CHAPTER 14

Basic Emotions, Psychological Construction, and the Problem of Variability

ANDREA SCARANTINO

The most influential recent challenge to the scientific viability of basic emotion theory (BET) comes from psychological constructionists, who have argued that, contrary to BET's claims, the empirical data do not support the existence of coordinated packages of biological markers—either in the body or in the brain—associated with candidate basic emotions such as anger, fear, happiness, disgust, and so forth (Russell, 2003; Barrett, 2006a, 2006b).

The take-home message of this chapter is that the empirical data undermine what I call *Traditional BET* but are not fatal to BET as a research program. I will show how a *New BET* can be formulated so as to answer the constructionist critique (see also Scarantino, 2012a, 2012b, 2012c; Scarantino & Griffiths, 2011). Furthermore, I will highlight some challenges that psychological constructionists face in their quest to account for emotion episodes without invoking specialized emotion mechanisms.

Seven Commitments for Traditional BET

BET finds inspiration in Darwin's (1872) theory of emotional expressions.[1] In the 1960s, Darwin's evolutionary approach was embedded into a general theory of affects by Tomkins (1995), and powerfully articulated into modern-day BET by Ekman (1980, 1999), Izard (1977, 1992), Panksepp

1. ***If emotions are psychological events constructed from more basic ingredients, then what are the key ingredients from which emotions are constructed? Are they specific to emotion or are they general ingredients of the mind? Which, if any, are specific to humans?***

According to the version of basic emotion theory I present in my contribution to this volume – the new BET – basic emotions are evolved programs that coordinate more basic ingredients such as facial muscle responses, autonomic blood flow, subjective experiences, respiratory and vocal changes, motor patterns, thoughts, memories, images, and so on. In this limited sense, the new BET is *compositionally constructionist*, because emotions are analyzed in terms of more elemental building blocks (programs, physiological changes, expressions, behaviors, etc.). I emphasize that the only building block specific to emotion is the program associated with each basic emotion. All building blocks elicited and coordinated by the program are instead also involved in non-emotional processes. For example, motor patterns and autonomic blood flow are involved not only in basic emotions but also in the non-emotional activity of doing push-ups. Some of the ingredients of human basic emotions are specific to humans, because they reflect capacities only humans have. Three ingredients of human basic emotions that are good candidates for being specific to humans are conceptual thoughts (e.g., "This is the third time I have been disrespected by my boss in front of my wife"), imaginings (e.g., to imagine hitting the boss without doing it), and at least some subjective experiences (e.g., the meta-experience of perceiving oneself as being angry at the boss). On the other hand, the basic emotion programs that elicit and coordinate emotional responses exist not only in humans but also (in homologous form) in related species. Finally, since I do not consider basic emotions to exhaust the domain of emotions, I am open to the possibility that there may be some nonbasic emotions that are exclusively human (e.g. regret).

2. ***Which processes bring these ingredients together in the construction of an emotion? Which combinations are emotions and which are not (and how do we know)?***

According to the new BET, what brings the ingredients of a basic emotion together is an evolved and specialized basic emotion program. The program was selected to coordinate organismic resources to deal successfully with fundamental life tasks such as avoiding dangers, removing obstacles, coping with losses, and so on. In this view, what is distinctive about basic emotions is not the specific responses they involve but the programs that recruit such responses in a task-oriented and (often) context-dependent fashion.

There are two ways in which we can determine whether a certain combination of ingredients counts as an emotion or not. The first is through what I call the *Folk Emotion Test*. The test is passed by a certain configuration *C* of ingredients insofar as *C* is sufficiently similar to the prototype (or exemplar, or other suitable construct supported by psychologists of concepts) of

(continued)

a vernacular emotion such as *anger, fear, regret, guilt, shame, awe*, and so forth. Passing the Folk Emotion Test entails being an emotion in the ordinary sense of the term. The second is what I call the *Natural Emotion Test*. The test is passed by a certain configuration *C* of components insofar as *C* satisfies the condition of membership of a natural emotion kind, which is (roughly) a theoretically homogenous class of emotions about which a great many explanatory and predictive generalizations can be formulated in affective science. The account of basic emotions I offer in this chapter aims to individuate a combination of ingredients that passes the Natural Emotion Test. I have also argued that passing the Folk Emotion Test is not sufficient for passing the Natural Emotion Test. This is to say that a certain combination *C* of ingredients can qualify as fear, or anger, or happiness in the ordinary sense, without individuating a natural kind of emotion. This is due to the heterogeneity of folk emotion categories, which collect members that are arguably too different from one another for purposes of scientific investigation. For this reason, I have recommended that basic emotion theorists stop designating basic emotions using folk psychological emotion categories, which fail to capture natural kinds.

3. ***How important is variability (across instances within an emotion category, and in the categories that exist across cultures)? Is this variance epiphenomenal or a thing to be explained? To the extent that it makes sense, it would be desirable to address issues such as universality and evolution.***

I have distinguished among three types of variability: variability across instances of a given folk emotion category, variability in the folk emotion categories that exist across cultures, and variability in the manifestation of a basic emotion. All forms of variability must be explained. But it is important not to confuse them. The variability we find within folk emotion categories such as "fear" or "anger" is a threat to their scientific status, because it stands in the way of formulating scientific generalizations that apply to all members of the category. The variability we find across cultures, instead, has no impact on the scientific status of folk emotion categories, because it does not affect the set of members such categories contain in any given language. The variability we find in the manifestation of a basic emotion, finally, has two main sources. The first is that basic emotions evolved to deal with fundamental life tasks that take on different forms, and require different responses on different occasions to be dealt with successfully. As a result, basic emotion programs need to be able to produce flexible responses. The second is the fact that basic emotions interact with other emotions and other mental states, which also leads their manifestations to differ from occasion to occasion. Because of these two sources of variability, the new BET holds that what is universal when it comes to basic emotions are first and foremost the evolved programs that run them. On this view, the program evolved to deal with dangers or the program evolved to remove obstacles are universals, in the sense that they are found in every culture and in related species in homologous form. But the specific

(continued)

manifestations of the activation of such programs are not necessarily universal. This marks a key difference with traditional BET, which has looked for universality in the actual bodily and neural changes associated with each basic emotion. Although some universal responses are going to be found in the (rare) occasions in which the elicitors of a basic emotion activate a rigid cascade of mandatory responses, most instances of basic emotions lack mandatory outputs and will only lead to irruptive and prioritized response tendencies whose actual manifestations are highly variable and context-sensitive.

4. ***What constitutes strong evidence to support a psychological construction to emotion? Point to or summarize empirical evidence that supports your model or outline what a key experiment would look like. What would falsify your model?***

I consider psychological constructionism to include both a positive program and a negative program. The negative program holds that there are no hard-wired emotion mechanisms uniquely associated with anger, fear, happiness, sadness, disgust, and so forth that are causally responsible for coordinating patterns of tightly associated components with a one-to-one correspondence with the relevant folk emotion categories. I think this component of psychological constructionism is strongly supported by the empirical evidence on variability with respect to the neural circuitry, physiological changes, expressions, behaviors, and phenomenological changes associated with anger, fear, happiness, sadness, disgust, and so forth.

The positive program of psychological constructionism would need to achieve two objectives in order to be fully supported: accounting for the relevant empirical data better than all competing models and providing a viable account of how emotion episodes come about without invoking specialized emotion mechanisms. I have argued in this chapter and in Scarantino (2012a) that, although both Russell's version and Barrett's version of psychological constructionism account for the empirical evidence better than traditional BET, they do not provide a viable account of how emotion episodes come to be.

I have offered the new BET as an alternative model that combines insights from the constructionist camp with insights from traditional BET. The new BET, just like psychological constructionism, predicts that instances of folk emotion categories like anger, fear, happiness, sadness, and disgust, will not all manifest the same neurophysiological signature, nor will display high correlation between changes in physiology, expressions, behaviors, phenomenology, and so forth. In this sense, the new BET is compatible with the empirical data that falsify traditional BET.

Unlike psychological constructionism, however, the new BET makes a further prediction, namely that we will find evidence for the existence of specialized emotion programs designed to solve evolutionary problems in a context-dependent and yet task-oriented fashion. The new BET is going to be empirically supported insofar as we can find evidence in three domains (conversely, absence of evidence in all such domains would falsify the new BET):

(continued)

(i) *Evidence of hard-wiredness*: Evidence that there are hardwired neural circuits designed to orchestrate solutions to evolutionary problems such as dealing with dangers, removing obstacles, expelling noxious substances, suffering losses, and so forth.

(ii) *Evidence of distinctness and continuity of responses*: Evidence that the presentation of focused/powerful/sudden/prototypical elicitors leads to a cascade of highly coordinated responses specific to distinct basic emotions, functional to the solution of a fundamental life task and continuous across cultures and species

(iii) *Evidence of distinctness and continuity of functional variants*: Evidence that the presentation of less focused/ powerful/sudden/prototypical elicitors leads to context-dependent yet task-oriented combinations of functional variants that are specific to distinct basic emotions and continuous across cultures and species

Although the bulk of recent empirical research has been aimed at either confirming or falsifying traditional BET, some of the evidence collected so far can be interpreted as supporting the new BET, and I predict that significantly more evidence in its favor will be forthcoming.

Re (i): We have evidence for the existence of hardwired circuits for orchestrating responses to the sorts of challenges basic emotions evolved to solve, including distinct circuits for responding to unconditioned threats and conditioned threats (Le Doux, 2012), distinct circuits for producing defensive aggression and predatory aggression (Moyer, 1976; Blanchard & Blanchard 2003), distinct circuits for responding to body-boundary violations and repulsive foods (Harrison, Gray, Giarnos, & Critchley 2010), and many others (see Panskepp, 1998 and LeDoux, 2012, for two alternative taxonomies of adaptive neural circuits shared across mammals). These circuits combine with learning and other forms of higher cognition to give rise to the full panoply of context-dependent manifestations of basic emotions.

Most of these circuits have been studied in rodents, but there are reasons to think that they are at least to some extent conserved across species and found in similar form in humans (Le Doux, 2012; Panskepp, 1998). This being said, it is an open question whether any phenomenological changes are associated with the activation of such circuits in non-human animals. What is quite clear is that multiple circuits can underlie solutions to the same evolutionary life task, which suggests the need for moving away not only from folk psychological categories (there is no neural circuit for fear, no neural circuit for anger, etc.), but also from simple basic emotion categories. (There are multiple circuits underlying basic fear, so we should distinguish between conditioned basic fear and unconditioned basic fear; there are multiple circuits underlying basic anger, so we should distinguish between defensive basic anger and offensive basic anger; and so forth.)

Re (ii) & (iii): The study of the evidence for distinctness has been set back by failure on the part of basic emotion theorists to acknowledge that basic

(continued)

emotion programs have two different modes of operation. When the basic emotion program is activated by elicitors that are "focused, powerful, sudden, and closely match prototypical antecedent conditions" (Levenson, 2011, p. 382), it coordinates a rigid cascade of facial, autonomic, behavioral, and phenomenological changes of the sort posited by traditional emotion theorists.

Testing for this hypothesis requires exposing individuals to elicitors such as loud sounds or deadly predators for basic fear; physical restraint or sudden pokes in the back for basic anger; dead insects in one's soup or feces for basic disgust; and so on. What the new BET predicts is that in such cases, and only in such cases, the basic emotion program will lead to a rigid cascade of responses that are specific to each basic emotion.

For basic fear, it is predicted for instance that exposure to a charging bear will lead humans to manifest autonomic changes such as increased heartbeats, increased myocardial contractility, vasoconstriction, and electrodermal activity (Kreibig, 2010); expressions such as eyes and mouth open, lips retracted, and eyebrows raised (Matsumoto et al., 2008); behaviors such as freezing and then running away and unpleasant feelings. These responses are expected to be found in different cultures and across species, but they are not expected to be found in all manifestations of basic fear, let alone in manifestations of fear that are not basic.

