psychological literature on person perception and with literature on embodied conceptual knowledge as it has recently been applied to social psychology.

Humans experience emotion. For many, experience serves as an emotion’s central and defining aspect. We feel the heat of anger, the despair of sadness, the dread of fear. Most days, at least in North America, each of us asks and answers about our emotional state. We talk about our experiences of emotion over coffee, or during a telephone conversation, or in a therapist’s office. All things being equal, we assume that experiencing an emotion gives us introspective access to the emotion itself. When an emotion is triggered, we feel it, in much the same way that we feel hunger before a meal.

In psychological science, researchers often assume that people experience emotion because people have “emotions”—internal mechanisms for a small set of reactions (typically happiness, anger, sadness, fear, disgust, and interest) that, once triggered, can be measured in a clear and objective fashion. Emotions such as anger, sadness, and fear are treated as entities that scientists can make discoveries about—and for better or worse, this assumption shapes the scientific treatment of emotion experiences. It is assumed that people feel fear when their fear mechanism has triggered, such that this feeling shapes perception and decision making and can have important consequences for social relationships.

Although there is accumulating evidence that the experience of emotion can have great consequence for subsequent thoughts, decisions, and behaviors, it remains unclear whether such experiences issue from separate mechanisms for anger, sadness, fear, and so on. The goal of this article is to explore how scientists might understand the experience of emotion if this view of emotions is set aside. I sketch a framework for one such model that is more consistent with the empirical evidence on emotion and incorporates existing research on psychological mechanisms more broadly. The result is a model of emotion that has greater explanatory power for understanding the richness and diversity of emotional life.

To begin the article, I argue that many contemporary models of emotion are guided by the assumption that emotions are entities, and I examine the consequences of this assumption for the scientific study and measurement of emotion experience. Next, I briefly review the literature and conclude that these assumptions are not well grounded in the available empirical evidence. This frames an emotion paradox: Our everyday experiences of anger, sadness, fear, and several other emotions are compelling, but they are scientifically elusive and defy clear definition.

The next major section of the article then explores the idea that people experience emotion in the same way that they see color or the way that they perceive behaviors in others: People use knowledge to parse and conceptualize the bottom-up information that is sensorially given. I sketch a framework where the ex-
Emotions as Entities

In the history of psychological science, there are several notable examples where psychologists have progressed from thinking about psychological phenomena as unitary faculties of the mind—entities, if you will—to thinking about them as emergent phenomena that vary with the immediate context. This is most clearly the case in the study of memory (e.g., Johnson, 1992; Loftus, 1992; Schacter, 1996). No one would deny that “memories” exist, but they are no longer thought of as intact entities like computer files that are stored in the brain to be retrieved when needed. Psychologists no longer search for the engram (the place in the brain where a memory resides). A memory is not an entity to be found. If I had coffee with a friend yesterday, and I remember that event today, the content of my memory may differ from what happened in the actual event, and my memory for this event several weeks from now may differ to some extent again. A similar argument has been made for the idea of “personality” (Mischel, 1984; Mischel & Shoda, 1995), and more recently for “concepts” (Barsalou, Simmons, Barbey, & Wilson, 2003). Each was once assumed to be a fixed entity with an identifiable causal mechanism or essence. Now, each is accounted for by distinct but interacting systems. A given instance of a memory, personality, or concept, is sensitive to context, so may be somewhat different from what occurs at some other instance. This variability, rather than being viewed as a source of error or bias, is a valid reflection of a person’s psychological state that must be modeled and explained.

Despite arguments to the contrary (e.g., Ortony & Turner, 1990), the dominant scientific paradigm in the study of emotion is grounded in the assumption that emotions are entities. This idea is schematically depicted in a very simple way in Figure 1: Some event (usually external) triggers a kind of emotion, producing a set of recognizable behavioral and physiologic outcomes (including a set of facial movements, a vocal signal, changes in peripheral physiology, and some voluntary action) that are coordinated in time, correlated in intensity, and constitute the components of an emotional response. Different kinds of emotion (e.g., anger, sadness, fear) are presumed to be associated with different coordinated response patterns. Presumably, these patterns are how we (as people and as scientists) recognize that an emotion has occurred. They allow us to know an emotion when we see it.

Figure 1 is simple, but it is not simplistic. The general idea is that emotions are unlike other psychological phenomena (say, cognition or attention), and each category of emotion, often referred to in psychological science by such English words as anger, sadness, and fear, is a natural kind (Barrett, in press-a). Instances of each kind of emotion cluster together in a meaningful way because they have something real in common.

Figure 1 distills the assumptions that are common to many scientific models of emotion, regardless of how those models differ in their surface features. Substitute “affect program,” “internal signals,” or a literal neural circuit defined by a specific brain area or a neurotransmitter system for “emotion” and Figure 1 summarizes the “basic” emotion approach (e.g., Buck, 1999; Ekman, 1973, 1992; Izard, 1977, 1993; Johnson-Laird & Oatley, 1992; LeDoux, 1996; Lundqvist & Öhman, 2005; Panksepp, 1998; Plutchik, 1980; Tomkins, 1962). Substitute “appraisals” for “emotion,” and Figure 1 is consistent with the basic assumptions of some appraisal models of emotion, particularly those models where appraisals are assumed to be literal cognitive processes that produce an emotional response (e.g., Lazarus, 1991; Roseman, 1984, 1991; Scherer, 1984). In principle, appraisal models acknowledge the enormous variety in emotional responding and do not assume that particular emotions are basic in any biological way, but these models do tend to organize emotional responding into the familiar set of discrete categories, such as anger, sadness, fear, and so on (e.g., Frijda, 1986; Lazarus, 1991; Roseman, 1991; Scherer, 1984; C. A.
The Experience of Emotion

A correspondingly simple perspective on the experience of emotion falls out from the scientific models that resemble Figure 1: When an emotion is triggered, a person senses it, resulting in the experience of a distinct kind of feeling state. More formally, the experience of an emotion is presumed to be the vertical sensory detection associated with triggering an emotion mechanism (or, in some models, the experience of emotion is the sensation of the other outputs, such as facial muscle movements). The emotion is an object of consciousness like a table or a chair—the object causes your experience of it. When emotion sensations register, the experience of emotion results.

In Figure 1, as in many scientific models, emotion and its experience are not synonymous. The emotion itself is thought to be largely unconscious, causing people to behave in ways of which they are not aware. The feeling associated with the emotion is conscious, by definition, and is available to be experienced, although it may not be a focal point for attention (e.g.,

Smith & Ellsworth, 1985; for a review, see Barrett, Ochsner, & Gross, in press).

Of course, Figure 1 is not comprehensive. Several scientific models do not assume that distinct kinds of emotion cause tightly coordinated suites of outputs (e.g., Averill, 1980; Harré, 1986; James, 1884; Kagan, 1979; Mandler, 1975; Ochsner & Barrett, 2001; Russell, 2003; Schachter & Singer, 1962; Shweder, 1994; Solomon, 2003). Furthermore, Figure 1 does not portray all of the complexity in the basic emotion and appraisal models that it is consistent with. Some models focus on certain outputs (e.g., facial movements and peripheral physiology) but ignore others (e.g., experience). And there appears broad acknowledgment that epigenetic influences (such as context and learning history) play a role in emotional responding (e.g., Ekman, 1992; Keltner & Haidt, 2001; Panksepp, 1998). Yet, despite these caveats, it is not uncommon for emotion researchers to make reference to specific kinds of emotions that are represented as invariant responses to particular kinds of antecedents.

When boiled down to their fundamental assumptions, the models that make up the dominant scientific paradigm in the psychological study of emotion suppose that traditional emotion categories exist as real entities in nature, or natural kinds, whether nature is defined as residing in the brain or body or in the deep structure of the situation (depending on your preferred level of description; Barrett, in press-a). Once an emotion is triggered (whether computed by an emotion program or a set of appraisals), the presumed result is an automated set of synchronized changes in response systems that produce the signature emotional response (such as the output side of Figure 1).

The Objective Measurement of Emotion Experience

Figure 1 represents the idea that there is some objective indicator of a person’s experience of emotion. If aspects of an emotional response (that is, the outputs in Figure 1) are connected by a single, common cause, it should be possible to measure more easily observable aspects of emotion (e.g., the facial movements, the vocal expression, the voluntary action, peripheral physiology) to learn something about experience (which itself is not observable). Lambie and Marcel (2002) made this argument when they stated that first-order emotion experience could be measured by observing expressive behavior (what an emotion feels like can be determined by how a person behaves). The more observable aspects of emotion could then be used to validate a person’s self-report of their own experience. Lack of correspondence between verbal reports and behavior would presumably indicated that the verbal reports are invalid. For example, if a person says he is angry, but moves his face in a way that we easily label sad, then we believe him to feel sad. Behavior trumps verbal report. To see a behavior in others is to see evidence of a causal mechanism (or so we believe).

A quick look at the empirical evidence suggests that it is difficult, if not impossible, to find an objective means of measuring the experience of emotion, however. As I show, none of the responses that scientists use to measure emotion (i.e., the outputs in Figure 1) consistently distinguish between supposed instances of anger, sadness, fear, and so on.

1Feeling does not necessarily serve as evidence of the causal mechanism to observers, however. For the observer (observing the emoter), the emotion is revealed by expressive behavior, over and above anything else (including the emoter’s verbal report of experience).
Perhaps the most compelling idea in the psychology of emotion is that emotional states have specific and unique patterns of somatovisceral changes. Although individual studies sometimes report distinct autonomic correlates for different emotion categories (e.g., Christie & Friedman, 2004; Ekman, Levenson, & Friesen, 1983; Levenson, Ekman, & Friesen, 1990), meta-analytic summaries generally fail to find distinct patterns of peripheral nervous system responses for each basic discrete emotion (Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 2000). Peripheral nervous system responses configure for conditions of threat and challenge (Quigley, Barrett, & Weinstein, 2002; Tomaka, Blascovich, Kelsey, & Leitten, 1993; Tomaka, Blascovich, Kibler, & Ernst, 1997) and for positive versus negative affect (Cacioppo et al., 2000; Lang, Greenwald, Bradley, & Hamm, 1993), but do not robustly and unambiguously distinguish between emotion categories such as anger, sadness, and fear.

Evidence from facial behavior has yielded the same result. Facial electromyography measurements coordinate around positive versus negative affect (Cacioppo et al., 2000) or intensity of affect (Messinger, 2002), rather than discrete emotion categories. Participants can assign posed facial muscle configurations to emotion categories with some reliability, but these findings are open to alternative explanations (e.g., Russell, 1994; Russell, Baborowski, & Fernandez-Dols, 2003), including that perceivers are imposing, rather than detecting, categorical distinctions in the facial configurations that they rate (I will return to this point later).

Evidence from instrumental behavior is similar. Behavioral responses such as flight or fight correspond to situational demands (Bouton, 2005), rather than to specific categories of emotion. Behaviors are specific, context-bound attempts to deal with a situation (Cacioppo et al., 2000; Lang, Bradley, & Cuthbert, 1990). Functional demands vary with situations, making it likely that a range of behaviors will occur with instances of the same emotion category will involve. For example, Lang et al. noted that the behaviors associated with fear can range from freezing to vigilance to flight. Not only are different behaviors associated with the same emotion category, but also one type of behavior can be associated with many categories. For example, attack behaviors (e.g., defensive or offensive) are associated with different types of stimulus situations assumed to connote fear or anger (Blanchard & Blanchard, 2003).

