Core Affect and Natural Affective Kinds

Andrea Scarantino†‡

It is commonly assumed that the scientific study of emotions should focus on discrete categories such as fear, anger, sadness, joy, disgust, shame, guilt, and so on. This view has recently been questioned by the emergence of the "core affect movement," according to which discrete emotions are not natural kinds. Affective science, it is argued, should focus on core affect, a blend of hedonic and arousal values. Here, I argue that the empirical evidence does not support the thesis that core affect is a more "natural" category than discrete emotions. I conclude by recommending a splitting strategy in our search for natural affective kinds.

1. Introduction. For approximately 25 centuries, theorists of emotions have disagreed about the nature of emotions. These disagreements have taken place on the background of a shared methodological presupposition, namely, that the scientific study of emotions should focus on discrete emotion categories such as fear, anger, sadness, joy, disgust, shame, surprise, guilt, and so on.

In recent times, this position has been questioned by the emergence of the "core affect movement" (Russell 1980, 2003; Feldman 1995; Russell and Barrett 1999; Barrett 2006a). Core affect theorists have recommended a reorientation of focus from discrete emotions to core affect, defined as a combination of hedonic and arousal values. The trouble with discrete emotion categories, psychologist Lisa Feldman Barrett (2006a) has influentially argued, is that they are not natural kinds. Core affect, on the other hand, "constitutes the most basic building block of emotional life" (48), and it is more likely to qualify as a natural kind.

In this paper, I consider the empirical evidence amassed by Barrett

†To contact the author, please write to: Andrea Scarantino, Department of Philosophy and Neuroscience Institute, Georgia State University, P.O. Box 4089, Atlanta, GA 30302-4089; email: ascarantino@gsu.edu.

‡I would like to thank Eddy Nahmias, George Graham, and especially Jason Clark for their very useful comments on previous drafts of this paper.

Philosophy of Science, 76 (December 2009) pp. 940–957. 0031-8248/2009/7605-0029$10.00

(2006a) to make her case and argue that it does not support the recommended shift of focus from discrete emotions to core affect. The evidence suggests that neither core affect nor folk discrete emotions qualify for natural kind status. The methodological implication I draw is that finding natural kinds in affective science will demand splitting both discrete folk emotion categories and core affect into inductively and explanatorily homogeneous subcategories.

2. The Empirical Evidence Introduced. Until recently, the debate on the natural kind status of emotions has been a philosophers' game (Griffiths 1997, 2004; Charland 2002; Prinz 2004b). This state of affairs changed when psychologist Barrett (2006a) launched a groundbreaking attack on the "natural kind paradigm" in affective science, leading to vigorous debate among emotion scientists (e.g., Barrett et al. 2007b; Izard 2007; Panksepp 2007). Central to this debate is the commonly invoked, yet slippery, notion of a natural kind.

According to Barrett (2006a, 29), a natural kind is a "category . . . given to us by nature . . . discovered, not created, by the human mind." The instances of a natural kind "cluster together in a meaningful way because they have something real in common" (29). The contrast here is the one between kinds whose defining properties cluster together by virtue of a natural process of some sort (e.g., kinds such as "dog" or "galaxy") and kinds whose defining properties cluster together merely by virtue of an arbitrary convention among language users (e.g., kinds such as "pet" or "constellation Aries"). Crucially, natural groupings of objects are assumed to form suitable scientific categories, whereas arbitrary groupings are not (e.g., astronomers study the nature of galaxies, but they do not study the nature of constellations).

Barrett (2006a) suggests that there are two ways to cash out the general idea that some kinds are natural rather than conventional. On one interpretation, a "category constitutes a natural kind if every instance of the kind . . . shares a collection of features or properties that co-occur" (29). For instance, "for anger to be a natural kind, all instances of anger should have a characteristic facial display, cardiovascular pattern, and voluntary action that are coordinated in time and correlated in intensity" (29). On another interpretation, a category may be considered natural by virtue of "some underlying causal structure or mechanism that makes a set of instances the kind that they are (and not some other kind)" (29). In this case, what gives identity to the kind is the activation of a causal mechanism, which may produce different sets of co-occurring properties in different circumstances.