The best evidence we have for continuity of responses across cultures and species concerns facial expressions. There is evidence that the kinds of "exaggerated"[2] facial responses associated with basic emotions in prototypical cases are recognized beyond chance by perceivers from different cultures (Ekman, 1972; but see Russell, 1994 for worries about the methods used), involve similar facial muscles across cultures (Ekman, 1972; Matsumoto & Willingham, 2006), are displayed by children born blind (Matsumoto & Willingham, 2009), and have homologous expressions in non-human primates (Chevalier-Skolnikoff, 1973; Waller et al., 2006).

The mistake of traditional basic emotion theorists has been to suggest that the "exaggerated" expressions associated with basic emotions in the presence of focused/powerful/ sudden/prototypical elicitors are universally found whenever "fear" or "anger" or "disgust" are activated. This prediction has been widely and rightly debunked by psychological constructionists, who have demonstrated that the facial expressions associated with fear, anger, disgust, and so forth vary with cultures; are perceived differently depending on context; and are often not present at all in fear, anger, disgust, and so forth (see Russell, Chapter 8, this volume).

As noted by Levenson (2011, p. 382), the sorts of elicitors used in social science experiments "are typically mild, gradual in onset, diffuse, and do not closely match prototypes."When the basic emotion program is activated by elicitors that are less focused/powerful/ sudden/prototypical, we should expect it to coordinate a flexible, yet task-oriented, combination of facial, autonomic, behavioral, and phenomenological changes.

(continued)

In this case, there will be a range of what I have called *functional variants* (i.e., different ways in which facial, behavioral, autonomic, and phenomenal changes can be manifested while preserving the task-oriented nature of the response). What psychological constructionists have in my view not fully appreciated is that the variability characteristic of basic emotions often involves functional equivalents. For instance, it is true that we can respond to dangers by running away from a bear, freezing when seeing a snake, shooting through a closed door toward an intruder, and by calling a doctor when feeling gravely ill, but these are all functional variants of avoidance behaviors.

Although it is true that basic emotions will involve different behaviors, and with them different patterns of expressions, autonomic changes, and feelings, basic emotions can still be associated with distinct patterns of functional variants that will be continuous across cultures and species. Further support of the new BET will come from studies specifically designed to unveil the functional variants associated with the same basic emotion, namely the sets of facial expressions, sets of autonomic changes, sets of behaviors and sets of phenomenological changes that tend to be associated with the activation of a given basic emotion program in different contexts.

Central empirical challenges for future basic emotion theory will include figuring out which contextual cues determine which specific functional variants are instantiated, what patterns of mutual change functional variants display, how such patterns are related to the notion of emotional intensity for basic emotions, and how basic emotions, nonbasic emotions and core affect are related.

(1982, 1998), Levenson (1988, 1992), and many others. Individual authors differ on several dimensions of analysis, but they share a number of commitments that define BET as a research program.[3] I present the following seven commitments as lying at the heart of *traditional BET*, the version of BET most prominently associated with Ekman and his colleagues:

- **TB1:** ***Basic emotions are evolutionary adaptations, selected for because they are efficient solutions to fundamental life tasks.*** As Ekman and Cordaro (2011, p. 364) put it, "Each basic emotion prompts us in a direction that, in the course of our evolution, has done better than other solutions in recurring circumstances that are relevant to our goals." The proposal is that basic emotions were selected over evolutionary time because they are efficient solutions to fundamental life tasks such as dealing with dangers, removing obstacles, suffering losses, being frustrated, and so on (cf. Tooby & Cosmides, 2000; Ekman, 1999).

- **TB2:** ***Basic emotions are associated with programs.*** Ekman, following Tomkins, assumed that basic emotions are managed by *affect programs*

that control when the emotions are elicited and how they unfold. Other researchers speak of basic emotion *systems* (Panskepp & Watt, 2011) or basic emotion *mechanisms* (Levenson, 2011). Despite these terminological differences, these accounts all agree that basic emotion programs–systems–mechanisms are causally responsible for the coordination of organismic resources toward the goal of efficiently dealing with an evolutionarily relevant emotion-specific life task.

• **TB3:** ***Basic emotions are associated with emotion-specific hardwired neural circuits.*** Basic emotion theorists generally assume that basic emotions have a one-to-one correspondence with hardwired neural circuits. Levenson (2011, p. 382) points out that "the central organizing mechanism would have to be initially hard-wired in the nervous system in at least a primitive form." A hardwired circuit is one that is "built in the nervous system" (p. 382), inherited, present at birth, and homologous to the brain circuits present in evolutionarily related species. Ekman and Cordaro (2011, p. 366) mention "the central brain mechanisms that are organizing and directing our emotional responses." Izard (2011, p. 375) argues that "each of the small set of emotions . . . I have called 'basic' . . . have dedicated . . . neural systems." Panskepp (1998) has explored basic emotion networks in mammalian brains in great detail, arguing that they are evolutionarily old, subcortical, and capable of being stimulated chemically and electrically in ways that directly result in emotional responses.

• **TB4:** ***Basic emotion programs are elicited by automatic appraisals, and generate automatic and mandatory responses.*** On the input side, it is assumed that basic emotion programs are activated by automatic appraisals.[4] The automatic appraisal mechanism scans the environment for "prototypical situations that have profound implications for the organism's immediate well-being and long-term survival" (Levenson, 2011, p. 381), activating a basic emotion program when such situations are found. Each basic emotion will be activated by a distinctive appraisal, which aims to detect cues reliably associated with the life task the emotion evolved to solve.[5]

On the output side, it is assumed that the activation of a basic emotion program leads to a suite of automatic and mandatory responses. As Ekman and Cordaro (2011, p. 366) put it, as soon as the basic emotion program is activated a

> cascade of changes (without our choice or immediate awareness) occurs in split seconds in: the emotional signals in the face and voice; preset actions; learned actions; the autonomic nervous system activity that regulates our body; the regulatory patterns that continuously modify our behavior; the retrieval of relevant memories and expectations; and how we interpret what is happening within us and in the world.[6]

• **TB5:** ***Basic emotions are associated with emotion-specific responses.*** The mandatory nature of the responses associated with basic emotion programs means that each basic emotion is associated with a distinctive *pattern of responses* (or components). Basic emotion theorists further assume that each component of the pattern is *emotion-specific*. In other words, it is generally presupposed that each response associated with a basic emotion corresponds *one-to-one* with that emotion.

Tomkins (1995, p. 58) initially argued that affect programs control "facial muscle responses, autonomic blood flow, respiratory, and vocal responses," suggesting that "these correlated sets of [distinctive] responses will define the number and specific types of primary affects." Contemporary basic emotion theorists have expanded the set of emotion-specific responses beyond expressions and autonomic bodily changes to include, among others, instrumental behaviors, subjective experiences, thoughts, memories, and images.

Even though there is a mandatory core to the responses associated with basic emotions, basic emotion theorists acknowledge that such responses can be partially regulated. The prime example of regulation is constituted by culturally specific *display rules* that affect whether the automated signals in the face and body are quickly inhibited or allowed to unfold without interference (Ekman & Friesen, 1969).

• **TB6:** ***Basic emotions are pancultural, present across species, and emerge early in development.*** Since basic emotions are evolutionary adaptations, basic emotion theorists expect to find them in different human cultures and in homologous form in related nonhuman species. Additionally, it is expected that basic emotions will emerge early in development, prior to the development of sophisticated cognitive capacities. Some have argued that most basic emotions emerge by the ninth month of life in human infants (Campos, Barrett, Lamb, Goldsmith, & Stenberg, 1983; see also Lewis, 2007).

The bulk of the evidence for evolutionary continuity across cultures and species pertains to facial expressions (Matsumoto, Keltner, Shiota, O'Sullivan, & Frank, 2008). Darwin suggested that emotional expressions were pancultural, but added that they did not evolve for their communicative function. Ekman (1999) proposed instead that the communicative function of emotional expressions is crucial to their evolutionary origin. As he put it, "I believe it was central to the evolution of emotions that they inform conspecifics, without choice or consideration, about what is occurring" (p. 47). This has turned the assumption of emotion-specific facial expressions into a non-negotiable tenet for traditional basic emotion theorists.

• **TB7:** ***Basic emotions are designated by folk psychological emotion categories such as anger, fear, happiness, disgust, and so forth.*** Although

contemporary basic emotion theorists disagree to some extent on which specific emotions are basic, the lists they propose show significant overlap. On pretty much everyone's list we find *anger* and *fear*. *Happiness, sadness*, and *disgust* are also widely invoked as examples of basic emotions (Ekman & Cordaro, 2011; Levenson, 2011). More idiosyncratic examples include *surprise* (Ekman & Cordaro, 2011; Levenson, 2011) and *interest* (Izard, 2011). Several more instances are judged by various authors to be candidate basic emotions for which we are likely to find empirical evidence in the future (e.g., *contempt, guilt, shame, amusement, pride, embarrassment*, and *relief*).

The Problem of Variability

The seven commitments introduced in the previous section lead to a number of empirical predictions. The viability of traditional BET hinges on whether such predictions are supported by evidence. Here, I focus on emotions that are basic on most theorists' lists, and on the two predictions that have elicited the lion's share of empirical work:

Prediction 1: There should be hardwired neural networks with a one-to-one correspondence to anger, fear, happiness, sadness, disgust, and so forth.

Prediction 2: There should be coordinated packages of responses with a one-to-one correspondence to anger, fear, happiness, sadness, disgust, and so forth.

Even though basic emotion theorists are convinced that the evidence supports both predictions, psychological constructionists such as Russell (2003), Barrett (2006a, 2006b) and Lindquist, Wager, Kober, Bliss-Moreau, and Barrett (2012) have made a strong case to the contrary. They have published several meta-analyses of the empirical literature and have concluded that they support the following two theses, which are incompatible with traditional BET:

No one-to-one correspondence (NOC) *thesis*: There is no one-to-one correspondence between anger, fear, happiness, sadness, and so forth, and any neurobiological, physiological, expressive, behavioral, or phenomenological responses.

Low coordination (LC) *thesis*: There is low coordination between neurobiological, physiological, expressive, behavioral, or phenomenological responses among instances of anger, fear, happiness, sadness, and so forth.

Basic emotion theorists have responded to NOC and LC in a variety of ways. They have either reasserted their original position despite the contrary evidence (e.g., Ekman & Cordaro, 2011) or added minor qualifications to their original accounts to accommodate the contrary evidence (e.g., Izard, 2007, 2011), or invoked specific empirical studies that are compatible with BET (e.g., Panskepp & Watt, 2011) or published alternative meta-analyses that seem more favorable to BET (e.g., Kreibig, 2010; Stephens, Christie, & Friedmana, 2010; Lench, Flores, & Bench, 2011; but see Lindquist, Siegel, Quigley, & Bararett, 2013). I have argued elsewhere that none of these strategies is likely to succeed (Scarantino, 2009, 2012a, 2012b, 2012c).

The core problem is that Predictions 1 and 2 are *a priori* unreasonable. There is no good reason to expect hardwired neural networks with a one-to-one correspondence to anger, fear, happiness, sadness, disgust, and so forth, or packages of highly coordinated responses with a one-to-one correspondence to anger, fear, happiness, sadness, disgust, and so forth. On the contrary, variability should be expected with respect to the folk categories of anger, fear, happiness, and so on, for at least three reasons: (1) because of how such categories are formed in natural languages, (2) because of the flexibility required for basic emotions to deal successfully with life tasks, and (3) because basic emotions interact with other mental states.

Acknowledging these three sources of variability will lead to what I call the *new BET*, an updated version of BET that acknowledges the constructionist critique while preserving the notion that basic emotions are specialized and evolved programs for dealing with fundamental life tasks. This modification of BET is required in order to account for the empirical data, and I argue that it is in keeping with how scientific theories should be modified to accommodate anomalies.

From Variability to Psychological Constructionism

Compositional Constructionism versus Psychological Constructionism

Psychological constructionists have interpreted the evidence for variability as supporting the following thesis:

> *No programs thesis* (NPT): There are no hardwired programs associated with anger, fear, happiness, sadness, disgust, and so forth, that are causally responsible for coordinating patterns of emotion-specific responses.