Many theorists assume that kinds of emotion have specific neural essences (e.g., Buck, 1999; Damasio, 1999; Dolan, 2002; Ekman, 1992; Izard, 1993; LeDoux, 1996; Panksepp, 1998). Yet, two recent meta-analyses of neuroimaging studies (Murphy, Nimmo-Smith, & Lawrence, 2003; Phan, Wager, Taylor, & Liberzon, 2002) failed to find consistent evidence for particular neural correlates for anger, sadness, disgust, and happiness (Barrett, in press-a). A fear–amygdala correspondence was noted across both analyses, but can be accounted for by alternative explanations (see Phan et al., 2002). For example, the fear–amygdala correspondence (which occurred in about 60% of the studies sampled) may be due to the fact that fear stimuli (such as facial depictions of fear) differ from other emotional stimuli on features such as novelty (e.g., Schwartz et al., 2003; Wilson & Rolls, 1993; Wright et al., 2003) and uncertainty (Davis & Whalen, 2001; Whalen, 1998; Whalen et al., 1998, 2001). Facial configurations of surprise produce levels of activation in the ventral amygdala that are similar to those observed for fear (Kim, Somerville, Johnstone, Alexander, & Whalen, 2003; Kim et al., 2004), consistent with the view that the amygdala may be computing the predictive value of a stimulus (i.e., the likelihood it will predict threat or reward), rather than specifically responding to fear stimuli per se.

Although there are a number of methodological and theoretical factors that currently limit our ability to draw inferences about the neural bases of emotional responses, the failure to find neural signatures for distinct emotions thus far is consistent with the behavioral evidence. A different way to think about the findings is that there is good evidence that specific behaviors (e.g., freezing) may depend on specific brainstem and subcortical nuclei (e.g., Panksepp, 1998), but there is little evidence to suggest that each behavior is necessarily or sufficiently associated with any single emotion category. Because people can effortlessly assign such behaviors to emotion categories does not necessarily make doing so scientifically accurate or useful. So whereas freezing may be an innate behavior, and may be part of the English script for the category fear, it is not necessarily innately linked to some module of fear responding.

Most important, physiological, behavioral, and experiential outputs for each emotion category are only weakly intercorrelated (Bradley & Lang, 2000; Lang, 1968; Mandler, Mandler, Kremen, & Sholiton, 1961; Mauss, Wilhelm, & Gross, 2004), undermining the claim that the various responses in Figure 1 derive from a single, common cause. In studies where moderate correspondences are found (e.g., correspondences between facial behaviors and ratings of experience for anger and sadness, with effect sizes ranging from .34 to .52; Bonanno & Keltner, 2004), alternative explanations due to differences in other properties, such as arousal or intensity, were not ruled out. Moreover, facial behaviors, reports of experience, and peripheral nervous system activity show stronger correspondences for the affective properties of valence and intensity (effect sizes range from .76 to .90; Lang et al., 1993). Without strong correlations at the level of specific emotions, then it is difficult to justify using one
measure (e.g., facial behaviors) as a proxy for another (e.g., the experience of anger, sadness, fear, etc.).

Several different caveats have been offered to account for the absence of biobehavioral signatures for different emotion categories. One argument suggests that social factors, like display rules (Ekman, 1972; Ekman & Friesen, 1969) or other regulation processes might mask or inhibit prepotent responses that would otherwise be correlated. A second argument suggests that scientists routinely fail to observe signatures because response systems differ in their temporal dynamics, sensitivity, and reliability of measurement (e.g., Bradley & Lang, 2000). A third argument is that laboratory studies of emotion do not employ emotion-eliciting stimuli that are strong enough to produce a prototypical emotion episode, where responses would be correlated with one another (Tassinary & Cacioppo, 1992). Although any of these explanations may be correct, an equally plausible explanation is that scientists have failed to observe stable and reliable response clusters because they are not really there. At this point, enough evidence has accumulated for some theorists to conclude that the lack of coherence within each category of emotion is empirically the rule rather than the exception (Bradley & Lang, 2000; Ortony & Turner, 1990; Russell, 2003; Shweder, 1994).

Taken together, then, the available body of evidence suggests that there is no clear objective way to measure the experience of emotion. Scientists are not able to use any single measurement, or profile of measurements, to indicate when a person is in a state of anger or fear or sadness, and so on. The fact that the available data fail to configure into categories with clear boundaries is not necessarily evidence that the data are flawed. Rather, the data might be viewed as instructive because they reflect something important and true about the nature of emotional responses.

In sum, scientists may be able to assess a person’s affective state (i.e., pleasure and displeasure) by more indirect (see Berridge & Winkelman, 2003) or objective means (in the face or body, e.g., Cacioppo et al., 2000), but these measurements cannot be used to assess feelings of anger, sadness, fear, per se. Verbal report, even with all of its failings, may be the only means of assessing the experience of emotion. If we want to know whether a person is experiencing an emotion, we have to ask them.

**Experiencing “Kinds” of Emotion**

Figure 1 not only depicts the idea that there are qualitatively different emotion states that can be measured objectively, but it represents the idea that experiences of those states should differ qualitatively as well. If anger and sadness and fear are distinct states with discrete boundaries in nature, then the sensations that derive from those states—experience of those emotions—should be structured in kind. The taxonomic structure of experience should reflect the structure of emotions as they really exist. As a result, everyone should know the difference between a sad feeling, an angry feeling, a guilty feeling, and so on.

If self-reports of emotion experience have any validity at all, then when projected into geometric space, those reports should exhibit a simple structure (Thurstone, 1935), with one factor each for anger, sadness, fear, and so on. This would provide evidence that each kind of emotion is associated with an experiential primitive feeling, meaning that the feeling cannot be broken down into component parts or reduced to anything else psychological. If self-reports fail to show simple structure, then this can be taken as evidence that those reports are not valid. Figure 1 makes clear that experiences should bear a one-to-one correspondence with the emotions themselves, but words for feelings may not. As a result, dimensional properties (such as pleasantness and arousal) that can be identified in self-reports of emotion experience are assumed to represent the artificial influence of language. In general, verbal reports are considered a fallible way of studying the experience of emotion. They are “shackles” from which scientists must be freed because they tell scientists more about emotion language than they do about the experience of emotion or the emotions from which those experiences are derived.

There is no clear evidence for qualitatively different kinds of experiences, however, just as behavioral and biological findings have not produced strong evidence for kinds of emotions. In the following sections, I review evidence that people vary in their tendency to represent their experiences as categorically distinct events, that reports of experience give evidence of more basic psychological properties of valence and arousal, and that these properties are note merely due to the artificial influence of language. Furthermore, I discuss how self-reports, despite their obvious limitations (Stone & Turkkan, 2000), do appear to contain valid information about fundamental aspects of experience.

**Knowing the difference between experiences.** Contrary to popular belief, it is far from clear that everyone necessarily experiences anger, sadness, fear, and so on, as qualitatively different states (Carstensen, Pasupathi, Mayr, & Nesselroade, 2000; Lane, Quinlan, Schwartz, Walker, & Zeitlin, 1990; Lane & Schwartz, 1987; Larsen & Cutler, 1996). When people report on their experiences in the course of everyday life over several weeks, and those self-reports are analyzed as verbal behaviors, it becomes clear that some people make categorical distinctions, characterizing their experiences in discrete emotion terms, whereas others characterize their experiences in broad, global terms (Barrett, 1998; Barrett, Gross, Conner, & Benvenuto,
2001; Feldman, 1995a). I have called this phenomenon *emotional differentiation* (Barrett et al., 2001) and, more recently, *emotional granularity* (Barrett, 2004).

Individuals who are low in emotional granularity report their experiences in global terms. They use discrete emotion labels such as *angry, sad,* and so on, to represent only the most general aspects of their internal state (typically pleasure and displeasure). For example, in response to the events of September 11, 2001, a student who was lower in emotional granularity said, “I felt a bunch of things I couldn’t put my finger on. Maybe anger, confusion, fear. I just felt bad on September 11th. Really bad.” This person was using discrete emotion terms to communicate an unpleasant or negative state. More generally, individuals who are lower in emotional granularity use emotion-related words such as *happy* and *excited* to mean *pleasant,* and the terms *sad* and *angry* to mean *unpleasant.*

Individuals higher in emotional granularity report their experiences in more precise, differentiated terms, using discrete emotion labels in a way that captures the distinctiveness in the words’ meaning. For example, a student who was higher in granularity said, “I felt a bunch of things I couldn’t put my finger on. Maybe anger, confusion, fear. I just felt bad on September 11th. Really bad.” This person was using discrete emotion terms to communicate an unpleasant or negative state. More generally, individuals who are lower in emotional granularity use emotion-related words such as *happy* and *excited* to mean *pleasant,* and the terms *sad* and *angry* to mean *unpleasant.*

The fact that people differ in emotional granularity indicates that not everyone knows the difference between a sad feeling, an angry feeling, a guilty feeling, and so on. It is tempting to assume that differences in granularity reflect differences in accuracy, that some people are better than others when reading off their internal emotional states and translating them into words. This interpretation, however, presupposes that these distinctive internal emotional states exist, and that they can be quantified with some scientific criterion. I have already presented evidence, however, that there is currently no way to scientifically confirm when a person is happy or angry or sad.

The taxonomic structure of self-reported experience. The taxonomic structure of self-reported experiences of emotion does not support the view that anger, sadness, fear, and so on, are qualitatively distinct and experientially primitive. When reports of emotion experience are projected into geometric space, they do not generally assume a simple structure configuration, with one factor each for anger, sadness, fear, and so on (for a recent review, see Barrett, in press-a). Instead, they take on a roughly circular (or circumplex-type) shape. Self-reports taken from a group of individuals at one point in time configure into a circumplex shape (Feldman, 1995b; Russell, 1980; Yik, Russell, & Barrett, 1999; for a review, see Barrett & Russell, 1998, 1999; Russell & Barrett, 1999), as do more idiographic reports that are taken over time and modeled separately for each person (Barrett, 1998, 2004; Feldman, 1995a).²

Psychological measurements only take on a circular shape if they are heterogeneous, meaning that those measurements can be decomposed into a least two (but perhaps more) more fundamental psychological properties (Guttman, 1957). The fact that self-reports of experienced emotion configure into a circumplex-like shape is evidence, therefore, that those reports are not psychologically primitive (for a discussion of why the emotions are not psychological primitives, see Ortony & Turner, 1990). Whether people use emotion words in a highly granular manner or not, self-reports of experience can themselves be broken down further into component parts. These components are valence (pleasure–displeasure), and to a lesser extent arousal (activation–deactivation). Valence and arousal are usually called *dimensional* properties of experience.

When people report experiences of emotion, they are able to explicitly describe feelings of valence (pleasure or displeasure) and activation (feeling sleepy or excited; Russell, Weiss, & Mendelsohn, 1989). People also implicitly communicate valence and activation in their self-reports of experienced emotion (Barrett, 2004, 1998; Feldman, 1995a). All humans, it seems, can tell the difference between a pleasant affective state and an unpleasant affective state (Barrett, in press-b). Many, but not all, people also characterize their affective states as high or low in activation.

**Affect dimensions and language.** The dimensional properties observed in self-reports of emotion experience (such as valence and arousal) are often thought to represent the artificial influence of language or cognition. It has been argued that self-reports reflect people’s beliefs about what they feel, rather than the contents of conscious feeling (Dennett, 1991), or that valence and arousal derive from the contaminating influence of emotion language (that is, the words used in the rating process; e.g., Frijda, Markham, Sato, & Wiers, 1995; Ortony, Clore, & Collins, 1988). The conclusion has been that valence and arousal are artifacts of the self-report process and do not directly inform on emotion experience. There is mounting evidence, however, that valence, and to a lesser extent arousal, represents more elemental aspects of a person’s internal state and are not merely an artifact of language or belief.