Barrett's argument, in a nutshell, is that the empirical data show that the "the psychological events referred to by the English words anger,

sadness, fear, disgust [etc.]" (Barrett et al. 2007b, 297–298) lack both an emotion-specific profile of measurable affective responses and an emotion-specific causal mechanism. I will call the categories formed by the psychological events referred to by English words such as "anger," "sadness," "fear," and so on *discrete folk emotion categories*. To be an instance of a discrete folk emotion category E (shortly, to be a folk emotion) is simply to be appropriately called "E" in ordinary English. Let us now turn to the empirical data.

2.1. Data on Phenomenology and Autonomic Changes. Consider first the phenomenology and physiology associated with discrete folk emotion types. Up until the beginning of the twentieth century, the feeling theory constituted the dominant paradigm in emotion theory. According to it, the essential component of a discrete emotion is the phenomenological one: emotions are essentially ways of feeling. William James offered the most detailed version of the feeling theory, in that he tried to spell out what the feelings of emotions were constituted by. His influential proposal was that the way emotions feel is the way perceptions of bodily changes feel: "*our feeling of* [bodily] *changes as they occur* IS *the emotion*" (James 1884, 189–190). Under this view, fear is the feeling resulting from the perception of autonomic bodily changes such as "quickened heart-beats," "shallow breathing," "weakened limbs," and "visceral stirrings." Neo-Jamesian theorists such as Damasio (1999) and Prinz (2004a) have recently defended a neurobiologically updated version of James's theory of emotions.

Barrett (2006a) points out that the thesis that discrete emotions such as anger, fear, or sadness have either phenomenological or autonomic signatures is at present not well supported by the empirical data. The first problem is that subjective experiences do not seem to cut as finely as folk emotion categories do. Lexically, we commonly distinguish between negative emotions such as sadness, fear, anger, indignation, and annoyance. But self-reports of "negative emotion experience tend to correlate so highly that measures of sadness, fear, and other categories of negative emotions often fail to capture any unique variance" (Barrett 2006a, 35). This is to say that the subjective experiences associated with distinct negative emotions, say indignation and annoyance, arguably do not differ enough to type-identify the specific discrete emotions they are. As philosopher Bedford (1957, 79) put it, "I certainly find no feeling, or class of feelings, that marks off indignation from annoyance, and enables me to distinguish them from one another."

There is a further reason, not mentioned by Barrett (2006a), why subjective experience is unlikely to have a one-to-one correspondence with discrete folk emotion types. Emotions such as fear, anger, guilt, shame,

sadness, and many others can be instantiated in the absence of subjective experiences, that is, unconsciously. Even though some have questioned the very possibility of emotions without subjective experience (e.g., Clore 1994; Hatzimoysis 2007), I believe that the case for unconscious emotions is strong (Scarantino n.d.).

Granting that there is no one-to-one correspondence between types of subjective experiences and folk emotion types, however, is not granting that there is no one-to-one correspondence between types of autonomic changes and folk emotion types. One may argue that on some occasions emotion-specific autonomic changes give rise to distinctive feelings and on other occasions they do not, for instance, when they are unconsciously perceived. The problem is that even this more limited claim is hard to maintain in light of the empirical evidence.

Barrett (2006a, 41) points out that "the issue of autonomic specificity for different categories of emotion remains controversial, even after 100 years of research." Admittedly, a number of studies have presented evidence for autonomic specificity (e.g., Ekman, Levenson, and Friesen 1983; Levenson, Ekman, and Friesen 1990; Stemmler et al. 2001; Christie and Friedman 2004). At the same time, the most systematic recent metareview of the literature on autonomic differentiation has concluded that "even a limited set of discrete emotions such as happ[iness], sad[ness], fear, anger, and disgust cannot be fully differentiated by visceral activity alone" (Cacioppo et al. 2000, 184).