Since NPT is incompatible with traditional BET, psychological constructionists have suggested that we should give up on BET all together,

and take the constructionist alternative seriously. The orienting thought of psychological constructionism is the assumption that the scientific understanding of psychological phenomena requires breaking them up into their most primitive components. These are what Russell (2003, p. 146) has called "elemental—but still psychological—building blocks."

In an article on the history of psychological constructionism, Gendron and Barrett (2009) have traced its origins back to Spencer (1855), and suggested that prior to the 20th century the approach "was most clearly articulated by William James and Wilhelm Wundt" (p. 319). Here I focus on James, singled out by Mandler (1990, p. 180) as the proponent of "the first constructionist psychology, attempting to understand the processes that generate and construct behavior and conscious experiences." Focusing on James's version of constructionism allows me to distinguish between two notions of "construction" that have often been conflated.

James famously argued that emotions were "constructed" by means of two building blocks: perception and bodily changes. Furthermore, he suggested that the perception of these bodily changes follows *directly* the perception of some exciting fact. This contradicts common sense, according to which "the mental perception of some fact excites the mental affection called the emotion, and that this latter state of mind gives rise to the bodily expression" (James 1890, p. 449).

In James's account, there is no *psychic entity* that mediates between the mental perception of the exciting fact and the bodily expression, in the sense that "*the bodily changes follow directly the* PERCEPTION *of the exciting fact*" (p. 449) and "*our feeling of the same changes as they occur* IS *the emotion*" (p. 449, emphasis and capitalizations in original).

James's approach expresses a methodological position I call *compositional constructionism*:

> A theory is *compositionally constructionist* with respect to emotion episodes if an only if such episodes are analyzed into building blocks.

James's theory is compositionally constructionist because it takes emotion episodes to be feeling episodes analyzed in terms of perception and bodily changes. Basic emotion theory is also compositionally constructionist, because basic emotions are analyzed in terms of physiological, expressive, behavioral, and phenomenological building blocks coordinated by basic emotion programs.

To qualify properly as *psychologically constructionist*, I submit, a model of emotions must do more than break emotions apart into building blocks: It must also hold that the building blocks are not specific to emotions, and that there are no mechanisms specific to emotions that bring

such building blocks together into an emotion episode. When both conditions apply, the occurrence of an emotion episode can be explained in terms of non-emotional processes.

Basic emotion theory is clearly not psychologically constructionist in the sense just described, because an underlying affect program is assumed to coordinate the physiological, expressive, behavioral, and phenomenological building blocks into which emotions are analyzed.

My view is that James's theory is also not psychologically constructionist. Although he did not consider perception and bodily changes to be specific to emotions, he did posit the existence of emotion-specific mechanisms that couple the perception of an exciting fact with the occurrence of the bodily changes whose perception constitutes the emotion. This interpretation is admittedly contentious (see Gendron & Barrett, 2009, for an alternative interpretation), but I think it is backed up by ample textual evidence. Consider the following excerpt:

> The love of man for woman, or of the human mother for her babe, our wrath at snakes and our fear of precipices, may all be described similarly, as instances of the way in which peculiarly conformed pieces of the world's furniture will fatally call forth most particular mental and bodily reactions, in advance of, and often in direct opposition to, the verdict of our deliberate reason concerning them. (James 1890, p.191)

In this passage, James appears to be arguing that emotion mechanisms independent of "our deliberate reason" are automatically activated by "peculiarly conformed pieces of the world's furniture" and mandatorily—"fatally"—cause the bodily reactions whose perception gives rise to the feelings of love or wrath or fear. This interpretation is further bolstered by the fact that James draws an analogy between the way pieces of the world's furniture "call forth" bodily reactions and the way *keys* open *locks*: "Every living creature is . . . a sort of lock, whose wards and springs presuppose special forms of key" (James 1890, p. 192).

For example, the lock associated with delight makes it so that "no woman can see a handsome little naked baby [a key for delight] without delight," and the lock associated with fear makes it so that "in advance of all experience of elephants no child can but be frightened if he suddenly finds one trumpeting and charging upon him [a key for fear]" (James 1890, p. 192). In this picture, which I find much closer to basic emotion theory than to psychological constructionism, there exist causally powerful mechanisms specific to emotions, whose job is to "call forth" the specific bodily reactions whose perception amounts to love, wrath, and fear.

If so, we should not mistake James's opposition to the idea that there is a "psychic entity" mediating between pieces of the world furniture and

bodily reactions for an opposition to the idea that there is a specialized "causal entity" mediating between pieces of the world furniture and bodily reactions.[7] It is precisely the existence of this specialized causal entity that psychological constructionists deny, suggesting instead that episodes of love of man for woman, or wrath at snakes or fear of precipices, can be explained without invoking specialized emotion mechanisms. As I argue below, the viability of psychological constructionism hinges on how convincing these alternative explanations turn out to be for making sense of emotion episodes.

Two Varieties of Psychological Constructionism

Several influential psychological constructionists have emphasized the importance of one primitive in particular, namely *core affect*, which is defined as a blend of hedonic and arousal values and hailed as "the most basic building block of emotional life" (Barrett 2006a, p. 48). Core affect is ubiquitous, because one is always in a state characterized by some degree of pleasure ranging from extreme unpleasantness (e.g., agony) to extreme pleasantness (e.g., ecstasy), and by some degree of arousal ranging from extreme deactivation (e.g., sleep) to extreme activation (e.g., frantic excitement).

I want to introduce a critical but often neglected difference between varieties of psychological constructionism that hinges on how an emotion episode is supposed to be constructed out of core affect. The difference concerns the role played by *conceptualization*:

> *Nonconceptualist Psychological Constructionism* (NCPC): Episodes of anger, fear, happiness, sadness, disgust, and so forth, occur independently of conceptualization, but the meta-experiences of anger, fear, happiness, disgust, and so forth, require conceptualization of an underlying state of core affect.
>
> *Conceptualist Psychological Constructionism* (CPC): Episodes of anger, fear, happiness, sadness, disgust, and so forth, require conceptualization of an underlying state of core affect in order to come about.

NCPC is the position held by Russell (2003, 2012), and CPC is the position held by Barrett (2006a, 2006b, 2012). Whereas Russell thinks that conceptualization affects the perception of oneself as having a certain emotion (the meta-experience of the emotion) but not whether the emotion occurs, Barrett considers conceptualization necessary to generate an emotion episode in the first place.[8] Let us consider the two models in turn.

Russell's Nonconceptualist Psychological Constructionism

Psychological construction is for Russell (2012, p. 82) "an umbrella term for three sets of processes, those that produce: (a) the [emotion] components, (b) associations among these components, and (c) the categorization of the pattern of components as a specific emotion." The components of discrete emotions include the familiar ones invoked by basic emotion theorists (appraisals, expressions, autonomic changes, instrumental actions, subjective experiences) plus some new ones: the perception of affective quality of an antecedent event (i.e., whether the antecedent event is pleasant–unpleasant and arousing–not arousing), the change in core affect resulting from this perception, the attribution of this change to some antecedent event (e.g., the event of encountering a charging elephant), meta-experience (the experience of categorizing oneself as afraid or angry or happy, etc.), and regulation (the deliberate attempt to self-regulate that follows the categorization of oneself as having a certain emotion).

Russell points out that the emotion components associated with fear, anger, happiness, and so on, do not correlate to the extent that basic emotion theorists have predicted, but he acknowledges that they correlate to some extent. In fact, whether a discrete emotion episode is instantiated hinges on the extent to which such components correlate. If they correlate sufficiently to match the mental script (or prototype) for some folk emotion category *E*, an episode of *E* is instantiated (whether or not anyone categorizes the episode as *E*) (Fehr & Russell, 1984).

Russell's model explains the variability within folk emotion categories in terms of the fact that several different combinations of components will match the script associated with each folk emotion category. For example, there will be cases of "fear" that include facial signals and cases of fear that do not, cases of fear that include autonomic bodily changes and cases that do not, and so on. Furthermore, when a given component is instantiated, variability in the way it is instantiated will be the norm. Among instances of fear associated with physical actions, some will involve running, whereas others will involve hiding, shooting, climbing trees, making phone calls, and an open range of other possible actions.

In the rare cases in which all or most components are instantiated, the instance of a folk emotion category *E* will become *prototypical*. Most members of folk emotion categories will be nonprototypical members, instantiated by virtue of a fairly weak correlation among components. Finally, when the components are neither sufficiently many for clear membership nor sufficiently few for clear nonmembership, instances will become borderline cases of emotion.

The central challenge for NCPC is to explain what underlies the correlations among components of fear, anger, and so on, in both prototypical and nonprototypical cases. Three possibilities must be considered. The

first is that there is a unique emotion mechanism or program associated with each folk emotion category *E* that is causally responsible for bringing about the components that instantiate *E*. The second is that multiple emotion mechanisms associated with each folk emotion category *E* are causally responsible for bringing about the components that instantiate *E*. In such case, different instances of *E* will be caused by different emotion mechanisms. The third possibility is that there are no emotion mechanisms at all, and the components that instantiate folk emotion categories are brought together by non-emotional means.

Traditional BET favors the *unique emotion mechanism* assumption, but I have argued that this assumption is at odds with the empirical data. Russell favors the *no emotion mechanisms* assumption, and proposes three causes other than emotion mechanisms why the emotion components correlate: "(a) features in the environment have a correlational structure, which then creates correlations among components, (b) one component can influence another, and (c) two components are correlated when they are both influenced by the same central mechanism such as attention" (Russell, 2012, p. 83).

I have considered these alternative explanations elsewhere and argued that they do not successfully explain why the components correlate to the (limited) extent that they do (Scarantino, 2012a). My proposal is that a better explanation for the existing correlations among components is the presence of *multiple causal emotion mechanisms* associated with the same folk emotion category.

In his reply, Russell (2012b) has pointed out that my proposal is largely speculative, in that I have not provided details on what these multiple causal mechanisms are, and on how they are supposed to work. This is a fair criticism. In the next section, I start providing some of the missing details, illustrating how the new BET can put theoretical flesh around the multiple emotion mechanisms assumption.

The third leg of Russell's constructionism is the idea that the categorization of the pattern of components that instantiates a specific emotion generates a *meta-experience* of emotion. This is the experience associated with categorizing oneself under a certain folk emotion category (e.g., the experience associated with categorizing oneself as "angry" or "afraid"). Russell (2012, p. 105) is clear that "emotional episodes can occur unaccompanied by an Emotional Meta-Experience." This is the case for infants and animals, on the assumption that they do not possess folk emotion concepts. It is also the case for adult human beings who are deeply engrossed in an emotion episode and lack the attentional resources required to categorize themselves under an emotion concept.

Finally, based on Russell's theory, emotion categorizations can be mistaken. For example, an episode of fear may be instantiated by virtue of

the fact that enough fear components have co-occurred to match the fear script, but the fearful person may wrongly categorize him- or herself as "angry." Russell (2012, p. 105) gives the example of alexithymics, namely diseased patients whose defining feature is their inability to categorize correctly the emotions they are undergoing.

Barrett's Conceptualist Psychological Constructionism

Russell's proposal differs from Barrett's (2006a, 2206b) CPC, according to which concept use is constitutive of emotion episodes. Whereas Russell endorsed a prototype theory of concepts, Barrett endorses Barsalou's (1999) theory of concepts, according to which concepts are *goal-related* (we conceptually represent things in order to do things with them) and *situated* (things are not represented in isolation but in a setting that will make inferences about what do to with them more effective).

A defining feature of Barsalou's theory is that concepts are not amodal collections of features either classically or prototypically organized but rather multimodal integrations of modality-specific memory traces—perceptual symbols—stored in long-term memory and "organized into a simulator that allows the cognitive system to construct specific simulations of an entity or event" (Barsalou 1999, p. 586). The fact that perceptual symbols are organized into a simulator is one of the elements distinguishing Barsalou's theory from a straightforward exemplar theory of concepts, in which memory traces of encounters are not integrated into a unified representation.