Although the definitive experiments have yet to be done, valence, in particular, is are closer to meeting the criteria for natural-kind status, suggesting that the appearance of valenced content in self-report of experience may not be an artifact. Of course, emotion lan-

---

²Self-reports rarely take on a perfect circular shape, but a circumplex need not be perfectly circular with equally spaced elements (Fabrigar, Visser, & Browne, 1997).
guage, and language in general, can be characterized in terms of valence and arousal (e.g., Osgood, 1969; Russell, 1991; Wierzbicka, 1992). But it is important to note that so can more objective measures of emotion (i.e., the output side of Figure 1). Judgments of facial behaviors (for a recent review, see Russell et al., 2003), autonomic physiology (for a meta-analytic review, see Cacioppo et al., 2000), and expressive behavior (for a review, see Cacioppo & Gardner, 1999), including electromyography recordings of facial movements (for a meta-analytic review, see Cacioppo et al., 2000) and acoustic components of voices (Bachorowski, 1999) can all be described in terms of pleasure–displeasure or its intensity. Moreover, although the evidence for such coordinated signature responses for categories of emotion, such as anger, sadness, fear, and so on, is slim, there is some evidence for such coordination in affective responses (Lang et al., 1993; for a recent review, see Bradley & Lang, 2000).

Furthermore, recent evidence suggests that a person’s cognitive structure of emotion language does not strongly dictate the way that he or she uses emotion words to verbally represent experience (Barrett, 2004). Three studies compared individual differences in the extent to which people emphasize valence and arousal in their self-reported emotion experience (termed valence focus and arousal focus, together called affective focus), with individual differences in the extent to which those same individuals emphasize valence and arousal in the structure of emotion language (semantic focus), allowing a direct test of their relation. In all three studies, individual differences in the cognitive structure of emotion language were observed, but these individual differences in semantic focus were weakly related to valence focus and moderately related to arousal focus.

Perhaps most important, then, are several studies documenting that both valence focus and arousal focus have distinctive external correlates. Valence focus and arousal focus, because of the way that they are estimated, reflect the affective content that is implicitly contained in self-reports of emotion experience. Each is uniquely related to more basic psychological processes, suggesting that valence and arousal in self-reports are not artifacts of words or beliefs.

Valence focus is related to the frequency and intensity of evaluative processing. When compared to those lower in valence focus, those who focused more on feelings of pleasure and displeasure showed greater perceptual sensitivity to facial affect in others (Barrett & Niedenthal, 2004). Individuals higher in valence focus perceived the onset of angry and sad facial expressions earlier than did those lower in valence focus. Individuals higher in valence focus experienced more labile self-esteem in response to positive and negative interpersonal cues, providing evidence that individuals who are strongly valence focused are more sensitive to evaluative cues around them (Barrett, in press-b). Individuals higher in valence focus also showed enhanced automatic evaluations in a sequential priming procedure, showing that they are generally prepared to evaluate, regardless of the circumstances or context (Conner, Barrett, & Bliss-Moreau, 2005).

Arousal focus, too, has an external correlate that provides evidence of its validity as an index of actual feeling. Individuals higher in arousal focus were more interoceptively sensitive and less biased during a heartbeat detection task when compared to those who were lower in arousal focus (Barrett, Quigley, Bliss-Moreau, & Aronson, 2004). These findings suggest that individual differences in the ability to accurately perceive internal bodily cues provided one avenue for subjective feelings of activation and deactivation that, in turn, lead to higher arousal focus.

Summary

A brief review indicates that there is little evidence to support the hypothesis, depicted in Figure 1, that the experience of emotion is the simple, veridical sensation of an emotion mechanism. Instead, the empirical evidence supports two conclusions about the study and measurement of emotion experience. First, the experience of emotion cannot be measured objectively. Second, discrete emotion experiences are not psychologically primitive. The research also supports two additional observations that must be accounted for by any model of emotion experience. People differ in their emotional granularity, or precision with which they represent experiences of emotion. And, self-report data, despite their fallibility, provide evidence that people experience, and report experiencing, feelings with at least two properties—valence and arousal. Physiological and behavioral measures also communicate something about a person’s affective state, rather than categories of emotion such as anger, sadness, and fear per se.

The Emotion Paradox

Thus far, one can make several observations about human emotion. People parse the world into things emotional and nonemotional, and they further divide the emotional world into discrete categories. In Western culture, these categories correspond to the English words anger, sadness, fear, disgust, happiness, and possibly a few others (guilt, shame, and so on). It seems natural and easy to find these emotions in others people. In scientific experiments, perceivers effortlessly and confidently see these emotions in the facial muscle movements of other people. There are caveats (Russell et al., 2003). People agree on which emotion
is seen in a face more when they are from the same culture. They agree on the emotion more when viewing posed, caricatured muscle configurations than when viewing naturally moving faces. And they easily adjust the emotion that they see to fit the immediate context if one is provided. The experience of looking at another person’s face, and seeing emotion there, is familiar and compelling to most everyone, even though instrument-based measures of facial movements more consistently configure to represent affect, rather than discrete categories of emotion.

People not only see emotions in others, but they experience emotions themselves. In everyday language, people use words such as angry, sad, afraid, and so on to describe their internal feeling states, and they assume that these states are natural and obvious. These feelings have a given quality. People experience anger as erupting or “happening to” them, so that it seems as if an anger mechanism hijacks the mind and body, presumably causing people to behave in ways that they would rather not (e.g., “I yelled because I was angry”). There are caveats. Some people characterize their experiences of emotion with more granularity than do others. And instrument-based measures of internal states yield an estimate of valenced affect, rather than discrete emotions per se.

When taken together, these observations frame a fundamental emotion paradox: People are compelled by their own experiences to believe that emotions exist as natural-kind entities, yet a century of research has not produced a strong evidentiary basis for this belief. To date, there is no clear, unambiguous criterion for indicating the presence of anger or sadness or fear.

There are two ways to solve this emotion paradox that correspond to Greenwald’s discussion of a disconfirmation dilemma (Greenwald & Ronis, 1981). One solution is to argue that the blueprint for each kind of emotion will emerge with better experiments, using more sophisticated research tools and more precise designs. A second solution, and the one that I pursue here, argues that emotions are not biologically given, but are constructed via the process of categorization. Emotions exist, but only as experiences. Specifically, the experience of feeling an emotion, or the experience of seeing emotion in another person, occurs when conceptual knowledge about emotion is brought to bear during the act of categorization. The experience of feeling emotion occurs when a person categorizes his internal state. The experience of seeing emotion in another person occurs when that person’s behavior is categorized as emotional.

In the next section, I begin to develop this view with a brief overview of how conceptual knowledge shapes perceptual information via the process of categorization. Specifically, I discuss how people use category knowledge about color to shape the perception of wavelengths of light into the experience of color (Davidoff, 2001) and how category knowledge of people shapes perceptions of behavioral actions into meaningful acts (Gilbert, 1998; E. R. Smith & DeCoster, 2000). I argue that the process of categorization is fundamental to both color and person perception and draw a parallel between these categorization processes and the way in which people use their knowledge of emotion categories to shape an experience of emotion. After outlining the general framework for how emotion knowledge might influence the experience of emotion via categorization, I present my solution to the emotion paradox in more detail and discuss its implications for developing a research agenda to investigate the richness and diversity of emotional life as it is experienced.

Categorization and Experience

Categorizing is fundamental cognitive activity. A category is a class of things that are treated as equivalent. To categorize something is to determine what it is, why it is, and what to do with it. A concept can be thought of as a collection of mental representations for a category that people draw on during the process of categorization. Once conceptual knowledge is brought to bear to categorize something as one kind of thing and not another, the thing becomes meaningful. It then becomes possible to make reasonable inferences about that thing, predict about how best to act on it, and communicate our experience of the thing to others. Although volumes of scientific literature have been devoted to the topic of categorization, and a full review is beyond the scope of this article, I briefly review the role of conceptual knowledge and categorization in color perception and person perception, to help describe a framework for understanding the role of conceptual knowledge in the experience of emotion.

Seeing Color

The light spectrum is a continuum of wavelengths, yet colors are experienced categorically. The English words red, green, blue, and yellow correspond to sets of wavelengths that are experienced as qualitatively different. There are some embodiment constraints on color categorization. We sense light at a particular wavelength in a way that is constrained by how the retina registers sensory information, how this information is transduced, and how low-level visual processing takes place. Although the issue of color realism is still very much a matter of debate (e.g., Byrne & Hilbert, 2003) the processes that transduce reflected light alone are thought by many to be insufficient to explain how people experience the color corresponding to a particular wavelength.
The color that people experience for a given wavelength is influenced by conceptual knowledge of color. Early in childhood, English-speaking people learn about what constitutes the category blue (green, red, and so on), including knowledge of different hues that are treated as equivalent, what objects are typically associated with those hues, what colors match with blue, and so on. There is variability in conceptual knowledge of color, so the way that one person categorizes a certain wavelength may differ from someone else’s, leading to a different experience of color. This is most clearly the case when looking at language-based differences in color categorization. Color categories are created by language, and languages differ in how they name parts of the color spectrum (Özgen, 2004). For example, faced with an object reflecting light at 450 nm, a Berinmo speaker (from Papua New Guinea) would experience the object as green, because Berinmo distinguishes between green and brown categories. An English speaker might see the object as blue, because English distinguishes between green and blue (Roberson, Davies, & Davidoff, 2000). A Turkish speaker, on the other hand, would experience the object as lacivert or dark blue, which in Turkish is a different category than mavi or blue (Özgen & Davies, 1998). In each instance, the information hitting the retina is the same (light at 450 nm), and the detection of that wavelength is dictated by what the visual system produces in its early stages of processing. The experience of color that corresponds to this wavelength depends on the conceptualization of the sensory information, however. For a given object in a given instance, such as a shirt or a flower, a person might experience the color blue, navy blue, or even green, depending on what he or she knows about color.3

The effect of category knowledge on color experience is not a mere case of verbal labeling. English speakers can learn to perceptually distinguish between two hues that are usually categorized as blue, indicating that color perception is changed by learning hue categories (Özgen & Davies, 2002). Further, behavior toward the blue-like object, be it a shirt (what pants to wear with) or a flower (whether or not to plant it) or mold on a piece of cheese (whether or not to eat the cheese), will depend, in part, on what a person knows about color and how that knowledge is used during categorization of the object’s color. Presumably, as a person becomes more of a color expert and learns more about color (perhaps by taking an interior design course), the way he or she experiences and acts on blue-like objects will change.

**Seeing Behavior**

Conceptual knowledge not only influences our experience of color, but it also influences our experience of other people. Social psychology has accumulated a large and nuanced body of research on how we perceive other people’s behavior and infer causes for them (for a historical review, see Gilbert, 1998). In this literature, the word perception is used to refer to the process of assigning someone (or his or her behavior) to a meaningful category so that a perceiver “sees” an instance of that category and can infer something about the person’s internal state or enduring disposition or both. In essence, perception here refers to the process of categorization (Allport, 1954; Macrae & Bodenhausen, 2000). A person’s actions are categorized into discrete meaningful behaviors, and that person is categorized as to what kind of person he or she is. The general idea is that our knowledge of people and situations automatically and effortlessly shapes what we “see” people doing and gives rise to our explanations for that behavior. In a sense, person perception is the study of theory of mind—how the categorization process allows people to attribute mental states to identify, explain, and predict behavior.