It appears that instances of folk emotion categories share general autonomic nervous system (ANS) *tendencies*, without any folk emotion type having a one-to-one correspondence with a distinctive ANS *signature*. Cacioppo et al. (2000, 180) reported several studies showing that heart rate increase tends to be higher in fear than in anger, higher in anger than in happiness, higher in both fear and anger than in sadness, and higher in anger, fear, happiness, and sadness than in control conditions. Disgust, on the other hand, does not seem to differ from control conditions with respect to any autonomic measure. Cacioppo et al. also found that diastolic blood pressure tends to be higher in anger than in fear, sadness, or happiness and higher in sadness than in happiness. Anger also appears to differ from fear because it tends to be associated with larger increases of nonspecific skin conductance responses, facial temperature, and finger pulse volume and smaller increases in stroke volume and cardiac output. Finally, Cacioppo et al. suggested that visceral differentiation is "clearer when positive and negative emotions are contrasted than when discrete emotions are contrasted" (184), a view core affect theorists widely endorse.

2.2. Data on Instrumental and Expressive Behaviors. Consider now the idea that we can find a one-to-one correspondence between discrete folk

emotion types and behavior patterns. We can distinguish two versions of this idea, depending on whether we focus on *physical behaviors* (e.g., running, striking) or *expressive behaviors* (e.g., grimacing, crying). The conviction that emotions can be distinguished from one another in terms of physical behaviors lies at the foundation of behaviorist theories of emotions (Ryle 1949; Skinner 1953). In an updated form, it is still popular in some quarters (e.g., Frijda 1986, 2007). As Skinner (1953, 162–163) put it, "when the man in the street says that someone is afraid or angry or in love, he is generally talking about predispositions to act in certain ways."

The difficult part is to offer a detailed account of these predispositions to act. The problem is that the same type of discrete folk emotion can be associated with different types of instrumental behaviors in different circumstances. This flexibility is related to the function of (many) discrete emotions, which appear to work as "decoupled reflexes," with a latency period between the registration of a stimulus and the execution of a behavioral response (Scherer 1994). The latency period allows for the selection of a behavior appropriate to the stimulus, and it may lead to significantly different behaviors depending on the circumstances of stimulus elicitation.

There is empirical evidence, for instance, that rats will choose behaviors related to their distance to the threat, *orienting* when the threat is remote, *freezing* when it is proximal, and engaging in *fight-or-flight* behaviors when the threat is imminent (see Fanselow 1994; Bouton 2005). Systematic correspondences between folk emotion types and types of instrumental behaviors can be found only when the latter are functionally described. For example, fear tends to lead to threat-coping behaviors, anger to attack behaviors, guilt to reparation behaviors, and so on. Any attempt to find more specific, nonfunctionally described behavioral predispositions with a one-to-one correspondence with discrete folk emotion types appears doomed to failure.

Expressive facial (and vocal) behaviors are another standard candidate for emotion specificity. Tomkins (1995) argued, for instance, that the best way to taxonomize discrete emotions is by facial expressions, because "affects are primarily facial behaviors and secondarily outer skeletal and inner visceral behavior" (217). This view led to the emergence of *affect program theory*, developed principally by Izard (1977, 1994) and Ekman (1994, 1999a, 1999b). On this view, anger, fear, disgust, surprise, happiness, sadness, and other discrete emotions are *affect programs*, namely biologically based pan-cultural suites of short-term, coordinated responses evolved to deal with fundamental life tasks in ancestral environments.

Barrett (2006a) acknowledges that there is some evidence for the universality of the facial expressions associated with some discrete emotions.