Perceptual symbols in a simulator span every experiential modality in which previous encounters with instances of the category have occurred, including sensory experience, motor experience, and emotional experience. For instance, the simulator for CAR will include memories of how cars looks and sound, memories of the actions involved in interacting with cars, and memories of the emotions elicited by cars.

Simulators produce simulations in working memory, namely activations of a subset of the information stored in the simulator in the form of a partial reenactment that may be conscious or unconscious. Every simulation counts as a specific *conceptualization* of a given *concept*, which according to Barsalou's (1999) theory is the integrated collection of multimodal memory instances of a certain category organized into a simulator.

When objects and events are perceived, they are categorized as members of a certain category just in case the simulator associated with such category produces a simulation that "fits" the perceived object or event. The same simulator can produce many distinct simulations depending on context, which accounts for the variability that characterizes the instances of most lexical concepts, whose members share nothing more than family resemblances.

Once a categorization has taken place, the simulator for the categorized object or event is updated with a new memory of an encounter with a category instance. Finally, producing the categorization activates inferences and possibly bodily states that help interact with the category instance in the circumstances at hand. Which inferences and bodily states are activated will depend on the specific simulation triggered and on the situational demands of the context.

Barrett (2006a, 2006b) puts Barsalou's theory of concepts at the core of her own *conceptual act theory* (CAT) of emotions, the most careful and detailed proposal to emerge so far from the CPC camp. CAT's central thesis is that emotions are *situated conceptualizations* (cf. Wilson-Mendenhall, Barrett, Simmons, & Barsalou, 2011). This is to say that "categorizing the ebb and flow of core affect into a discrete experience of emotion corresponds to the colloquial idea of 'having an emotion" (Barrett 2006a, p. 49). What is being claimed here is not that conceptualizing oneself under an emotion concept *E* produces an emotional meta-experience of *E* (NCPC), but, more strongly, that it produces emotion *E* (CPC). To put it in slogan form, no conceptualization, no emotion.

Consider an episode of fear. According to Barrett and her colleagues, "*fear* cannot be understood independently of an agent conceptualizing his [sic]- or herself in a particular situation" (Wilson-Mendenhall et al., 2011, p. 1108). How so? In a nutshell, a fear conceptualization is a simulation produced by the FEAR simulator. As we have seen, this is an integrated, multimodal collection of perceptual, motoric, and affective memories of fear experiences. Barrett's proposal is that producing a fear conceptualization of an underlying state of core affect *is* having fear.

The causal role that basic emotion theorists give to *affect programs* and that Russell gives to a heterogeneous variety of *non-emotional factors* (e.g., the correlation of features in the environment, the causal connections among components, and the presence of non-emotional mechanisms such as attention) is in Barrett's theory given to *situated conceptualizations* of folk emotion concepts:

> Although a person is always in some state of core affect . . . a situated conceptualization has the capacity to shift core affect toward a state typically experienced during emotion episodes for a particular kind of situation. Along with core affect, the situated conceptualization produces related changes in bodily states, such as muscle tension and visceral activity. Additionally, the situated conceptualization may initiate relevant actions that are typically associated with the emotion in this situation, with core affect and bodily states often motivating and energizing these actions. Finally, the situated conceptualization may produce perceptual construals of the current situation, biasing and distorting perception toward typical experiences associated with the respective type of situation. (Wilson-Mendenhall et al., 2011, p. 1109)

This passage makes it clear that the physiological, expressive, behavioral, and phenomenological components commonly associated with fear (muscle tension, visceral activity, avoidance actions, perceptual changes, etc.) are assumed to be "produced" by a *conceptual act*, namely a situated FEAR simulation. The variability of instances of folk emotion categories is explained by the variability of context-dependent simulations associated with the same folk emotion category.

Consider the difference between the fear one may undergo when lost in the woods at night and the fear one may undergo when realizing that one's work presentation is not ready (Wilson-Mendenhall et al., 2011). These two instances of fear will presumably differ in terms of the components associated with them at the level of physiological changes, expressions, behaviors, and phenomenology, as well as at the neural level. Barrett thinks that they do because two different FEAR simulations have produced them.

CAT is remarkably original and thought provoking, but it is also potentially problematic on a number of fronts. Here, I briefly introduce three conceptual challenges for CAT, in the spirit of fostering further discussion:

1. *How do we transition from a world without the FEAR concept to a world with the FEAR concept?* An obvious requirement for producing a fear conceptualization is having a FEAR concept, which according to Barrett's theory is an integrated, multimodal set of memories of fear experiences. The problem is that CAT holds that every fear experience presupposes an act of conceptualization, which is to say that it presupposes having a FEAR concept in the first place. This makes a mystery of how we ever transition from a world without a FEAR concept to a world with a FEAR concept. The formation of the FEAR concept according to Barrett's theory requires that someone experiences fear (no one can have memories of fear without someone having had fear experiences), but experiences of fear simply cannot happen according to CAT in a world where no one has the FEAR concept. So how is the FEAR concept supposed to emerge?

2. *What exactly is being categorized as fear if fear does not exist prior to the categorization?* In standard cases of categorization—say, the categorization of a car under the CAR concept—the concept user compares a perceived instance *X* with a situated CAR simulation, and categorizes *X* as a car if the simulation "fits" *X*. Importantly, whether or not *X* is a car does not hinge on whether it is categorized as such: cars are not situated CAR conceptualizations.

Barrett's claim is that things are different when it comes to emotions. An instance of fear is supposed to occur *by virtue of* a fear simulation. This creates a puzzle, namely that it is hard to see how there could be a fit between a perceived instance *X* and a FEAR simulation if *X* is not fear until it has been categorized as such. Analogously, if something became a

car by virtue of being categorized as such, there could not be a fit between a perceived instance *X* and a CAR simulation, because *X* would not be a CAR until so categorized.

Another way to put the problem is that if what is causally responsible for the activation of the components associated with fear is a FEAR simulation, as Barrett argues, there are no components to be "fitted" by the FEAR simulation, because such components are not present until a satisfactory fit has been provided. But what could possibly ground such fit then?

Note that a fear categorization is not an evaluation that fear "fits" the circumstances at hand (e.g., the circumstances of being in the forest all alone at night or being unprepared for a work presentation). This sort of evaluation can certainly occur prior to the fear components being in place, and it is what basic emotion theorists call an *appraisal* of danger. Rather, a fear categorization aims to determine whether a FEAR simulation "fits" a perceived event sequence already under way in the circumstances at hand. How is this latter evaluation going to take place if the fear components have yet to be produced?

3. *How can a fear categorization be necessary and sufficient for having fear?* CAT holds that categorizing an underlying state of core affect as fear is *necessary* for having fear. It follows that no humans who lack the FEAR concept, no infants, and no animals can have fear.[9] The problem with this position is that creatures without the relevant concepts appear perfectly capable of manifesting the combinations of components we associate with fear.

Consider a patient with alexithymia who systematically misapplies the FEAR concept, an infant, and a dog. Suppose that they are suddenly thrown into a cage with an elephant that starts trumpeting and charging. My prediction is that, just like James suggested, all three creatures would automatically and mandatorily manifest the prototypical components of fear at the level of physiology, expressions, behavior and phenomenology, with the exception of the *meta-experience* of fear. Since such creatures by assumption lack the ability to correctly apply the FEAR concept, they cannot have the experience associated with categorizing themselves under such a concept.

CAT goes well beyond this claim, and commits us to saying that even though we perceive them as being afraid, these creatures are *not* truly afraid. As Barrett (2012, p. 420) puts it with respect to anger, "if some people do not have a concept of anger, then [a] constellation [of components such as a scowl, blood pressure increase, and a feeling of offense] will never exist as anger for those people (i.e., it is not that they are truly angry and don't know it)." This position is unpersuasive.

First, it is unclear why such creatures would not be truly angry or afraid if they fitted, respectively, the ANGER and the FEAR simulations of

creatures endowed with the relevant concepts. When we travel to another country where no one has the ROSE concept and find in a garden something that fits our ROSE simulation, we correctly conclude that we have encountered a rose abroad, even though the locals do not know it. Why would things work any differently for emotion concepts?

Second, it is unclear why CAT assumes that situated categorizations are necessary for bringing about the components of anger or fear, if creatures that lack the ability to engage in such categorizations can still manifest such components. This calls into question the causal role allegedly played by conceptual acts in producing the constellations of components associated with discrete emotions.[10]

CAT also holds that categorizing an underlying state of core affect as fear is *sufficient* for having fear. This proposal is also hard to swallow, because it would seem to prevent the possibility of categorization errors. According to CAT, emotion concepts are such that emoters cannot apply them wrongly to their own emotions. For example, if the alexithymic patient in the elephant example categorized him- or herself under the ANGER concept while manifesting all the components of fear, we would presumably not conclude that he or she is angry rather than afraid, contrary to CAT. Any good account of emotion episodes should allow for the possibility of introspective error concerning which emotions one is having, and it is unclear how CAT can account for that.

Although I have raised some challenges for CAT, I want to emphasize that I do not deny the important role that categorizations play in affecting the unfolding of an emotion episode. Among other things, the ability to self-categorize as being angry or afraid will have an impact on the experience associated with the emotion episode, on whether and how a memory of the episode is formed, and on whether and how the emotion episode is regulated over time.[11]

From Variability to a New BET

As Barrett (2009, p. 1290) puts it, "[d]uring the late nineteenth century . . . and mid-twentieth century . . . many psychological constructionist models of emotion were proposed, all of them inspired by the observation of variability in emotional responding and the failure of basic emotion approaches to account for this variability."

This quotation usefully emphasizes that a primary motivation for constructionist proposals is the conviction that BET does not have the resources to account for variability. We must also note that worries about the variability of the bodily changes associated with emotions were the primary motivators for two of the revolutions in affective science of the

20thcentury, namely the *behaviorist revolution* and the *cognitivist revolution*. Behaviorists like Skinner (1953) argued that "[i]n spite of extensive research it has not been possible to show that each emotion is distinguished by a particular pattern of responses of glands and smooth muscles", a view Cannon (1929) had influentially attacked. Skinner had the same worry about facial expressions of emotions, as he said that "it has not been possible to specify given sets of expressive responses as characteristic of particular emotion" (1953, p. 161).

The cognitivist model of emotions proposed by Schacter and Singer (1962) was also driven by the view that "the variety of emotion, mood, and feeling states are by no means matched by an equal variety of visceral patterns" (p. 379). The absence of physiological differentiation raised the question of what distinguished from one another emotions associated with undistinguishable physiological changes. This led Schachter and Singer (1962) "to suggest that cognitive factors may be major determinants of emotional states" (p. 379).

These quotes reveal how accounting for variability has historically been a central challenge for models of emotions, and one that basic emotion theory will need to successfully address in order to stay competitive. According to Barrett (2009), basic emotion models assume that "observed variability in emotional responding is the result of epiphenomenal social factors, like display rules or other regulation processes that mask or inhibit pre-potent, stereotyped responses" (p. 1288). Additionally, basic emotion theorists often "explain the variability away as error or failure of experimental design" (p. 1288).

Regulation processes and experimental error do have a role to play, but I agree with Barrett that they fall short of explaining the massive amount of variability we find associated with basic emotions. Unlike Barrett, however, I am convinced that BET has the resources to account for variability. In what follows, I distinguish between three sources of variability—concept-dependent, context-dependent, and interaction-dependent variability—and explain how taking them into account can lead to a promising new version of BET.

Concept-Dependent Variability

As we have seen, the standard lists of basic emotions provided by basic emotion theorists comprise folk psychological categories such as *anger, fear, happiness, sadness, disgust*, and so on. This terminological choice reveals the conviction that all the items we call *anger, fear, happiness*, and so forth, in English are basic emotions. The assumption then is that all members of these folk emotion categories evolved to solve fundamental life tasks, are implemented by an emotion-specific neural program, involve a highly coordinated set of emotion-specific responses, and so on.[12]

Since the empirical evidence has been unfavorable to this hypothesis, psychological constructionists have concluded that the folk emotion categories used by basic emotion theorists to designate basic emotions fail to designate *natural kinds*. A natural kind is (roughly) a theoretically homogenous class of items about which a great many explanatory and predictive generalizations can be formulated in a certain scientific discipline (cf. Boyd, 1999; Scarantino, 2012c).