Research on person perception has produced a highly developed psychological model that is empirically well grounded (for a review of evidence, see Gilbert, 1998; Lieberman, Gaunt, Gilbert, & Trope, 2002). The process of person perception has been decomposed into three components that were initially assumed to occur in sequence (for a review, see Gilbert, 1998) but that are now assumed to proceed more or less in parallel (Kunda & Thagard, 1996; Lieberman et al., 2002; Reed, Vanman, & Miller, 1997; E. R. Smith & DeCoster, 2000). For the purposes of this article, we will concentrate on one element in the person-perception process known as categorization or behavioral identification: the way in which ongoing actions are categorized into discrete, psychologically meaningful behaviors.

For one person to understand the behavior of another, the behavior itself must first be identified. Physical movements are continuous in the same way that the visible light spectrum is continuous. People are constantly moving and doing things—that is, they are constantly engaging in “behavioral actions.” Yet, behaviors are perceived as discrete acts (Newtonson, 1973; Newtonson & Engquist, 1976; Vallacher & Wegner, 1987), much in the same way that colors are experienced is categorical terms. The role of category knowledge about people (based on group membership, physical features, prior behaviors, and so on) in behavioral identification or categorization is similar to the role of

---

3It is not the case that a Berinmo speaker will mistake light reflected at 500 nm (in the green range in English) for light reflected at 450 nm (in the blue range in English). Rather, the Berinmo speaker would experience them as shades of the same color, in the way that some English speakers might experience light reflected at 620 nm and 670 nm as two shades of red.
category knowledge of color in color perception. A perceiver automatically and effortlessly partitions continuous movements into recognizable, meaningful, discrete acts, using category knowledge about people.

The idea that prior information structures incoming information is generally accepted in psychology. This is very clearly the case in person perception, where the knowledge that is active in the mind of a perceiver influences what behavior is seen, usually without awareness. What we know about people, and how we use what we know, influences what we see that person doing at a particular point in time, and how we explain that behavior (for a review, see Gilbert, 1998). For example, if a professor meets her student’s gaze, bobs her head up and down as the student talks, contracts the muscles around her mouth, and speaks encouraging words in soft tones, the student will see her engaging in nurturing behavior, in large part because the category “female” was helping to shape the student’s perception of her behavioral actions into a distinct, nurturing act. If, however, the professor knits her brow, fails to contract the muscles around her mouth, tilts her head to the side, and utters some words in a neutral tone, the student may see her as engaging in an act of criticism, again because category knowledge about “female” was active during the person-perception process (and the professor failed to produce the expected actions).

Seeing Emotion

In the position that I am advancing here, experiencing an emotion is less like seeing an object such as a chair and more like experiencing color or experiencing another person as nurturing. I am proposing that instances of what has been termed self-focused emotion (Lambie & Marcel, 2002), such as “I am angry at Bob,” or world-focused emotion (Lambie & Marcel, 2002), such as “Bob is an idiot” are experiences that are constructed in a similar fashion to the experience of blue-ness in a flower (“I see blue” or “that flower is blue”), or to the experience of another person as behaving in a nurturing way (“I think she is supportive” or “she is nice”).

It is not difficult to imagine how the experience of seeing anger in another person might result when knowledge about anger shapes the conceptualization of a person’s ongoing behavior. For example, if Joe moves his feet heavily as he walks, an observer might categorize this movement as the behavioral act of stomping. Category knowledge about anger might be primed because of the category that Joe belongs to (Joe is male, and the Western stereotype is that men get angry), or because of other features in the situation or context (perhaps someone just insulted Joe). Observers prefer to identify behavioral acts in terms of the target’s intentions (cf. Gilbert, 1998), which in turn gives meaning to the behavioral act. Part of seeing Joe stomp involves making the inference that Joe is in an intentional state (meaning his behavior is caused). It is easy, and typical, to assume that some sort of internal state caused Joe’s behavior—that is, Joe is behaving angrily, which the observers take as evidence that he is in an angry state. Essentially, the observer categorizes an instance of anger in Joe by seeing Joe’s physical movements as the act of stomping. Joe’s angry behavior might be caused because he is generally an angry person (this is a dispositional attribution or characterization of why Joe is angry in this instant), because he was insulted (this would be the situational attribution of why Joe is angry in this instant), or aspects other than Joe’s internal state might have caused his stomping (perhaps it is raining outside and Joe may have mud on his shoe, or perhaps he is merely stepping on ants). If people have the opportunity, motivation, and cognitive resources (Barrett, Tugade, et al., 2004; E. R. Smith & DeCoster, 2000) to pay attention to the fact that it is raining, or that someone left crumbs on the floor so that it is crawling with ants, they might change their impression of Joe. He’s not angry, he is just cleaning his shoes or killing ants.

Presumably, facial muscle movements in another person are experienced as facial “expressions” in much the same manner. If Joe contracts the muscles over his eyebrows, while pulling his mouth downward, people recognize this as the behavioral act of scowling. More typically, they will say “Joe is angry.” It is easy, and usual, for people to assume that Joe’s face is displaying anger, presumably because he is in an angry state. From a person-perception standpoint, I would argue that seeing Joe’s angry face was partly a function of categorizing his ongoing facial behaviors as an instance of the category of anger.

There is experimental evidence that conceptual knowledge about emotion seamlessly shapes the perception of emotion in others. Some of the initial research on behavior identification examined how emotion category information influenced the identification of emotion in photographs of facial configurations and in verbal descriptions of emotional reactions (Y. Trope, 1986). Participants were exposed to context information (in the form of situation scenarios) to prime a specific emotion category and, subsequently, rated the extent to which each facial configuration signaled anger, fear, sadness, happiness, and disgust. Results indicated that the situation produced strong contextual effects on emotion identification such that situation information influenced the identification of the emotion depicted on the actor’s face toward what was expected in that situation. Perceptions of faces depicting ambiguous emotional signals were influenced more than those caricaturing a specific emotion category, although the more extreme facial configurations were not immune from situational influence. Furthermore, similar
effects were obtained when participants were asked to identify the emotion in verbal descriptions of reactions (as opposed to facial depictions). Additional studies have replicated and extended these findings (Bouhuys, Bloem, & Groothuis, 1995; Carroll & Russell, 1996; J. Trope & Cohen, 1989).

Perhaps the most compelling evidence that conceptual knowledge shapes the perception of emotion in others comes from several studies where participants viewed ambiguous facial stimuli (e.g., blends of two different prototypical depictions such as happy and angry) when category knowledge about one or the other emotion was made more accessible (Halberstadt & Niedenthal, 2001). Later, participants were asked to identify the facial stimuli in computerized movies that morphed the previously seen faces from one prototypical facial configuration to the other (e.g., from a happy configuration into an angry configuration). In three studies, participants remembered the face stimuli in line with the conceptual knowledge that was active at encoding, strongly suggesting that conceptual knowledge about emotion shaped perceptual memory for the faces.

Although it may seem plausible that existing conceptual knowledge about emotion shapes the emotion that we see in others, it is somewhat more difficult to believe that our own experiences of anger, sadness, fear, and so on, issue from what we know about emotion. Knowledge of emotion and experiences of emotion seem like very different things. It has been argued that studying knowledge about emotion is not the same thing as studying the emotion itself (Clore & Ortony, 1991). Knowledge about emotion is cold and conceptual, whereas emotional states are hot and urge action. Knowledge can be deliberately called forth, whereas emotions are triggered and erupt without effort.

Yet, if we reject a naive realist view that our experiences are causally related to states that we experience, it is possible that the basic tenant of emotion and person perception holds true for the experience of emotion: Existing knowledge (about emotion) may thoroughly infiltrate perceptual processes by operating in an automatic fashion as a form of postattentive automaticity (Logan, 1992) to categorize an internal state as the experience (of emotion). The automatic nature of the process accounts for why experiences of emotion feel “given.” If processes that shape the phenomenal contents of color experience similarly shape emotion experience, or if the processes that enable us to know another also allow us to know ourselves, then conceptual knowledge about emotion may allow us to quickly and unconsciously categorize our own continuously changing internal state in much the same way that conceptual knowledge serves to categorize the continuous light spectrum or the continuous actions of other. The proposal, then, is that anger is experienced when a person brings knowledge about anger to bear on his or her own continuously changing internal state in the same manner that an experience of blue is the result of bringing conceptual knowledge of blue to bear on sensory information about wavelengths of light that are reflected off an object—or similarly, the way in which conceptual knowledge of woman is brought to bear on another person’s facial muscle movements, head movements, voice prosody, and so on.

**Solving the Emotion Paradox: Emotions as Conceptual Acts**

This section provides a more detailed account of how the process of categorization produces the experience of emotion. The goal is to account for experiences of anger, sadness, fear, and so on, without assuming that their phenomenological character derives from stereotyped, specific patterns of somatovisceral activity, brain activation, and behavior. When people say, “I am angry,” they are, of course, referring to some physical activity in their brain and body that is occurring as the result of their interaction with the world. A categorization account suggests that this activity is the result of (at least) two basic components—affect and conceptual knowledge about emotion.

This solution to the emotion paradox correspondingly involves two propositions. First, affect is a basic, biological substrate that is available to be categorized. Second, the conceptual knowledge that is called forth to categorize affect is tailored to the immediate situation, is represented in sensorimotor cortex, and is acquired from prior experience and supported by language. When combined, core affect and conceptual knowledge about emotion produce a highly flexible system that can account for individual variation in emotional granularity, cultural differences in emotion experience, and the full richness and range of experience that composes human emotional life.

**Core Affect**

One of the clearest findings from the scientific literature dealing with emotion is that all instrument-based measures of emotion give evidence of a person’s affective state. Objective measures of the face, voice, and body, as well as more subjective measures of experience, indicate whether a person is feeling pleasant or unpleasant. As a consequence, my first suggestion to account for the experience of emotion is that something called core affect is the basic building block of emotional life. Core affect has been characterized as the constant stream of transient alterations in an organism’s neurophysiological state that represent its immediate relation to the flow of changing events (Russell, 2003; Russell & Barrett, 1999). In a sense, core affect
is a neurophysiological barometer of the individual's relation to an environment at a given point in time. A person's momentary core affect is multiply determined and is an accounting of how events and objects influence his or her homeostatic state. It is similar to what Lambie & Marcel (2002) referred to as "emotion state." Core affect is the ongoing, ever-changing state that is available to be categorized during emotion conceptualization, much like the visible light spectrum is categorized in color perception and physical movements in person perception.

The term core in core affect is meant to refer to a specific construct that is distinct from the more general usage of the term affect (i.e., referring to anything emotional) and signifies several important ideas about this form of affective responding. Core affect may be a basic kind of "core" knowledge (as defined by Spelke, 2000). The hardwiring to support it is present at birth (Bridges, 1932; Emde, Gaensbauer, & Harmon, 1976; Spitz, 1965; Sroufe, 1979), so all humans are endowed with the ability to have core affective states. It is homologous in other mammalian species (Cardinal, Parkinson, Hall, & Everitt, 2002; Schneirla, 1959).

This form of affective responding is "core" because it is influenced by a very simple form of meaning analysis—whether stimuli or events are helpful or harmful, rewarding or threatening for a given person at a given point in time and whether an active behavioral response is required. This basic type of meaning is pancultural (Mesquita, 2003; Russell, 1991), so all human languages have words or concepts to communicate pleasure and pain, reward and threat (Wierzbicka, 1999). Although it can be communicated with words, core affect is not an artifact of verbal labeling. It can exist and influence behavior without being labeled or interpreted and can therefore function unconsciously (e.g., Berridge & Winkielman, 2003).

The term core also signals the idea that basic instrumental behaviors (such as approaching or avoiding) are part of this form of affective responding (for a review, see Cacioppo & Gardner, 1999). Objects and events have affective meaning to the extent that they can impact and change the homeostatic (core affective) state of the individual. The intensity of a core affective response (the degree of sympathetic and parasympathetic activation at a given moment in time) results in a perceived urgency to act that is independent of the specific action taken (the specific action being tailored to the particular situation at hand).

Finally, the term core also signifies that this form of affective responding forms the "core" of experience. Core affect (i.e., the neurophysiological state) is available to consciousness and is experienced as feeling good or bad (valence) and to a lesser extent as activated or deactivated (arousal). Extreme changes that capture attention and deliberate introspection allow core affect to always be represented verbally as feelings of valence, and it is often represented as arousal. These feelings are primitive (psychologically irreducible) and universal, although the extent to which one or the other property characterizes a conscious feeling varies within a person over time (Barrett, forthcoming), across people (Barrett, 1998, 2004; Feldman, 1995a), and across cultures (Mesquita, 2003). When people are asked about their experiences of emotion, they unintentionally communicate their core affective feelings (Barrett, 1998, 2004; Feldman, 1995a). All individuals, without exception, are able to distinguish feeling good from feeling bad (Barrett, n.d.), but the same is not true of feeling aroused or sleepy; most, but not all, individuals characterize themselves as energized or slowed down (Barrett, Quigley, et al., 2004; Feldman, 1995a). One hypothesis is that a person's core affective state can be characterized as having some level of arousal (associated with the uncertainty with which a stimulus will predict threat or reward, the need to pay more attention, or the urge to engage in active coping), but people are better or worse at attending to this property of their core affective state and representing it as a feeling. This last idea awaits empirical test.

**Core Affect and Valuation**

Although core affect can have noncognitive causes (e.g., diurnal rhythms, hormonal changes, satiety or hunger, pharmacological agents, too much or too little sleep, etc.; Russell, 2003), changes in core affect largely result from the process of valuation. Organisms continually and automatically evaluate situations and objects (see Bargh & Ferguson, 2000) for their relevance and value—that is, whether or not their properties signify something important to well-being (but for a contrasting view see Storbeck & Robinson, 2004). An object is valuable when it is potentially important to survival (Davis & Whalen, 2001), salient and meaningful (Phan et al., 2002), or relevant to immediate goals (Rogers, 1959; C. A. Smith & Kirby, 2001).

Objects rarely have intrinsic value or meaning (i.e., they rarely act directly on the nervous systems without involving prior learning; see Owen & Bachorowski, 2003; Owen & Rendell, 2001). Typically, meaning is determined by a particular person in a particular context at a particular point in time. This is the basic point made by appraisal models of emotion. People are continually assessing situations for personal relevance, beginning with an evaluation of whether or not the stimulus is "good for me/bad for me" (Arnold, 1960; Lazarus, 1966; Ortony et al., 1988; Roseman, 1984; Scherer, 1984; C. A. Smith & Ellsworth, 1985). The result of this evaluative processing (also called primary appraisal; Lazarus & Folkman, 1984) at each moment in time is some change in a person's core affective state.
Core affect is influenced by a distributed computation of value that derives from the neural circuitry that performs evaluation. Based on the neuroanatomical evidence for connectivity between these brain regions (e.g., Amaral, Behniea, & Kelly, 2003; Amaral & Price, 1984; Gashghaei & Barbas, 2002), as well as a host of behavioral neuroscience research (e.g., LeDoux, 1996; Rolls, 1999), it is clear that evaluations occur continuously at various levels of processing, and simultaneously for multiple objects. Evaluation begins as the thalamus transmits sensory information about the world to the amygdala and the orbital frontal cortex (OFC), which in turn directs attention, autonomic function, and behavior. Evaluation continues into the earliest stages of perceptual processing during object recognition (via connections between the amygdala, OFC, and ventral stream). When a stimulus is first encountered, the available sensory information is extracted, configured, and matched with stored representations in perceptual memory (see Kosslyn, 1995). When viewing a snake, for example, bottom-up (or stimulus-driven) processes extract information about perceptual features of the object. These features constrain one another during a matching process, leading to stimulus recognition (i.e., “there is an object”). The result is that the sensory input implicates a single representation in perceptual memory, and the existence of the stimulus is recognized as familiar or not. Even though the object is not yet named, or identified as belonging to a specific category, its affective value is computed, potentially leading the organism to produce simple, evolutionarily tuned behaviors to deal with threat or reward. As a consequence, even identifying an object as familiar or not can produce an initial evaluation that can influence (and be influenced by) the additional processing as the object is specifically categorized. The evaluation of an object is influenced by later stages of perceptual processing when an object is interpreted (Kosslyn, 1995). This is when conceptual knowledge begins to play a role in affective processing, because the stimulus is assigned to a category (e.g., the object is categorized as an instance of snake), can be named (e.g., “that is a snake”), although it need not be, and an action plan is generated to deal with it (e.g., the perceiver prepares to walk in the other direction). Finally, the evaluation of an object can also be influenced by more complex conceptual processing that alters a person’s behavioral plan. If I would act on my propensity to walk away from the snake, I might actually approach it if I was visiting a zoo and the snake was behind glass. My 6-year-old daughter, on the other hand, would likely approach the snake regardless, because she is very curious about snakes.

The neural consequences of evaluation have a neuromodulatory effect on a wide array of output systems, from those that are typically associated with emotional responding (such as autonomic and endocrine changes, voluntary behavior, facial movements, etc.) to those that are not typically considered to be part of an emotional response (such as selective attention, memory, etc). These output response systems are influenced by nonaffective processes as well (e.g., autonomic shifts occur with simple changes in posture, endocrine changes occur after eating, facial movements occur for social communication), such that there is no specific class of “emotional behaviors,” no specific “action tendency,” facial “expression,” or autonomic nervous system “patterning” that is unique to each kind of emotion. As discussed earlier, freezing may be an innate behavior and may be part of the Western script for fear, such that we automatically categorize freezing as an instance of fear, but this is not evidence that freezing behavior is caused by some module of fear responding or gives evidence of fear. The output of any given response system (behavior, attention, facial movements) is multiply determined and can vary in the extent to which it is caused by the affective value of an object in the environment, such that it can be more or less affectively infused.

Conceptual Knowledge About Emotion

When a state of core affect is categorized with knowledge from the conceptual system, the result is the experience of discrete events that people (at least in English) characterize as anger, sadness, and fear. What people know about emotion and how they use what they know during the categorization process will influence what they feel. For example, compared to individuals who are higher in emotional granularity (and who represent their experiences as discrete emotional events), those who are less granular in their experiences (and who represent their experiences as simple affective states) may be less efficient in categorizing their core affective states, or they may be less motivated to categorize. A complete failure to categorize core affect into the experience of emotion would result in alexithymia (defined as an inability to identify and name their emotional states; Taylor & Bagby, 2000). Cultural differences in experience would also derive from the categorization process, both in terms of what knowledge is brought to bear (i.e., the content of the conceptual system) and whether an affective state is experienced as emotional (i.e., how frequently affective states are categorized). From this viewpoint, questions about the organization, representation, and acquisition of emotion concepts become central to understanding emotion itself.

Most scientific treatments of emotion concepts have assumed that knowledge about emotion is represented in semantic memory as abstract, decontextualized distillations of invariant properties and features. A concept for a particular emotion, say, for anger, is thought to contain a feature list, including what triggers an instance of anger, what it feels like to be angry, the rela-
tional theme likely to be present, the physiological changes to be expected, what voluntary movements, vocal cues, and facial movements are typically involved, and the social rules for expressing the emotion. The concept might also include the causal sequence for these features. It is typically assumed that emotion concepts are ideographically stable so that a given person uses the same representation for anger on different occasions. Emotion concepts are also thought to be nomothetically stable so that everyone within a culture shares roughly the same representation. To know the script for anger in your culture is to know what anger is (Fehr & Russell, 1984).

**Instances of an Emotion Concept Are Context Sensitive**

In the view that I am developing here, however, emotion concepts are not static, invariant representations but instead are constituted as a collection of situated conceptualizations (Barsalou, 1999, 2003, 2005; Barsalou, Simmons, et al., 2003) that guide the emotion categorization process in a much more flexible and situationally sensitive manner. The concept for anger, for example, is not a single representation of information that organized classically (e.g., Johnson-Laird & Oatley, 1989; Ortony, Clore, & Foss, 1987) or as a prototype (Russell, 1991, 2003) that can be retrieved from long-term memory when needed. Instead, an anger concept is a set of representations or packets of conceptual knowledge, each one different from the other. Each conceptualization of anger (or representation of anger) is situated, in that it is a highly specialized package of conceptual knowledge that is tailored to the needs of the person in a given context. This idea is consistent with the broader view that cognition is socially situated (Niedenthal, Barsalou, Winkielman, Krauth-Gruber, & Ric, 2005; E. R. Smith & Semin, 2004).

Context is particularly important to constructing a situated conceptualization of abstract concepts (Barsalou & Wiemer-Hastings, 2005) such as anger. Knowledge for anger is established via a distributed collection of context-specific memories captured across all instances of anger. From the perspective being developed here, this means that knowledge for anger is established by capturing context-specific memories for instances when core affect has been labeled as anger (such as when the category is being learned). A given emotion label (such as anger) is used to refer to a variety of instances. Core affect can be categorized as anger on the highway (when a person might speed up, yell, or shake a fist), in a boardroom (when a person might sit quietly), or on the playground (where a child might make a scowling face, stomp, or throw a toy). In each case, the situational context (both the physical and the relational context) will, in part, determine what behaviors will be performed, such that the context is an intrinsic element of any anger episode.

As a consequence, situated conceptualizations of anger are heterogeneous. Packets of conceptual knowledge about anger will vary within a person over instances as context and situated action demand. No single situated conceptualization for anger need give a complete account of the category anger. There is not one script for anger, but many. On any given occasion, the content of a situated conceptualization for anger will be constructed to contain mainly those properties of anger that are contextually relevant, and it therefore contains only a small subset of the knowledge available in long-term memory about the category anger.

The situation, then, will largely determine which representation of anger will be constructed to conceptualize a state of core affect, with the result that the experience of anger (or of any emotion) will be sculpted by the situation. This idea is, in principle, consistent with the fundamental assumption of appraisal views of emotion: The meaning of a situation to a particular person at a particular point in time is related to the emotion that is experienced.

**Emotion Concepts Are Represented Modally**

Situated conceptualizations are not abstracted from sensorimotor events and stored in some sort of propositional form, like in an encyclopedia (that is, they are not “amodal”). Instead, situated conceptualizations are perceptual symbols, or partial reenactments or simulations of the sensorimotor states that occurred with previous instances of the category. A conceptualization serves to construct, or simulate, the perceptual experience of a category member, bringing an instance of category knowledge into existence.

Evidence for a “modal” view of conceptual knowledge comes from a range of different studies. For example, some studies demonstrate modality-specific processing during cognitive tasks. People are slowed when conceptually processing a statement that describes tasting if on a previous trial they processed a statement describing hearing; this modality-switching cost would only happen if conceptual processing was dependent on sensory processing (Pecher, Zeelenberg, & Barsalou, 2004). Other studies show that activating knowledge about a social concept generates associated bodily states. People walk more slowly after the concept elderly has been activated, and behave more aggressively after the concept “African American” has been activated (Bargh, Chen, & Burrows, 1996).