I have emphasized (Scarantino, 2012c) that the assumption that folk emotion categories designate natural kinds is typical not only of BET but also of the great majority of theories of discrete emotions. The problem is that no theory has so far been able to unveil the scientifically interesting explanatory and predictive generalizations that are true of all members of any folk emotion category. For every candidate generalization at the level of bodily changes, neural circuits, origins, current function, and development, we seem to be able to find members of any given folk psychological category *E* to which the generalization applies and members to which it does not apply.

One interpretation of this failure is that emotion theorists have not been sufficiently ingenious so far. The other interpretation, which both psychological constructionists and I favor, is that folk emotion categories are highly heterogeneous, to the point that *no* scientifically interesting generalizations are likely to apply to all of their instances. If this is so, a fundamental rethinking of the categories on which affective science relies is in order.

The methodological approach I recommend differs from the one championed by psychological constructionists. Whereas psychological constructionists have rejected the view that there are specialized emotion mechanisms causally responsible for the coordination of physiological, expressive, behavioral, and phenomenological responses, I think the search for such mechanisms is exactly the way to go, provided that we stop assuming that there is a *unique* specialized emotion mechanism associated with each folk emotion category.[13] Rather, a *multiplicity* of such mechanisms should be expected to correspond with anger, fear, disgust, and so on.

Whereas traditional BET assumes that all items included in the folk categories of *anger, fear, happiness*, and so forth, share an emotion-specific neural program or an emotion-specific package of coordinated responses, the new BET rejects this assumption, proposing instead that only a subset of the members of the folk categories of *anger, fear, happiness*, and so forth, are basic.[14] The transition from traditional BET to the new BET is modeled after similar transitions that have occurred in other scientific domains over time. For example, whereas the initial assumption in memory science was that the folk psychological category of *memory* designates a unique information-retention mechanism, it is now commonly acknowledged "that memory can be divided into multiple forms or systems–collections of processes that operate on different kinds of information and according to different rules" (Schacter, 2004, p. 644). The received view currently is that

multiple memory systems exist, are activated by distinct tasks, and differ on a number of important theoretical dimensions (e.g., duration, storage modality, capacity, neural underpinnings).[15]

This has led to the proliferation of a fine-grained non-folk psychological taxonomy, which distinguishes, for instance, between *long-term memory* and *short-term memory* (a.k.a. *working memory*), and between varieties of each (e.g., *declarative long-term memory* and p*rocedural long-term memory*). Note that the idea that memory is not a theoretically homogeneous category "was hardly acknowledged until a window was opened into the normal operations of the mind through the study of cognitive losses suffered by brain-damaged patients" (Rosenbaum et al., 2005, p. 990). This evidence conclusively showed that brain damage can impair some forms of memory but not others, eventually leading to the now "widely accepted idea of multiple memory systems" (p. 990).

My view is that the time has come for emotion science to undergo a similar transition. The empirical data on neural circuitry, phenomenology, physiology, expressions, and behavior should lead us to take seriously the idea of multiple emotion systems, multiple anger systems, multiple fear systems, multiple disgust systems, and so on (Scarantino, 2012c). Just as the fact that the folk psychological category of "memory" is not a natural kind is compatible with the existence of natural kinds of memory (e.g., procedural long-term memory), the fact that the folk psychological categories of *anger, fear, disgust, happiness* and so forth do not designate natural kinds is compatible with the existence of natural kinds of anger, natural kinds of fear, natural kinds of disgust, natural kinds of happiness, and so forth. Emotion scientists will have to find out how many of these there are, what defines them, and what scientifically interesting explanatory and predictive generalizations are true of them.

The new BET I outline in what follows offers an account of *some* of the natural kinds into which folk emotion categories should split. Contrary to traditional BET, the new BET does not aim to capture what *all* emotions or *all* angers or *all* fears are like, but only what some relevant subsets of such folk categories are like. I propose that we initially designate such subcategories as *basic anger, basic fear, basic disgust, basic happiness*, and so on, in order to emphasize that the predictive and explanatory generalizations formulated by the new BET are not meant to apply to all members of the folk emotion categories (as I will argue below, there are reasons to further refine the basic emotions taxonomy). This once again replicates the model of memory studies, in which scientifically interesting generalizations are taken to apply only to instances of theoretically motivated subcategories such as *short-term memory* or *long-term memory*, rather than to the folk psychological category of *memory* writ large.

I conclude by discussing a different type of variability that has no impact on the natural kind status of folk emotion categories. It is a

well-known fact that various cultures differ in terms of the folk emotion categories their languages contain. For example, the superordinate category of *emotion* is absent from several languages. Some languages also use subordinate categories that English lacks and lack subordinate categories we commonly use. Some lexical categories, finally, are hard to translate between cultures, because they capture a combination of components not labeled lexically in another culture (cf. Russell, 1991; Mesquita & Frjida, 1992; Mesquita, 2003).

What are we to make of this *linguistic* variability? Russell (2012a) and others have considered it a reason to conclude that folk emotion categories in English are scientifically unsuitable. While I agree with the conclusion, I disagree with the rationale. What matters for the scientific suitability of an emotion category is whether it designates a theoretically homogeneous category for the explanatory and predictive purposes of affective science (i.e., a natural kind). The fact that other cultures lack a category equivalent to, say, "fear" does not affect whether "fear" designates a natural kind. This is because it does not affect the extension of the category in English, and it is this extension alone that determines whether the explanatory and predictive generalizations that substantiate natural kind status can be formulated about instances of fear.

Russell (2012b, p. 286) has replied that "it would be pure coincidence if English got it right, and all languages that categorize emotions differently got it wrong" when it comes to parsing the affective domain into suitable scientific categories. It would indeed be a pure coincidence, and we should definitely not expect that all folk emotion categories in English are scientifically suitable, whereas all non-English folk emotion categories are scientifically unsuitable. This reply, however, does not address my point that whether a folk emotion category *E*—in English or in any another language—captures a scientifically suitable category is independent of what folk emotion categories exist in other languages, because so is the extension of *E*.

What we should expect is rather that every natural language "got it wrong" when it comes to the scientific suitability of its folk emotion categories (give or take a few possible exceptions). Since natural languages do not generate folk psychological categories with the intent of capturing homogeneous domains of scientific investigation, the discovery for such domains, in English as in every other language, generally requires substantial linguistic refinement.

Context-Dependent Variability

A second important reason why variability should be expected is not due to the fact that folk emotion categories have theoretically heterogeneous extensions, but to a central commitment of BET, namely that basic emotions

evolved to deal efficiently with fundamental life tasks, such as dealing with danger, fighting, suffering losses, being frustrated, and so forth. What basic emotion theorists have neglected (with a few exceptions) is that, in many cases, what such tasks require is not a rigid cascade of responses but a set of flexible response *tendencies* that are adaptable to the context at hand.

The key difference between traditional BET and the new BET is that the latter, unlike the former, allows basic emotions to have highly flexible manifestations. More precisely, the new BET draws a distinction between two types of activations of a basic emotion program/mechanism: output-rigid activations and output-flexible activations. Whereas output-rigid activations involve *mandatory responses* to stimuli, just as traditional basic emotion theorists have posited, output-flexible activations only involve *irruptive* and *prioritized* response *tendencies* to stimuli (Frjida 1986, 2007).

The first thing to emphasize with respect to output-rigid activations is that different theorists have interpreted the rigidity of the output in different ways. Some output-rigid basic emotions are *unconditioned reflexes*. The unconditioned fear reflex, automatically elicited in a variety of species by, among other things, sudden loss of support and loud sounds, is an example of an output-rigid basic emotion. Some basic emotion theorists have suggested that unconditioned reflexes are the *only* emotions truly deserving of the qualifier "basic." According to Panskepp and Watt (2011, p. 388, emphasis in original), "*basic emotions* can only exist clearly at primary-process levels, namely before learning and higher order thoughts add rich developmental and cultural complexities." Panskepp (2012, p. 32) describes *primary-process levels* as "intrinsic (unconditioned) neuropsychological functions of the brain, responsive initially to only limited sets of environmental events (i.e., unconditional stimuli)."[16]

Other basic emotion theorists have been more inclusive with what they count as basic emotions. Izard (2011), for instance, has proposed that both conditioned and unconditioned reflexes qualify as basic or, as he put it, "first-order emotions." Once higher cognition rather than learning enters the picture, however, a basic emotion is turned into what Izard (2011, p. 372) calls an *emotion schema*: "Emotion schemas always involve interactions among emotion feelings [i.e., basic emotions] and higher order cognition—thoughts, strategies, and goals that complement and guide responding to the emotion experience."

Ekman's view is even more liberal, in that the intervention of both learning and higher order cognition is compatible with the basic status of an emotion. Ekman and Cordaro (2011, p. 367) considers basic emotions to be *open programs* in Mayr's (1974) sense, namely sets of instructions that allow for additional input from experience during the lifetime of the individual. However, according to Ekman's view, basic emotions are also output-rigid. This is because the open programs associated with basic emotions generate what Ekman has characterized as "inescapable" changes

in facial signals, in the autonomic nervous system, in preset and learned actions, and in other emotion components.

The responses are inescapable (or mandatory) in the following sense: "Once set into motion through automatic appraising, the instructions in the affect programs run until they have been executed; that is, they cannot be interrupted" (Ekman & Cordaro, 2011, p. 367). Different components will be uninterruptible for different periods. For facial signals, the period lasts less than a second, after which emoters can deliberately affect their facial expressions. On the other hand, "[t]he changes in our respiration, perspiration, and cardiac activity . . . have a longer time line, some stretching out to 10 or 15 seconds" (p. 367) during which they cannot be interrupted.

The assumption that basic emotions are rigid on the output side has led traditional BET to posit the presence of a pattern of highly coordinated components associated one-to-one with each basic emotion. But there is no good reason to form such an expectation, if we consider the fact that basic emotions evolved to deal with fundamental life tasks that take on different forms and require different adaptive responses. According to the new BET, the rigidity of the output is the exception rather than the rule for basic emotions.

Consider the task of avoiding dangers. Dangers differ in terms of how serious they are, in terms of how distant they are, and in terms of the responses required to avoid them. Some dangers are relatively negligible and quite distant (e.g., a big dog barking in the distance while tied to its owner's leash), and they demand nothing more than orienting and getting ready for unspecified actions if the danger increases. Some dangers are significant and imminent (e.g., being run over by a car), and they demand very specific and reflex-like actions (e.g., immediately jumping away from the car's trajectory).

Some dangers are also significant and fairly imminent (e.g., being attacked by an unleashed dog charging from a faraway distance), but they can be dealt with successfully by a nonspecific range of actions that requires some degree of planning and bodily control (e.g., finding a tree and climbing it, getting a long stick and keeping the dog at bay with it, reaching for a gun and shooting the dog with it).

This is to say that, in order to serve its evolutionary function, a basic emotion that evolved to deal with danger needs occasionally to work as an output-rigid program (the suddenly looming object case), but most of the time as an output-flexible program. Output-flexible basic emotions are best understood as *irruptive* and *prioritized* response *tendencies*.

The response tendencies are *irruptive* in the sense that they are automatically activated by the appraisal system, and they are *prioritized* in the sense that they manifest what Frjida (1986) has labeled *control precedence*. Response tendencies with control precedence endow a specific

task/goal (e.g., avoiding a serious and imminent danger) with precedence over other possible tasks/goals of the organism, and exercise control over all organismic resources, until such task/goal is fulfilled (or its pursuit is inhibited). Frjida (1986, p. 78) suggests that action tendencies with control precedence "clamor for attention and for execution": They "tend to persist in the face of interruptions," they "tend to interrupt other ongoing programs and actions" and they "tend to preempt the information-processing facilities."