Some of the most convincing support for a perceptual symbol account of conceptual knowledge comes from studies that report neuroimaging evidence of sensorimotor activations during cognitive acts. For example, a brain circuit that underlies grasping a hammer
becomes active when participants actually grasp a hammer or when they view hammers, but this circuit did not activate when participants viewed buildings, animals, and faces (Chao & Martin, 2000). An area of the brain involved in representing taste became active when participants merely viewed pictures of food but did not actually ingest it (Simmons, Martin, & Barsalou, 2000). Areas in motor cortex involved with leg actions, arm actions, and head actions activated in a specific fashion when participants merely read the words kick, pick, and lick (Hauk, Johnsrude, & Pulvermüller, 2004).

When applied to representing knowledge about emotion, the idea is that the human brain captures every instance of core affect that is labeled as anger. Information is captured as it occurs in perception (represented in sensory cortices), action (represented in motor cortex), and interoception (represented as somatovisceral information in insular cortex). The word occurs is used here to refer to instances where affective behaviors or events are labeled as anger when the category anger is first being learned. Later, these modality-specific states are available to be reactivated to represent knowledge about anger. When retrieving information about anger, sensory, motor, and interoceptive states are partially reinstated in the relevant aspects of cortex, simulating an instance of anger.

More formally, then, a situated conceptualization of anger (that is, an instance of conceptualizing anger) is produced by a simulator (or set of simulators). A simulator for a category of knowledge, like anger, develops as sensory, motor, and somatovisceral features are integrated across instances and settings where instances of the category are identified and anger labeled. Sensory information about the object that is in the focus of attention (e.g., visual information about the person you are interacting with, auditory information about his or her voice, as well as that person’s relation to you in this instance), somatovisceral information about your core affective state (i.e., your current homeostatic state), motor programs for interacting with that person, and for regulating your own core affective behavior, associated with unpleasant, high-arousal states (e.g., facial movements, body movements, loudness of the voice), as well as the label anger (provided by yourself or others), and so on, bind together (via conjunctive neurons; Simmons & Barsalou, 2003) to form an instance of anger. Said more simply, properties that are pointed out by parents (or other speakers) or those that are functionally relevant in everyday activities will bind to core affect to represent anger in that instance. As instances of anger accumulate, and information is integrated across instances, a simulator for anger develops, and conceptual knowledge about anger accrues. The resulting conceptual system is a distributed collection of modality-specific memories captured across all instances of a category. These establish the conceptual content for the basic-level category anger and can be retrieved for later simulations of anger.

Once a simulator for an emotion category, such as anger, is established, it is available to reenact small subsets of its content as specific simulations as needed. When conceptual knowledge about anger is primed, either by an aspect of sensory environment, by a motor action, or by a deliberate need to explain core affect, the simulator becomes active and generates a representation of anger that is tailored to the particular context or situation. For example, the anger simulator might simulate a state of core affect with yelling on one occasion, core affect with running on another, and core affect with crying on yet another. All the experienced content for an anger episode resides within the simulator for anger, so different combinations can be simulated as the situation requires.

### Emotion Concepts Are Directed by Language

A situated conceptualization view of emotion is, on its own, agnostic on whether or not emotions exist as real entities in nature. Whenever people selectively attend with some consistency to components of experience, knowledge of a category develops (Schyns, Goldstone, & Thibaut, 1998). Each time a parent (or some other person) labels a child’s behavior with an emotion term, or a child observes the emotion term being used to label someone else’s behavior, the child extracts information about that instance and integrates it with past information associated with the same term that is stored in memory. It may be that people acquire simulators for anger, sadness, fear, and so on, because these categories reflect the statistical structure of emotional responding that is given by design—that is, people learn to represent emotions in a way that preserves some real, biological distinctiveness (even if such distinctiveness has yet to be discovered). Another possibility, and the one that is advanced here, is that people learn to represent emotion in the way that they learn about other abstract concepts for which there is no biological basis. In this view, language plays a strong causal role in the conceptual development (Gentner & Goldin-Medows, 2003) of emotion knowledge. Children acquire emotion categories that conform to their culture, not because there is some natural, biological reality to anger (or fear or sadness, etc.), but because this level of categorization is socially functional. The emotion words for anger (e.g., angry, hostile, irritated, and so on) serve as the glue that integrates a variety of different sensorimotor states into one category called anger.

Because there remains great debate over the source of conceptual knowledge about emotion (and category knowledge in general), a return to the lessons of color is instructive. Recent experiments using artificial intel-
ligence simulations suggest that language plays an important causal role in the development of color categories (Steels & Belpaeme, 2005). The evidence generated in these experiments suggests that neither biological constraints (the way that our visual system takes in and processes sensory information) nor the statistical structure of the world (the way that wavelengths are distributed in the physical world) can account for how people acquire color categories that are sufficiently shared to allow them to communicate about the colors that they experience. The evidence is more consistent with the notion that language drives the acquisition of color categories. So, too, may it be with conceptual knowledge of emotion categories.

To the extent that language drives category acquisition, it will shape the experience of emotion. Language will help determine which emotion categories people acquire and, therefore, which simulations are available for emotion categorization. Language might also allow people to produce novel simulations (representations of internal states, behaviors, and so on, that have never actually been encountered together; Barsalou, 2003). People may integrate in long term memory two representations from the same emotion category even when their surface features differ, because the label for the emotion links them in memory (see Gelman & Markman, 1987). Highly different instances for the same category can become integrated over time and become available to construct novel simulations that have never been experienced before. This, in part, may help to explain why people believe that emotions like anger, sadness, fear, and so on have specific response signatures. A simulation of anger could allow a person to go beyond the information given to fill in aspects of a core affective response that are not present at a given perceptual instance. In such a case, the simulation essentially produces an illusory correlation between response outputs, helping to explain why researchers continue to search for coordinated autonomic, behavioral, and experiential aspects of an anger episode.

**Emotions as Conceptual States**

Core affect and simulated conceptualizations of emotion are the two components needed for a solution to the emotion paradox. Categorizing the flux and flow of core affect into a discrete experience of emotion corresponds to the colloquial idea of “having an emotion.” Like beliefs and memories, however, emotions are not things (for a discussion, see Gilbert, 1992). They are states. Therefore, it is probably more correct to say that categorizing core affect leads a person to experience their core affective state as emotional. It is also incorrect to think about core affect as a thing that can be perceived. The experience of emotion is not the result of an “inner eye” perceiving an object called “core affect.” Instead, it is probably more correct to say that both valuation and categorization processes change the state of the person to create an emergent product that is at once affective and conceptual. It is very unlikely that these processes proceed in a strict sequential fashion that can be depicted in a flowchart such as Figure 1. Instead, sensorimotor representations of conceptual knowledge fuse with incoming sensory or motor information to seamlessly shape the perception of somatovisceral information about core affect into the experience of emotion, in much the same way that conceptual information seamlessly shapes perceptual processing (e.g., Hochberg, 1998). This process might proceed by constraint satisfaction logic, where affective and conceptual representations constrain each other to reach an overall emotional “solution” (see Barrett, Ochsner, & Gross, in press, for a discussion of constraint satisfaction logic applied to the computation of emotion).

When the category anger is represented conceptually, the neural systems that processed previous instances categorized as anger become active, instantiating an exemplar of the category. The sensorimotor simulation of anger has the potential to fuse with incoming sensory information, as well as with somatovisceral information about core affect, producing some experience of anger. Conceptual information about emotion can be thought of as “top-down” and core affect “bottom-up” constraints on the emerging experience of emotion. Because both category knowledge (i.e., simulations) and core affect share a representational format (both can be characterized as sensorimotor events), they could be seamlessly integrated during an act of categorizing core affect. The idea is that conceptual and affective processing proceed in parallel, with the processing in each limiting, shaping, and constraining the way in which the brain achieves a single coherent “solution”—an instance of experienced emotion that is organized into a coherent interpretation and action plan that suits the particular goals of the individual and constraints of the context. All this occurs in the blink of an eye. The result is an emotional episode that people experience more or less as a gestalt.

At times, core affect may provide more of a constraint on the system, such as when there is a quick, intense shift in core affect where a behavior response is required immediately (such as when a driver cuts you off on the highway). In these instances, categorization may occur after the fact if it occurs at all. At other times, conceptual knowledge may disproportionally constrain the system so that a situated conceptualization might overwrite a person’s current state of core affect, such as in standard emotion induction techniques where participants are asked to imagine a prior experience of emotion. Either way, constructing an emotion is a conceptual act.

There are several key propositions that result from a conceptual act view of emotion. First, the act of catego-
rization performs a kind of figure-ground segregation (Barsalou, 1999, 2003) so that the experience of an emotion will pop out as a separate event from the ebb and flow in ongoing core affect (where core affect is associated with the direction and urgency of initial behavioral responses). In doing so, people divide ongoing changes in core affect into meaningful experiences. Just as categorizing behavior involves parsing a stream of actions into discrete bits by assigning them an intention to render them meaningful, categorizing emotion may involve assigning an intention to an instance of core affect to render it meaningful. Core affect, in philosophical terms, is not intentional, not “about” any thing in particular. Core affect is caused—it represents the state of the person in relation to the immediate environment (in philosophical terms, this is its intension), but “cause” and “aboutness” are not equivalent. When we identify our core affect as being about something, it becomes intentional, and the experience of emotion begins. Just as we interpret or imbue behavioral actions with intention when we parse them into discrete behavioral acts, so we imbue core affect with intention or emotional “aboutness” when we parse it into discrete emotions. Situated simulations naturally link object representations to representations of mental states (Barsalou, 2003, 2005). As a result, during the categorization process, core affect becomes bound to the object that we believe to have caused the affective state in the first place. As a result, we become angry with someone, afraid of something, sad about something.

Second, conceptualizing core affect as emotion not only allows us to determine why core affect is meaningful at a given point in time, but it also allows us to make reasonable inferences about what to do next. It should be possible to research this process. Much of the behavior that we typically think of as “emotional” may result from core affective processes, but it is possible, even likely, that once a situated conceptualization is constructed, it will direct subsequent behavioral responses. In a sense, a situated conceptualization, because it is designed for action, provides you with a script to guide your future behavior in a specific context or situation. For example, across varied situations, different situated conceptualizations of anger will be computed. Sometimes it works to yell, sometimes to pound your fist, sometimes to cry or walk away, sometimes to hit. During a given act of conceptualizing core affect, the simulation can shape a person’s behavior in line with what has been experienced before in that sort of situation (or one very much like it). As a result, situated conceptualizations deliver highly specific inferences tailored to particular situations regarding what actions to take. In addition, conceptual knowledge about emotion constitutes expertise about how to deal with your own internal state—experienced as “an emotion”—and the situation or event that you believe caused that emotion in the first place. In this sense, emotion categorization is functional. Situated conceptualizations may be thought of as an inference about what will make for successful self-regulation or goal achievement. This is generally consistent with several existing ideas about emotion, including the idea that conceptual knowledge about emotion contains information about appropriate or effective forms of emotion regulation (Barrett & Gross, 2001), as well as the idea that emotions are functional for social behavior (Frijda, 1986; Keltner & Haidt, 1999).

Third, conceptualizing core affect as emotion also allows people to communicate their experience to others in an efficient manner. Humans’ major adaptive advantage is to live in social groups, and as a consequence, we have evolved the kind of mind where we attempt to infer the internal states of others (so that we can better predict their behavior), as well as communicate our own internal states to others when it is advantageous to do so. Communication can be thought of as a form of social action. The language and conceptual systems that support communication are tools that can be used to coordinate or direct social action (E. R. Smith & Semin, 2004). Basic-level emotion categories (such as anger, sadness, fear, and so on, in Western culture) may be optimal tools for communicating in the kind of social context that humans typically find themselves in (living in large groups with complicated relational rules). Categorizing an unpleasant state of high arousal as an instance of fear communicates something very different about the perceived cause of that state and implies different behavioral intentions than if the state is categorized as anger. Furthermore, two instances of core affect may be different from each other, but calling them both “anger” allows us to treat them as equivalent for the purposes of communication. In a sense, the words “anger” (or “sadness,” or “fear,” and so on) may be the names that best capture the functional structure of the social world (Brown, 1958).

Fourth, conceptualizing core affect as emotion, like conceptualizing in general, is a skill. Some people may be better than others at tailoring conceptual knowledge to meet the needs of situated action (Barsalou, 2003). This skill for wielding conceptual knowledge about emotion might be considered a core aspect of emotional intelligence. If conceptualizations of a given emotion category lead to the experience of emotion, then constructing such an experience is also a skill. Presumably, there is not one experience of anger, but many, and the one that emerges in a given instance is dependent on the content of the simulation. It is a skill to simulate the most appropriate or effective representation, or even to know when to inhibit a simulated conceptualization that has been incidentally primed. Presumably, this skill not only can be measured, it can also be trained.

Finally, I would suggest that conceptualizing core affect is a skill that we bring to bear when trying to in-
The Reconstruction of Emotion: A New Research Agenda

The conceptual act model of emotion is not a deconstruction of emotional phenomena as much as it is an attempt to sketch the construction of an emotional event from basic psychological processes and their neural correlates. Taken together, the basic propositions of the conceptual act model of emotion map a novel research agenda for the psychological construction of emotion with several distinctive features. First, it hypothesizes that the basic building blocks of emotional life are conceptual and affective. The evolutionary legacy to the newborn is not a set of modular emotion circuits that are hardwired into the brain, but rather a set of mechanisms that compute valuation and core affect, as well as those that allow category learning. In principle, it should be possible to study how the two combine to produce the experience of emotion. Moreover, it should also be possible to examine how affective and conceptual changes configure to produce emotional change (as in the treatment of emotional disorders). Some treatments (such as the nonspecific affects of the client-therapist relationship or pharmacotherapy) may be more effective at producing affective relearning, whereas others (such as the emotion resocialization that is thought to occur in more cognitive forms of psychotherapy) might be more likely to produce conceptual change, leaving affective responding untouched (Quigley & Barrett, 1999).

Second, the conceptual act model focuses on the need to understand the richness and diversity of emotional life in humans. In doing so, it will move the science of emotion away from its current focus on a small set of canonical forms that are seen rarely in everyday life. Specifically, it leads us to expect, rather than to treat as error and explain away, variations in the form and functions of emotional episodes. It not only helps to explain why some individuals are better able to distinguish discrete emotional states than are others (i.e., why they differ in emotional granularity), but it also predicts that any emotion, such as anger, will differ from one instance to the next, even within the same person.

Third, the conceptual act model rescues the experience of emotion from obscurity. Experience is no longer considered as epiphenomenal to the scientific study of emotion, but is given a central place in characterizing what emotions are. At its core, the conceptual act model of emotion assumes that emotion categories do not have ontological status separate from our perception of them.

Fourth, the conceptual act model suggests an intrinsic role for language in the emergence of emotional events. It is consistent with a strong version of linguistic relativity (Whorf, 1956), that is, that language forms the basis of experience. In the case of emotion, language shapes core affective phenomena into the emotional reality that we experience. Language not only enters into the categorization process, but it also directs the development of emotion category knowledge in the first place (by guiding which nonlinguistic information is included in an emotion category as it is being constructed during the learning process). As a result, the conceptual act model provides a means for understanding the role of language in cultural, as well as in individual, differences in the experience of emotion.

Finally, the conceptual act model leads us to reflect on why scientists typically theorize about and focus their empirical efforts on prototypical emotional episodes (i.e., what most people consider the clearest cases of emotion that necessarily have all of the component parts; Russell, 2003; Russell & Barrett, 1999), even though such episodes are quite rare and the nonprototypical cases are more frequent in our everyday lives. The answer may be that it is a natural consequence of the way that categories work. Emotion categories can be thought of as goal-directed categories that develop to guide action. The most typical members of a goal-directed category are those that maximize goal achievement, not those that are most frequently encountered (Barsalou, 2003; Barsalou & Ross, 1986). As a result, the most typical instances of a category contain properties that represent the ideal form of the category—that is, whatever is ideal for meeting the goal that the category is organized around—not those that most commonly appear as instances of the category.

Summary

In suggesting that emotions are experiences constructed with a conceptual act, I have argued that the way people learn about emotion categories and use conceptual knowledge determines what they see and feel. Variation in conceptualizing an instance of emo-
tion, whether because of language use, context, culture, or individual differences in prior experience, will produce variation in which emotion is experienced and how it is experienced.

At a sensory level, people have a continuous stream of homeostatic feedback from the body that delivers affective information about their current relation to the world. It is not a specific interoceptive readout of autonomic activity or anything so precise. Rather, it is a core affective state that gives rise to feelings of displeasure (or pleasure) and activation (or deactivation) that results from many sources, including ongoing automatic evaluations or primary appraisals of the world. The way that people conceptualize their affective state will depend on the knowledge about emotion that they bring to bear when categorizing it. Knowledge about emotion is context dependent, represented by sensory, motor, and somatovisceral information, and driven by emotion language. A person might experience his or her core affective state as a particular sort of sadness, anger, or nervousness, depending on the conceptual knowledge that he or she brings to bear in that situation. Categorizing core affect in this way is functional. It changes core affect into a meaningful experience, allowing people to make inferences about what caused the state, and how to deal with the situation. Emotion categorizations also allow people to efficiently communicate their experiences of core affect to others. Categorizing core affect into the experience an emotion can proceed with more or less skill. It is a skill that we bring to bear when representing and communicating about our own internal states, as well as the internal states of others, including nonhuman animals.

Finally, I suggested that the conceptual act model of emotion has explanatory power for understanding the richness and diversity of emotional life. It leads scientists to ask different questions about emotion and suggests new ways to take account of the evidence that already exists.

**Alignable Differences**

Thus far, I have argued conceptual knowledge and core affect, rather than discrete emotions, are the building blocks of emotional life. From this view, the experience of emotion does not issue from discrete bursts of activity in putative emotion mechanisms, but rather from an act of conceptualization. We experience anger, sadness, fear, and so on, when we categorize our core affective state in a given instant. In Table 1, I briefly contrast the categorization model of emotion to the natural-kind model according to a set of alignable differences, to show where the models are similar and where they diverge.

**Standing on the Shoulders of Giants**

The conceptual act model of emotion builds on key ideas from the psychological literature of emotion. To begin, it builds on the natural-kind view of emotion. The idea that emotions are real biological entities rescued the topic of emotion from the ashes of behaviorism, inspiring the scientific study of emotion for several decades (Ekman, 1992). In doing so, it produced the evidentiary basis for the emotion paradox. The conceptual act model of emotion, as one solution to that paradox, also has elements in common with other theoretical perspectives on emotion, although it cannot be reduced to any single one. I briefly highlight a few perspectives here.

**William James**

The idea of emotion as a perception began with William James. In the most general terms, James (1884, 1890/1950, 1894/1994) suggested that the experience of emotion (which he merely called *emotion*) results from the self-perception of information in the body. He focused primarily on the ways that the peripheral nervous system reacts to the external environment, producing somatovisceral and voluntary muscle activation that is sensed via interoceptive mechanisms, constituting the experience of emotion. A variety of modern accounts are inspired by the idea that the experience of emotion is a self-perception (e.g., Damasio, 1994; Dolan, 2002; Laird & Bresler, 1992; Russell, 2003), although they vary in the type and specificity of processes proposed. The perspective I am advancing here is very similar to what James actually wrote in certain ways, but is less consistent with how his ideas have been characterized by others.

James’s (1884, 1890/1950, 1894/1994) ideas can be broken down into several hypotheses that are consistent with the view that the experience of emotion is core affect plus emotion category knowledge. First is the idea that a stimulus situation will evoke an organismic response in an automatic, prespecified way, much as a key opens a lock. This is similar to the idea that core affective state, and its associated behaviors, can issue from the evaluative processing resulting from the early perceptual processing involved with object recognition.

Second, James (1884, 1890/1950, 1894/1994) proposed that the peripheral nervous system changes that result from reactions to the stimulus situation produce the antecedents, rather than consequences, of an emotional state. This is similar to the idea that core affective state is available for categorization via emotion knowledge. Furthermore, the extent to which such knowledge is embodied provides another way for sensorimotor representations to constitute, rather than result from, an emotional response.
### Table 1. Alignable Differences For The Natural–Kind and Psychological Construction Views of Emotion Experience

<table>
<thead>
<tr>
<th></th>
<th>Natural-Kind View</th>
<th>Categorization View</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>How many kinds of emotion?</strong></td>
<td>There are a few privileged emotion kinds (at a minimum, anger, sadness, fear, disgust, and happiness).</td>
<td>There are no natural kinds of emotion. A person parses core affect into categories based on the concepts that he or she has learned.</td>
</tr>
<tr>
<td><strong>Brain mechanisms</strong></td>
<td>Specific kinds of emotion are produced by distinct, dedicated neural circuits.</td>
<td>There are two distributed neural systems of concern—one for computing valuation and the other for implementing conceptual knowledge about emotion.</td>
</tr>
<tr>
<td><strong>Role of evolution</strong></td>
<td>The neural circuitry for discrete emotions is inborn and homologous in nonhuman mammals.</td>
<td>Humans have an inborn capacity for core affect that is homologous in nonhuman mammals. Humans also have the innate ability to learn categories, although the emotion categories they learn are not biologically given. Humans may also have an innate propensity to categorize their internal states, and the behaviors of others, because we evolved the kind of minds where we need to be proficient at communicating our own internal states and inferring those states in others.</td>
</tr>
<tr>
<td><strong>Distinctiveness</strong></td>
<td>Different emotion experiences can be distinguished at the sensory level.</td>
<td>Different emotion experiences are distinguished at the conceptual level (although there is probably no strict boundary between conceptual and perceptual representations).</td>
</tr>
<tr>
<td><strong>Role of top-down processing</strong></td>
<td>Top-down processing has little or no influence on the experience of emotion.</td>
<td>Top-down processing is necessary, but not always sufficient, for the experience of emotion.</td>
</tr>
<tr>
<td><strong>Organization of conceptual knowledge</strong></td>
<td>Categories can be classically organized, either as semantic representations or as embodied representations that show little situational variation within a category. Or, categories can be situated conceptualizations that are learned (permitting that people can learn to recognize and anticipate display rules or other forms of regulation that introduce variation into emotional responding).</td>
<td>The conceptual system is organized as situated conceptualizations that are context dependent.</td>
</tr>
<tr>
<td><strong>Origin of category knowledge</strong></td>
<td>Emotion concepts are either inborn or are derived from the statistical structure of emotional responding as it actually occurs.</td>
<td>The acquisition of emotion knowledge is language dependent.</td>
</tr>
<tr>
<td><strong>Variation in conceptual knowledge</strong></td>
<td>All humans have the same conceptual system for emotion that largely contains the same content.</td>
<td>Different learning histories produce conceptual systems for emotion that differ in content.</td>
</tr>
<tr>
<td><strong>Emotion perception</strong></td>
<td>Anger, sadness, fear, and so on, are real mechanisms that cause behavior, but people perceive them accurately or inaccurately.</td>
<td>There is no clear empirical criterion for judging the occurrence of anger, sadness, fear, and so on, so accuracy cannot be assessed. Instead, it is possible to assess whether the conceptual knowledge brought to bear is normative for a given situation within a given culture.</td>
</tr>
<tr>
<td><strong>Emotional granularity</strong></td>
<td>Differences in emotional granularity are produced by differences in the accuracy of self-reports. All people experience anger, sadness, fear, and so on, but only some report these experiences.</td>
<td>Differences in granularity are produced by differences in the conceptual knowledge that people bring to bear when categorizing core affect. Not everyone experiences anger, sadness, fear, and so on.</td>
</tr>
<tr>
<td><strong>Cultural variation</strong></td>
<td>There is cultural variation in the objects that trigger emotion and in the display rules that control the expression of emotion.</td>
<td>There is cultural variation in the experience of emotion that is intrinsically driven by cultural differences in emotion categories and concepts.</td>
</tr>
</tbody>
</table>
The idea that sensorimotor activations constitute the experience of emotion drives one of the most compelling ideas in the psychology of emotion: that emotional states have specific and unique patterns of somatovisceral changes, and the perception of these bodily events constitutes the experience of emotion. James is credited with the idea that invariant autonomic nervous system patterns and behaviors correspond to anger, sadness, fear, and so on. Interestingly, he did not appear to hypothesize invariant autonomic nervous system patterns for each category of emotion, however. In fact, he explicitly rejected the idea that there is a single set of bodily symptoms to describe instances of a given emotion category across individuals: “Surely there is no definite affection of ‘anger’ in an ‘entitative’ sense” (1894, p. 206).