Crucially, these prioritized action tendencies are flexibly manifested depending on the context, leading to high variability in the actual responses associated with any output-flexible basic emotion. Furthermore, different instances differ in terms of the degree of control precedence of their associated action tendencies, leading to more or less *intense* instances of the same basic emotion, which affect both the responses associated with the emotion and their degree of coordination.

Generally speaking, the higher the control precedence, the higher the intensity, and the more highly correlated the manifestations of a basic emotion will be. What holds the manifestations of an output-flexible basic emotion together is that they all share the same abstract life task (e.g., avoiding a danger, dealing with a loss, reacting to a frustration) and pursue such a task with (some degree of) control precedence.

Finally, the new BET gives up on the assumption that a unique neural circuit must underlie all instances of an emotion evolved to solve a given fundamental life task *T*. For instance, some evolutionary solutions to task *T* will rely on circuitry N_1 and others will rely on circuitry N_2. This is to say that not only there are instances of what we call "fear" in ordinary English that do not share neural circuitry with "basic fear" (this is what concept-dependent variability alone leads us to expect), but also that "basic fear," understood as an evolutionary solution to the fundamental life task of avoiding dangers, is likely to be too coarse-grained a category for purposes of neural investigation. Neurobiologists should adopt LeDoux's (2012) most recent "one emotion at a time" methodology, according to which neural circuits are studied one at a time within well-controlled behavioral tasks, without assuming that a single neural circuit will correspond to all instances of emotions with same adaptive function.[17]

I emphasize that we should expect differences at the level of neural circuitry not only between output-rigid and output-flexible solutions to life task T, but also between varieties of output-rigid and varieties of output-flexible solutions to life task T. For example, there is evidence that the neural circuit for *unconditioned basic fear* is different from the neural circuit for *conditioned basic fear*, even though both are instances of output-rigid basic fear.[18] The new BET expects that similarly fine-grained distinctions will apply to the neural circuitry associated with other basic emotions.

Interaction-Dependent Variability

A third important reason why variability should be expected is due to the fact that basic emotions interact with other mental states in ways that affect their manifestations. Roseman (2011) has provided some useful examples of this sort of variability.[19] The primary source of interaction-dependent variability is due to the interaction between *basic emotions* and *regulation* (Gross, 1998). Levenson (1999, 2011) has offered a compelling account of such interaction in terms of what he calls a *two-system design*, according to which the manifestations of basic emotions result from the interaction between a *core system* "designed early in evolution to cope effectively with a few very basic, ubiquitous problems" and a *control system* "more recently evolved . . . [and] . . . designed to influence the actions of the core system" (Levenson, 1999, p. 483). The job of the control system is to rely on learning and higher cognition to affect both what stimuli activate the core system and what responses the core system produces.

It is through regulation that individual and cultural variables affect the manifestation of basic emotions. The learning history of the individual, and his or her cognitive capacities and personality traits affect both the input and the output sides of the core system. For example, social rules about what is appropriate or inappropriate in the affective domain have a profound effect on how basic emotions are manifested. Display rules about facial expressions have been especially prominent in the debate on basic emotions (Ekman 1972), but a great many other social rules also likely affect how basic emotions are manifested.

Another source of interaction-dependent variability is connected to the "occurrence of multiple emotions in response to the same event" (Roseman, 2011, p. 435). For example, if my basic fear interacts with my basic anger, as it may happen when a menacing adolescent in a parking lot viciously insults me for no good reason, the manifestations of both emotions will be affected. Another source of interaction-dependent variability is connected to what Roseman (p. 436) called "other motivational, cognitive, and situational determinants," which may include "physical activity, eating, and sleep deprivation" and "can alter physiological responses that are also affected in emotion (such as heart rate, cortisol secretion, and serotonin levels)."

This list of potentially interacting mental states is not exhaustive, but it does point to the fact that the new BET should ideally make its predictions about the bodily and neural changes associated with basic emotions sensitive to which other mental states are activated while the basic emotion is under way. This is because the same basic emotion can lead to different manifestations depending on what it is co-occurring with it. This problem is pressing for output-flexible basic emotions, which unlike output-rigid

basic emotions may be significantly influenced by co-occurring mental states.

A central area of future research for the new BET concerns the interaction between basic emotions and core affect. Even though I have challenged some of the proposals of psychological constructionists, I am in agreement on the importance of core affect as a building block for our emotional lives. What is important in the context of the new BET is to understand how core affect is affected by the activation of basic emotion programs, and what role changes in core affect play in either facilitating or impeding the functioning of basic emotions.

The New BET Defended

The new BET replaces the seven original commitments of traditional BET with the following six commitments:

- NB1: *Basic emotions are evolutionary adaptations, selected for because they are efficient solutions to fundamental life tasks.*
- NB2: *Basic emotions are associated with programs.*
- NB3: *Basic emotions are associated with hardwired neural circuits, but such circuits do not correspond one-to-one with any folk psychological emotion category.*
- NB4: *Basic emotion programs are elicited by automatic appraisals and can be either output-rigid or output-flexible. Output-rigid activations are associated with automatic and mandatory responses, whereas output-flexible activations are associated with automatically elicited response tendencies with control precedence that lead to context-dependent responses. Both mandatory and context-dependent responses are oriented toward solving a specific fundamental life task.*
- NB5: *Basic emotions are pancultural, present across species and emerge early in development.*
- NB6: *Basic emotions are not designated by folk psychological emotion categories such as anger, fear, happiness, and so forth, but by theoretically motivated subcategories such as unconditioned basic fear, conditioned basic fear, body-boundary violation basic disgust, core ingestive basic disgust, defensive basic anger, and so forth.*

The core idea at the heart of the new BET is that basic emotions are programs evolutionarily selected to provide generalized solutions to recurrent evolutionary problems by coordinating, in a highly context-dependent

yet task-oriented way, clusters of biological markers driven by hardwired neural programs. The new BET differs from the traditional BET in a number of important respects. First, it is no longer assumed that there is a one-to-one correspondence between neural circuits and folk psychological categories such as anger, fear, disgust, and so on. Hardwired neural circuits are only expected to be found at a much finer grain of analysis (e.g., the unconditioned basic fear circuit).

Second, it is no longer assumed that the output of basic emotions must necessarily be a rigid cascade of mandatory responses. This will only be expected if the eliciting stimulus for eliciting stimuli that are "focused, powerful, sudden, and closely match prototypical antecedent conditions" (Levenson, 2011, p. 382). In the general case, each basic emotion will be associated with prioritized response tendencies geared toward solving a specific fundamental life task in a context-dependent way. This will lead to a range of what I call *functional variants* (i.e., different ways in which facial, behavioral, autonomic, and phenomenal changes can be manifested while preserving the task-oriented nature of the responses).[20]

Third, the folk psychological taxonomy on which basic emotion theorists traditionally rely is replaced with a theoretically motivated taxonomy that aims to track bodily and neural differences that exist between basic emotions. This non-folk psychological taxonomy aims to collect emotions into subcategories about which scientifically interesting explanatory and predictive generalizations can be formulated (i.e., natural kinds of emotions; Scarantino, 2012c).

The main advantage of the new BET is that it is compatible with the empirical data on variability that psychological constructionists have so aptly used against traditional BET. This is not surprising, because the new BET introduces a number of changes specifically designed to accommodate the empirical challenges faced by traditional BET. This fact can, in principle, be used against it. A critic may object that the new BET differs so significantly from the traditional BET that describing the former as a version of the latter is a bit like describing Darwin's theory of natural selection as the new creationism.[21] Relatedly, a critic may suggest that failure to abandon basic emotion theory after the empirical data have refuted it amounts to turning basic emotion theory into an article of faith rather than a scientific theory.

My reply is that the transition from traditional BET to the new BET is in keeping with how scientific theories should be modified over time to solve the empirical anomalies they inevitably face. The relation between the new BET and traditional BET, I suggest, is the relation between two different versions of the same research program, whereas the relation between Darwinism and creationism is the relation between two distinct research programs (one of which, incidentally, is not scientific). Whereas it is sleight

of hand to use the same label for two distinct research programs, it is both legitimate and advisable to use the same label for two different versions of the same research program.

The idea that the units of scientific progress are *research programs* has been influentially defended by philosopher of science Lakatos, who was responding to Popper's (1959) view that science proceeds through a cycle of conjectures and refutations. As Kuhn (1962) convincingly argued in light of the history of science, the idea that scientists abandon scientific theories as soon as they find empirical facts that falsify them is entirely unrealistic. It fails to acknowledge that all scientific theories are born refuted or falsified, in the sense that they face a number of empirical facts that contradict the theory's predictions.

Lakatos's view was that scientific theories are best understood as *research programs* rather than collections of easily falsifiable declarative statements. Lakatos distinguished two main parts of a research program: a nonrevisable *hard core* and a revisable *protective belt*. The hard core of a research program is a set of commitments that are essential to the research program, in the sense that abandoning them amounts to giving up on the research program as a whole. For example, the three laws of motion and the law of gravitation constituted the hard core of the research program of Newtonian physics in the 19th century. These commitments were eventually abandoned in the transition from Newtonian to relativistic physics in the 20th century.

Newtonian physics, however, faced a number of empirical anomalies even in its heyday. Three especially stubborn ones concerned the orbit of the terrestrial moon, the perihelion of Mercury, and the orbit of Uranus, none of which fit the motions predicted by Newton's laws. This is why research programs need a *protective belt*, namely a set of auxiliary hypotheses whose job is to protect the hard core from refutation in the face of anomalies.

As stated by Lakatos (1970, p. 133), "[i]t is this protective belt of auxiliary hypotheses which has to bear the brunt of tests and get adjusted and re-adjusted, or even completely replaced, to defend the thus-hardened core." The auxiliary hypotheses that Newtonian physicists modified over time concerned atmospheric refraction, the propagation of light, the number of planets in the solar system, and other hypotheses changed to "digest anomalies and even turn them into positive evidence" (Lakatos, 1998, p. 24).

Unlike Kuhn (1962), Lakatos (1970) emphasized that sticking with a research program is not always rational. He thought that scientific standards are upheld only when researchers stick with a *progressive* research program, whereas they are violated when they stick with a *degenerative* research program. A progressive research program is one in which earlier

versions of the program are replaced by later versions that predict more facts and have their predictions confirmed. A research program whose protective belt is exclusively devoted to protecting the hard core from refutation, without making any new observationally confirmed predictions, would instead be degenerative.

Lakatos exemplified the difference by contrasting Newtonian physics with Marxism. Whereas Newtonian physicists replaced old auxiliary hypotheses with new ones that ultimately increased the predictive power of the theory, Marxism "lagged behind the facts and has been running fast to catch up with them" (Lakatos, 1998, p. 25), ultimately turning into a pseudoscience.

I suggest that the replacement of traditional BET with the new BET involves changes in the protective belt of the research program, while leaving the hard core intact. I take the hard core to be constituted by the following commitments: basic emotions are evolutionary adaptations; they are associated with programs and with hardwired neural circuits (although not circuits corresponding one-to-one to folk emotion categories); and they are pancultural, present across species, and emerge early in development.

The remaining commitments of traditional BET I take to belong to the protective belt. I have proposed the modification of two auxiliary hypotheses in particular: I have argued that evolved emotion programs can also be output-flexible rather than just output-rigid and that basic emotions should not be designated by theoretically heterogeneous folk psychological categories like anger, fear, disgust, and so forth.

Only time will tell if the research program of basic emotion theory continues to be progressive or becomes degenerative. If it becomes degenerative, sticking with it will indeed turn basic emotion theory into a pseudoscience. What I have argued in this chapter is that we are not there yet. Not only has the BET research program been progressive so far, but the empirical evidence that contradicts traditional BET can be accounted for with plausible modifications that lead to a new BET while leaving the hard core of the research program intact.

I want to emphasize in conclusion that the changes I have recommended are not unprecedented within basic emotion theory itself. A good example of the rejection of folk psychological emotion categories by a basic emotion theorist is offered by Panskepp (2012, p. 33), who has "pointedly chosen not to use vernacular terms for primary-process [a.k.a. basic] emotional systems" in order "to avoid part–whole confusions." As an alternative, Panskepp (2012) favors using capitalized versions of ordinary lowercase English emotion terms such as SEEKING, RAGE, and FEAR (see Scarantino, 2012c, for further discussion).

The view that basic emotion programs can have flexible outputs is also present, although far from prominent, within basic emotion theory. As early as 1990, Nesse argued that "far more useful than fixed patterns of

response are patterns and regulatory mechanisms that adjust to the needs of the current environment" (Nesse, 1990, p. 280).[22]

In his more recent work, Levenson (2011) has stated that basic emotions may be less "deterministic" in their connection to emotional responses than commonly assumed: "The influence of basic emotions on behaviors and thoughts becomes most deterministic under those conditions in which antecedent conditions closely match prototypical elicitors. . . . When these conditions are *not* met, the plasticity and flexibility of the emotion system becomes more ascendant" (p. 382). This is precisely the idea I have defended in this chapter, in which I have suggested that flexibility should be expected if basic emotions are to fulfill their evolutionary functions, because of the differences that exist in the way the same fundamental tasks are instantiated in different circumstances.

Roseman (2011, p. 435), whose work I acknowledge as an inspiration for this chapter, has also argued that most of the "action tendencies hypothesized to be characteristic of emotions . . . are not fixed action patterns, but complex and flexible action programs", and correctly emphasized that "a contingent relationship between an emotion and a behavioral or physiological response is *very* different from an absence of relationship" (p. 436).

Finally, in their summary of contemporary basic emotion theory, Tracy and Randles (2011, p. 400) have suggested that "as individuals develop higher level cognitive and social capacities that allow for emotion regulation, these causal effects [of basic emotions on responses] become probabilistic, merely increasing the likelihood of emotion-congruent behavior."

What these quotes reveal is that there already is a minority position in basic emotion theory that demands changes along the lines I have recommended. I hope that the general framework I have offered here will turn this minority position into the majority view, leading to a version of BET that is informed and ultimately strengthened by the insights emerging from the critique of psychological constructionists.

Conclusion

Psychological constructionists have done a real service to affective science by bringing to center stage the variability of neural circuitry and physiological, behavioral, expressive, and phenomenological responses associated with anger, fear, disgust, and so on. As they have forcefully argued, this variability is incompatible with basic emotion theory as traditionally understood (traditional BET).

Constructionists have used the evidence for variability to support an entirely new approach to the scientific study of emotional phenomena. A defining tenet of psychological constructionism is the idea that emotion episodes can be explained without invoking specialized emotion mechanisms.

I have distinguished between two varieties of psychological constructionism, a nonconceptualist one (Russell) and a conceptualist one (Barrett), and presented some challenges for both (see also Scarantino, 2012a).

The take-home message of my chapter is that basic emotion theory can only survive the constructionist critiques if it introduces substantive rather than merely cosmetic changes. In particular, the new BET I have introduced gives up on two of the defining tenets of traditional BET: the idea that basic emotions correspond neatly with folk emotion categories, and the idea that their outputs must comprise mandatory and emotion-specific responses.

The payoff of this transformation is that the new BET is compatible with the empirical evidence that there are no signatures in the brain or body with a one-to-one correspondence with anger, fear, disgust, and so forth. Finally, I have argued that this transformation is not a sleight of hand but is instead a scientifically warranted attempt to preserve the hard core of the basic emotion research program by changing its protective belt.

Although I have defended basic emotion theory from the constructionist critique, I do not consider basic emotions as defined by the new BET to be the building blocks of all other emotional phenomena,[23] nor do I consider them theoretically more important than nonbasic emotions. Finally, I am convinced that core affect is an important building block of our emotional life, just as psychological constructionists have argued. So my ecumenical conclusion is that the new BET and psychological constructionism should engage in a cooperative venture for mutual advantage and explore which aspects of our emotional life involve changes in core affect, changes in basic emotions, and coordinated changes in both.

ACKNOWLEDGMENTS

I want to thank Jim Russell and Lisa Barrett for their meticulous commentary on the first draft of this chapter. I hope that trying to respond to their objections and requests for clarification has made the chapter better. The errors, of course, remain all mine. I also want to thank Luc Faucher for his helpful feedback on a previous draft.

NOTES

1. What BET has in common with Darwin (1872) is the idea that emotions are reliably expressed in the face, voice, and posture. Darwin's understanding of the evolutionary origins of emotional expressions, however, differed from the one endorsed by basic emotion theorists. In particular, whereas BET theorists assume that expressions are adaptations for purposes of communication, Darwin thought that expressions emerged as vestigial by-products of adaptive actions (principle of serviceable associated habits) or morphological opposites of vestigial by-products

(principle of antithesis) or direct effects of nervous excitation (principle of direct action of the nervous system).

2. The facial expressions used are "exaggerated" in a literal sense, in the sense that they are often faked by actors who overemphasize the expression.

3. In this chapter, by *basic emotion* I mean "biologically basic emotion." As clarified by Ortony and Turner (1990), there are at least two other notions of basicness: *conceptual basicness* and *psychological basicness*. I have discussed how they differ from the notion of *biological basicness* in Scarantino and Griffiths (2011).

4. Automatic processes are understood as processes that use limited cognitive resources, are quick, effortless, unattended to, and do not require volitional control (Shiffrin & Schneider, 1977). Most basic emotion theorists accept that there are exceptions to the rule that basic emotions are elicited by appraisals. Nonstandard cases of elicitation include direct stimulation of the brain, facial feedback, and drugs (Izard, 1993).

5. The idea that an organism's appraisal of the circumstances plays a key causal and differentiating role is at the heart of so-called "appraisal" theories of emotion. Appraisal theories, however, are committed neither to the automaticity of appraisal nor to the idea that what appraisals elicit are evolutionarily evolved affect programs. See Scherer, Schorr, and Johnstone (2001) for an informative collection of articles on appraisal theories.

6. Many basic emotion theorists also assume that these responses are short-lived, in the sense that they last "not hours or days" but "more in the realm of minutes and seconds" (Ekman, 1999, p. 54).

7. Since James thought that emotions are perceptions of bodily changes and that perception and bodily changes are not specific to emotions, he concluded that there are no "separate and special centres" in the brain that function as the "brain-seat" of emotions. On this view, "the emotional brain-processes not only resemble the ordinary sensorial brain-processes, but in very truth *are* nothing but such processes variously combined" (James, 1884, p. 188). At the same time, I have argued that James accepted that there are internal emotion mechanisms causally responsible for pairing "peculiarly conformed pieces of the world's furniture" with the "bodily reactions" whose perception is the emotion. From this, it follows that if we give up on James's narrow view that emotions are nothing but perceptions of bodily changes, and think of emotions more broadly as the mechanisms (the locks) that lead to bodily changes in the presence of the right stimuli (the keys), then the "brain centres" of emotions become the brain centers associated with such causal mechanisms. James never explored the possibility that the "brain-seat" of emotions may be associated with causal mechanisms rather than feelings, but basic emotions theorists have done so with inconclusive results (for further discussion, see Lindquist et al., 2012; Scarantino, 2012b; Hamann, 2012). I will argue later in the chapter (p. 000) that finding the brain-seat of basic emotions requires replacing our folk psychological affective ontology with a theoretically motivated ontology that relies on more fine-grained subcategories (e.g., unconditioned basic fear).

8. Gross and Barrett (2011, p. 13) have also distinguished between two varieties of psychological constructionism: *elemental psychological constructionism*, which "ontologically reduce[s] emotion[s] to their more basic psychological

ingredients" and *emergent psychological constructionism*, which "view[s] emotions as being more than the sum of their parts." At first blush, Barrett's model of emotions is an emergent model, whereas Russell's is an elemental model. The distinction I have introduced is different, even though it also distinguishes Barrett's model from Russell's model. My distinction hinges not on whether emotions can be reduced to their building blocks, but on whether the instantiation of an emotion requires an act of conceptualization. Conceptualist varieties of psychological constructionism say yes, and nonconceptualist varieties say no. Furthermore, both elemental and emergent varieties of psychological constructionism differ from what I have called *compositional constructionism*. This is because both elemental and emergent psychologically constructionist models deny the existence of specialized emotion mechanisms, whereas compositional constructionism allows for the existence of specialized emotion mechanisms. All that compositional constructionism requires is that emotions be analyzed in terms of their building blocks, which may include specialized emotion mechanisms. As I have defined it, compositional constructionism is compatible with both *latent variable models of emotion* (Coan, 2010), which assume that "the measured indicators of emotion covary by virtue of some common executive, organizing neural circuit or network in the brain" (p. 274) (e.g., basic emotion theory), and with *emergent variable models of emotion*, which assume that "emotions do not cause, but rather are caused by, the measured indicators of emotion, assuming no executive neural circuit or network" (Coan, 2010, p. 274; note the difference with Gross and Barrett's [2011] notion of emergence; e.g., psychological constructionism). This is to say that all psychologically constructionist models are compositionally constructionist, but not all compositionally constructionist models are psychologically constructionist.

9. I am here assuming that conceptual capacities require cognitive resources that are unavailable to infants and animals, an assumption that Barrett (2012, p. 423) appears to share.

10. Note that neither problem affects NCPC, according to which the emotions are instantiated if the components match the relevant prototype (whether or not the emoters themselves know it), and causal factors other than categorization are causally responsible for bringing about the components.

11. Gross and Barrett (2011) have argued that the notion of regulation properly applies only to what they have called *emergent* varieties of psychological constructionism, according to which an emotion cannot be reduced to its component parts. Barrett's CAT is an example of an emergent model. See footnote 7 for further discussion of this distinction.

12. This is not to say that basic emotion theorists assume that every folk psychological emotion category designates a basic emotion. For instance, Ekman (1999) considered some folk emotion categories to designate *emotional plots* (e.g., jealousy, love), others to designate *moods* (e.g., irritability), and still others to designate *personality traits* (e.g., hostility).

13. Although psychological constructionists deny that there are specialized emotion mechanisms causally responsible for coordinating physiological, expressive, behavioral, and phenomenological responses, they neither deny that there

exist causal mechanisms other than emotions behind such responses, nor that emotions themselves have causal powers.

14. I complicate this "subset" picture a bit (Scarantino, 2012a), but for the purposes of this chapter, this formulation will do.

15. Even though important neural differences have been unveiled between memory systems (e.g., see chapters on memory in Gazzaniga (2004), it is generally assumed that no memory system corresponds one-to-one to any brain region.

16. Panskepp (1998, 2012) has proposed that there are at least seven basic or *primary process emotions*: SEEKING, RAGE, FEAR, LUST, CARE, PANIC/GRIEF, and PLAY. *Secondary process emotions* emerge when classical and instrumental learning build on the unconditioned rewards and punishments associated with primary process emotions to enlarge the scope of stimuli that trigger emotions. *Tertiary process emotions*, finally, emerge when higher order cognitive processes such as thinking, ruminating, fantasizing, and so forth, interact with primary and secondary processes.

17. LeDoux (2012) assumes that basic emotion theory is committed to using folk psychological categories such as anger, fear, happiness, and so on, to designate basic emotions. Since such folk categories do not uniquely correspond to specific circuits in the brain and LeDoux wants to provide a brain-based taxonomy of affective phenomena, he rejects basic emotion theory. As an alternative, LeDoux offers an analysis of what he calls *survival circuits*, which are the "circuits involved in defense, maintenance of energy and nutritional supplies, fluid balance, thermoregulation, and reproduction" (p. 655). Although the terminology used by LeDoux differs from mine, what motivates his rejection of basic emotion theory is precisely what motivates my attempt to replace traditional BET with the new BET. Both proposals are responses to the realization that folk psychological categories are massively variable with respect to their neural circuitry, physiological changes, expressions, behaviors, and phenomenology. LeDoux's suggestion is to use neologisms (e.g., survival circuit) to capture affective phenomena that share neural circuitry, whereas my proposal is to follow the lead of memory studies and add qualifiers to folk categories (e.g., basic unconditioned fear) to capture affective phenomena that share neural circuitry, as well as physiological, expressive, behavioral, and phenomenological manifestations.