Although James’s (1884, 1890/1950, 1894/1994) writings are laced with detailed descriptions of the bodily symptoms that characterize anger, grief, fear, and the like, he explicitly stated in several places that variability within each emotion category is the norm. More important, James argued for the heterogeneity of instances within each emotion category. According to James, there can be variable sets of bodily symptoms associated with a single category of emotion, making each a distinct feeling state and therefore a distinct emotion. By the term emotion, James is referring to particular instances of feeling, not to discrete emotion categories. Different instances of an emotion, even if within the same category, will feel different if the somatovisceral activations are different. The idea of heterogeneity within an emotion category is very consistent with the view offered here, but it is less consistent with other embodiment views, such as Damasio’s (1994) somatic marker hypothesis, where specific emotion categories correspond to specific somatic markers (Damasio et al., 2000).

**Schachter and Singer**

The account of emotions as conceptual acts clearly owes a large intellectual debt to Schachter & Singer (1962) on two counts. First, they were the first to really challenge the idea that emotions were natural kinds, without redefining emotion out of existence. Second, they introduced the idea that a basic biological state (a general increase in sympathetic nervous system activation) was interpreted as meaningful by cognitive processes that are contextually driven (so that information from the surrounding environment is used to interpret the meaning of that general physiological arousal). Appraisal models of emotion have generally incorporated the idea that situation drives the cognitive processes that determine the experience of emotion. Other cognitive models of emotion, such as that proposed by Mandler (1975, 2002), develop this idea as well.

The differences between the Schachter & Singer (1962) model and the conceptual act model offered here are threefold. First, the idea of emotions as conceptual acts does not posit a general state of physiological arousal that must be consciously experienced before it can be labeled after the fact. Second, the cognitive influences on feeling need not be deliberate, nor are they necessarily produced by attribution. Context shapes the experience of emotion in a more direct and automatic fashion. Third, the cognitive processes of import are those involved with categorization. Separate cognitive mechanisms for determining meaning are not necessary. The conceptual act model of emotion is probably closer to what Mandler (1975, 2002) proposed when he suggested that the phenomenological quality of felt emotions arises from visceral activity (in response to sensory input from the environment) that is evaluated in light of the situation.

### Appraisal Models

The conceptual act model of emotion is similar in some respects to the level of processing views put forth by some appraisal theorists (Leventhal & Scherer, 1987; Power & Dalgleish, 1997; C. A. Smith & Kirby, 2001; Teasdale & Barnard, 1993). It is also broadly consistent with an appraisal perspective on emotion, to the extent that the content of appraisals can be seen as resulting from conceptual knowledge about emotion. What differentiates the conceptual act model from most appraisal perspectives is the emphasis on categorization processes as a core mechanism driving emotion experience. Separate cognitive mechanisms for computing a situation’s meaning (as found in some appraisal models, e.g., Lazarus, 1991; Roseman, 1984, 1991; Scherer, 1984) are not necessary to account for the experience of emotion. In this sense, the conceptual act model of emotion is more similar to appraisal models where appraisals are not literal cognitive mechanisms for assessing meaning, but instead represent dimensions of meaning that are associated with particular emotions (e.g., Frijda, 1986; Ortony, Clore, & Collins, 1988). In these models, the appraisals describe the set of rules for which emotions are felt, but they are agnostic as to the cognitive processes that compute meaning. The conceptual act model of emotion is probably closest to the ideas presented in Clore & Ortony (2000), where a situation’s meaning can be either reinstated or computed anew during each instance of conceptualization.

### The Core Affect Model

The idea that the experience of emotion is constructed from more basic psychological processes further develops Russell’s (2003) ideas regarding the psychological construction of emotion. The conceptual act model of emotion is similar to Russell’s (2003) psy-
chological construction model in three ways. First, emotional life is thought to consist of the continuous fluctuations in core affect. Second, ongoing changes in the face, the body, and the behavior are not necessarily coordinated into specific response profiles for discrete emotions. And third, core affect itself is subject to further meaning analysis.

The differences between the position advanced here and the Russell (2003) model are threefold. First, the conceptual act model specifies a broader, more fundamental role for categorization in the construction of emotion. In Russell’s (2003) view, the face, body, and behavior are associated with a core affective state and occasionally take the form of a prototypical emotion pattern, such that they are then categorized as an emotion. In a sense, Russell (2003) argued that emotions such as anger, sadness, and fear are superstitious perceptions (Gosselin & Schyns, 2003). In contrast, the conceptual act model suggests that core affect is not categorized as an emotion only when core affect is similar to a prototype, but rather that categorization has a more pervasive and intrinsic influence so that even nonprototypical conceptual representations will be used to conceptualize core affect as emotional when the situation demands it.

Second, the conceptual act model relies on emotion concepts that are more contextual and socially situated, rather than on static, symbolic representations such as prototypes. From a situated conceptualization viewpoint, prototypes do not exist as stored representations in memory, but can be constructed (or simulated) when needed (Barsalou, Niedenthal, et al., 2003).

Finally, in the conceptual act model of emotion, the binding of an object to core affect is a natural consequence of the categorization process, whereas Russell (2003) discussed this as an attributional process. Although intense core affective feeling may invoke a deliberate search for its cause, leading to its categorization as emotional, more often the feeling of emotion arises automatically and effortlessly from the conceptual processes involved in categorizing that core affective state.

Situated Conceptualizations of Emotion Concepts

Niedenthal and colleagues (Barsalou, Niedenthal, Barbe, & Ruppert, 2003; Niedenthal et al., 2005) have recently reinterpreted a good deal of evidence from existing research on attitudes, social perception, and emotion from a situated conceptualization view, to make a persuasive case that social knowledge (including emotion knowledge) is grounded in context-sensitive representations. In particular, their discussion emphasized the role of sensorimotor processing in the acquisition and performance of social knowledge. Their careful analysis of the organization and representational format of the emotion conceptual system figures prominently in the conceptual act model of emotion, although the model goes beyond their analysis in several respects (e.g., considering the role of language in orchestrating situational conceptualizations, as well as the role of situated conceptualizations in the experience of emotion).

Conclusions

People rely on their experiences to inform them about the world. We conduct ourselves as if experience gives us direct access to the world around us, on the assumption that the world as we feel (hear, taste, see, or smell) is identical to the physical world that exists apart from us. People see (at least in Western cultures) anger, sadness, fear, and so on, in other people’s behavior, and experience it themselves, leading to the assumption that these emotions are entities lurking somewhere within the brain or body. People assume that their experiences of emotion reveal an unbiased, internal reality. Many scientific models of emotion make similar assumptions, such that emotions such as anger, sadness, fear, and so on, are given by design. When verbal reports of emotion experience fail to support the idea that anger, sadness, fear, and so on, are given categories, it is assumed the data themselves are flawed.

A brief review of the scientific evidence indicates that when it comes to emotion, our experiences might be misleading. In this article, I reviewed scientific evidence to show that there is less than sufficient support for the idea that emotions exist as discrete entities. This has left scientists with a paradox: People from Western cultures experience anger, sadness, fear, and so on, even though scientists are without a clear empirical criterion for judging when these emotions have occurred. In the remainder of this article, I outlined a potential solution to the emotion paradox: When people have an experience of emotion, they categorize core affect with the help of conceptual knowledge about emotion via a process that is very similar to color perception and person perception.

The conceptual act view of emotion implies certain counterintuitive ideas. Although core affect is sensorially given and can be observed in behavior, emotions only exist when they are experienced (although this experience is not always represented in awareness). Emotions are not causal entities, but are states that are caused and can be explained. Affective responses are real. Their configuration in a given instance, whatever it is, is also real, but a particular configuration of face, body, and behavior is not required for a given instance of an emotion category. An emotional episode may involve specific behaviors (over and above the core affect, but emotions do not cause behavior.)
Although there is no direct evidence for the model that I have outlined here, there are reasons to grant its plausibility. First, there is some scientific precedence for dispensing with the view of emotions as entities and hypothesizing that emotions are states constructed from more basic psychological processes. Many psychological constructs that we once thought of as fixed, unitary causal entities with an identifiable essence (e.g., memory, personality, attitudes, concepts) are now thought of as emergent phenomena or by-products of distinct but interacting systems. Second, the conceptual act model proposes that the experience of emotion is constructed via psychological processes that are well established in other areas of psychology. The ability to explain emotion with familiar processes is not only a strength of the model, but it also contributes to psychology’s development as a cumulative science. Third, there is accumulating scientific evidence from different levels of inquiry from the brain to behavior that something like core affect exists, and that architecturally distinct circuits for discrete emotional episodes may not. Finally, there is a long empirical tradition demonstrating that category knowledge shapes the perception of color and the perception of other people, as well as more direct evidence that conceptual knowledge about emotion, in particular, shapes the emotions that we identify in others.

Of course, the value of this perspective will rest or fall with direct empirical, rather than circumstantial, evidence. Undoubtedly, a major research initiative is required to test these ideas. But this sort of research program will only be carried out if psychological scientists are willing to question the traditional view of emotions as entities, and take seriously the idea that emotions may not be real categories with mechanisms that are hard-wired into the brain, waiting to be found with better research tools. Doing so may allow us to ask different, and perhaps better, questions about the nature of emotion.

References


THE EMOTION PARADOX


BARRETT

Green (Eds.), Crying as a sign, a symptom, & a signal (pp. 8–22). New York: Cambridge University.  


James, W. (1884). What is an emotion?  


Null