18. LeDoux (2012) uses a different terminology to draw the same distinction, differentiating between *defense reactions elicited by unconditioned threats* and *defense reactions elicited by conditioned threats*.

19. Roseman (2011) has also mentioned the role of action tendencies and emotional intensity in accounting for variability. I have discussed both topics under the heading of context-dependent variability.

20. Similarly, Roseman (2011, p. 441) has characterized the responses associated with basic emotions as a "functional behavior class".

21. I thank the volume editors for pressing me to consider this objection.

22. I thank Luc Faucher for pointing me to this quotation.

23. To use a distinction introduced in Scarantino and Griffiths (2011), I do

not consider emotions that are *biologically basic* to also be p*sychologically basic* (i.e., to be building blocks of all other emotions and affective phenomena).

REFERENCES

Barrett, L. F. (2006a). Are emotions natural kinds? *Perspectives on Psychological Science, 1*, 28–58.

Barrett, L. F. (2006b). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review, 10*, 20–46.

Barrett, L. F. (2009). Variety is the spice of life: A psychological construction approach to understanding variability in emotion. *Cognition and Emotion, 23*, 1284–1306.

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

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences. 22*, 577–660.

Blanchard, R. J., & Blanchard, D. C. (2003). What can animal aggression research tell us about human aggression? *Hormones and Behavior, 44*, 171–177.

Boyd, R. (1999). Kinds, complexity and multiple realization. *Philosophical Studies, 95*, 67–98.

Campos, J. J., Barrett, K. C., Lamb, M. E., Goldsmith, H. H., & Stenberg, C. (1983). Socioemotional development. In M. M. Haith & J. J. Campos (Eds.), *Handbook of child psychology: Vol. 2. Infancy and developmental psychobiology* (4th ed., pp. 783–915). New York: Wiley.

Cannon, W. (1929). *Bodily changes in pain, hunger, fear and rage*. New York, Appleton.

Chevalier-Skolnikoff, S. (1973). Facial expression of emotion in nonhuman primates. In P. Ekman (Ed.), *Darwin and facial expression* (pp. 11–89). New York: Academic Press.

Coan, J. A. (2010). Emergent ghosts of the emotion machine. *Emotion Review, 2*(3), 274–285.

Darwin, C. (1872). *The expressions of emotions in man and animals* (1st ed.). New York: Philosophical Library.

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska Symposium on Motivation, 19*, 207–283.

Ekman, P. (1980). Biological and cultural contributions to body and facial movement in the expression of emotions. In A. O. Rorty (Ed.), *Explaining emotions* (pp. 73–102). Berkeley: University of California Press.

Ekman, P. (1999). Basic emotions. In T. Dalgleish & M. Power (Eds.), *Handbook of cognition and emotion* (pp. 45–60). Chichester, UK: Wiley.

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic? *Emotion Review, 3*(4), 364–370.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behaviour. *Semiotica, 1*(1), 86–88.

Fehr, B., & Russell, J. A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General, 113*, 464–486.

Frijda, N. H. (1986). *The emotions*. Cambridge, UK: Cambridge University Press.

Frijda, N. H. (2007). *The laws of emotion*. Mahwah, NJ: Erlbaum.

Gazzaniga, M., (Ed.). (2004). *The cognitive neurosciences*. Cambridge, MA: MIT Press.
Gendron, M., & Barrett, L. F. (2009). Reconstructing the past: A century of ideas about emotion in psychology. *Emotion Review, 1*, 316–339.
Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*, 271–299.
Gross, J. J., & Barrett, L. F. (2011). Emotion generation and emotion regulation: One or two depends on your point of view. *Emotion Review, 3*(1), 8–16.
Hamann, S. (2012). Mapping discrete and dimensional emotions onto the brain: Controversies and consensus. *Trends in Cognitive Sciences, 16*(9), 458–466.
Harrison, N. A., Gray, M. A., Giarnos, P. J., & Critchley, H. G. (2010). The embodiment of emotional feelings in the brain. *Journal of Neuroscience, 30*, 12878–12884.
Izard, C. E. (1977). *Human emotions*. New York: Plenum.
Izard, C. E. (1992). Basic emotions, relations amongst emotions and emotion–cognition relations. *Psychological Review, 99*, 561–565.
Izard, C. E. (1993). Four systems for emotion activation: cognitive and noncognitive processes. *Psychological Review, 100*, 68–90.
Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science, 2*(3), 260–275.
Izard, C. E. (2011). Cognition interactions—forms and functions of emotions: Matters of emotion. *Emotion Review, 3*(4), 371–378.
James, W. (1884). What is an emotion? *Mind, 9*, 188–205.
James, W. (1890). *The principles of psychology*. New York: Holt.
Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology, 84*(3), 394–421.
Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–195). Cambridge, UK: Cambridge University Press.
Lakatos, I. (1998). Science and pseudoscience. In M. Curd & J. A. Cover (Eds.), *Philosophy of science*: *The central issues* (pp. 20–26). New York: Norton.
LeDoux, J. (2012). Rethinking the emotional brain. *Neuron, 73*, 653–676.
Lench, H. C., Flores, S. A., & Bench, S. W. (2011). Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: A meta-analysis of experimental emotion elicitations. *Psychological Bulletin, 137*, 834–855.
Levenson, R. W. (1988). Emotion and the autonomic nervous system: A prospectus for research on autonomic specificity. In H. L. Wagner (Ed.), *Social psychophysiology and emotion: Theory and clinical applications* (pp. 17–42). Chichester, UK: Wiley.
Levenson, R. W. (1992). Autonomic nervous system differences among emotions. *Psychological Science, 3*, 23–27.
Levenson, R. W. (1999). The intrapersonal functions of emotion. *Cognition and Emotion, 13*, 481–504.
Levenson, R. W. (2011). Basic emotion questions. *Emotion Review, 3*(4), 379–386.
Lewis, M. (2007). Self-conscious emotional development. In J. Tracy, R. Robins,

& J. P. Tangney (Eds.), *The self-conscious emotions: Theory and research* (pp. 134–152). New York: Guilford Press

Lindquist, K. A., Siegel, E. H., Quigley, K., & Barrett, L. F. (2013). The hundred years emotion war: Are emotions natural kinds or psychological constructions? Comment on Lench, Flores, and Bench (2011). *Psychological Bulletin, 139*, 255–263.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences, 35*, 161–162.

Mandler, G. (1990). William James and the construction of emotion. *Psychological Science, 1*, 179–180.

Matsumoto, D., Keltner, D., Shiota, M. N., O'Sullivan, M., & Frank, M. (2008). What's in a face?: Facial expressions as signals of discrete emotions. In M. Lewis, J. M. Haviland, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 211–234). New York: Guilford Press.

Matsumoto, D., & Willingham, B. (2006). The thrill of victory and the agony of defeat: Spontaneous expressions of medal winners at the 2004 Athens Olympic Games. *Journal of Personality and Social Psychology, 91*(3), 568–581.

Matsumoto, D., & Willingham, B. (2009). Spontaneous facial expressions of emotion of congenitally and non-congenitally blind individuals. *Journal of Personality and Social Psychology,, 96*(1), 1–10.

Mayr, E. (1974). Behavior programs and evolutionary strategies. *American Scientist, 62*, 650–659.

Mesquita, B. (2003). Emotions as dynamic cultural phenomena. In R. J. Davidson, K. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 871–890). New York: Oxford University Press.

Mesquita, B., & Frijda, N. H. (1992). Cultural variations in emotions: A review. *Psychological Bulletin, 112*, 197–204.

Moyer, K. E. (1976). *The psychobiology of aggression*. New York: Harper & Row.

Nesse, R. (1990). Evolutionary explanations of emotions. *Human Nature, 1*(3), 261–289.

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

Panksepp, J. (1982). Toward a general psychobiological theory of emotions. *Behavioral and Brain Sciences, 5*, 407–467.

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

Panskepp, J. (2012). In defence of multiple core affects. In P. Zachar & R. Ellis (Eds.), *Categorical versus dimensional models of affect: A seminar on the theories of Panskepp and Russell* (pp. 31–78). Amsterdam: Benjamins.

Panskepp, J., & Watt, D. (2011). What is basic about basic emotions?: Lasting lessons from affective neuroscience. *Emotion Review, 3*(4), 387–396.

Popper, K. (1959), *The logic of scientific discovery*. London: Hutchinson.

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

Rosenbaum, R. S., Köhler, S., Schacter, D. L., Moscovitch, M., Westmacott, R.,

Black, S. E., et al. (2005). The case of K. C.: Contributions of a memory-impaired person to memory theory. *Neuropsychologia*, 43, 989–1021.
Russell, J. A. (1991). Culture and the categorization of emotions. *Psychological Bulletin, 110*, 426–450.
Russell, J. A. (1994). Is there universal recognition of emotion from facial expression?: A review of the cross-cultural studies. *Psychological Bulletin, 115*, 102–141.
Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*, 145–172.
Russell, J. A. (2012a). From a psychological constructionist perspective. In P. Zachar & R. Ellis (Eds.), *Categorical versus dimensional models of affect: A seminar on the theories of Panskepp and Russell* (pp. 79–118). Amsterdam: Benjamins.
Russell, J. A. (2012b). Final remarks. In P. Zachar & R. Ellis (Eds.), *Categorical versus dimensional models of affect: A seminar on the theories of Panskepp and Russell* (pp. 279–300). Amsterdam: Benjamins
Scarantino, A. (2009). Core affect and natural affective kinds. *Philosophy of Science, 76*, 940–957.
Scarantino, A. (2012a). Discrete emotions: From folk psychology to causal mechanisms. In P. Zachar & R. Ellis (Eds.), *Categorical versus dimensional models of affect: A seminar on the theories of Panskepp and Russell* (pp. 135–154). Amsterdam: Benjamins.
Scarantino, A. (2012b). Functional specialization does not require a one-to-one mapping between brain regions and emotions. *Behavioral and Brain Sciences, 35*, 161–162.
Scarantino, A. (2012c). How to define emotions scientifically. *Emotion Review, 4*(4), 358–368.
Scarantino, A., & Griffiths, P. (2011). Don't give up on basic emotions. *Emotion Review, 3*(4), 444–454.
Schachter, S., & Singer, J. E. (1962). Cognitive, social and physiological determinants of emotional state. *Psychological Review*, 69, 379–399.
Schacter, D. L. (2004). Introduction. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (Vol. 3, pp. 643–645). Cambridge, MA: MIT Press.
Scherer, K. R., Schorr, A., & Johnstone, T. (2001). *Appraisal processes in emotion: Theory, methods, research*. New York: Oxford University Press.
Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127–190.
Skinner, B. F. (1953). *Science and human behavior*. New York, Macmillan.
Spencer, H. (1855). *Principles of psychology*. London: Longman, Brown, Green & Longmans.
Stephens, C. L., Christie, I. C., & Friedmana, B. H. (2010). Autonomic specificity of basic emotions: Evidence from pattern classification and cluster analysis. *Biological Psychiatry, 84*, 463–473.
Tomkins, S. S. (1995). *Exploring affect: The selected writings of Silvan S. Tomkins* (E. V. Demos, Ed.). Cambridge, UK: Cambridge University Press.
Tooby, J., & Cosmides, L. (2000). Evolutionary psychology and the emotions. In

M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 91–116). New York: Guilford Press.

Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review, 3*(4), 397–405.

Waller, B. M., Vick, S.-J., Parr, L., Bard, K. A., Pasqualini, M. S., Gothard, K. M., et al. (2006). Intramuscular electrical stimulation of facial muscles in humans and chimpanzees: Duchenne revisited and extended. *Emotion, 6*(3), 367–382.

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