This evidence, interpreted by affect program theorists as supporting their theory, comes mostly from cross-cultural studies that use a technique pioneered by Darwin ([1872] 1998). It consists of showing pictures of emotional expressions and asking observers what emotions they express from a list of six to 10 emotion terms in the observer's language. As reported by Ekman (1999b, 305), experiments of this sort have so far been performed with observers from 21 literate countries, and they revealed "an extraordinary amount of agreement about which emotion was shown in which photographs across the 21 countries." This is true especially with respect to happiness, sadness, and disgust.

This being said, Barrett (2006a, 37) points out that "more recent times have seen an ongoing, lively debate over the relation of emotion to the face, and to a lesser extent the voice (e.g., see Ekman, 1994; Izard, 1994; Keltner & Ekman, 2000; Russell, 1994, 1995; Russell et al., 2003)." Some of the issues raised have to do with methodological concerns about the cross-cultural techniques used by affect program theorists (see the debate between Russell [1994] and Ekman [1994]).

Even bracketing these, there seem to be substantive reasons to reject the idea that we can find a one-to-one correspondence between discrete folk emotion types and types of facial expressions.[1] Fridlund (1997) and Russell (1997) have influentially argued that a great many emotional expressions are specifically directed toward a recipient, rather than automatically broadcast, and aim to generate an effect in the recipient that is advantageous to the sender (see also Griffiths and Scarantino, forthcoming). Since what message is strategically advantageous depends largely on the circumstances, "there may be no prototypes [*sic*] faces for each [discrete] category. Rather, displays exert their influence in the particular context of their issuance" (Fridlund and Duchaine 1996, 278).

In some contexts, anger may be expressed, as posited by affect program theorists, by fixed stare, widened eyes, contracted eyebrows, bared teeth, and compressed lips. In other contexts, anger may be expressed by facial stillness and lack of eye contact, as it sometimes happens in a lingering marital confrontation. In yet other contexts, anger may be expressed without any facial involvement at all, as in the case of unconscious anger

1. Affect program theorists such as Ekman (1999a) have given to the emotions that satisfy the definition of affect programs the same names as folk emotions (e.g., fear, anger, etc.), opening their theory to counterexamples consisting of instances of folk emotions that are not affect programs. A better strategy would have been to distinguish between folk fear and affect program fear, folk anger and affect program anger, and so on and clarify that affect program theory aimed only to shed light on the subset of emotions that are affect programs. I come back to this point below.

toward a more successful sibling expressed by not making him aware of an advantageous business opportunity.

I will add that some discrete folk emotions hardly ever have facial expressions associated with them, let alone emotion-specific ones. This seems to be the case, just to pick a few examples, for regret, resentment, and hope.

2.3. Data on Neural Circuitry. The Holy Grail of contemporary affective neuroscience is the search for separate neural circuits corresponding to different discrete emotions. Two recent meta-analyses have canvassed the data on emotion-specific neural circuitry reported in positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) studies conducted over the past 10 years (Phan et al. 2002; Murphy, Nimmo-Smith, and Lawrence 2003). These studies have "summarized the correspondence between each emotion category (as defined by everyday English words) and a frequency count of the number of significant peak voxel effects within particular brain locations" (Barrett 2006a, 43).

These meta-analyses have failed to provide strong support for the thesis that all instances of each discrete folk emotion category share an emotion-specific neural circuit. Consider fear, arguably the best-understood emotion at the neural level. The correlation between instances of fear and amygdala activation has been widely reported (Aggleton 1992; Adolphs, Tranel, and Buchanan 2005; see Barrett et al. 2007a for a meta-analysis). The problem is that the correlation may be explained in terms of psychological constructs other than fear, such as novelty (Wilson and Rolls 1993; Schwartz et al. 2003; Wright et al. 2003) and uncertainty (Whalen 1998; Davis and Whalen 2001). A case could be made that the amygdala differentially activates in response to affective significance broadly construed rather than specifically in response to threats.

Concerning disgust, Phan et al. (2002) reported that 60% of studies showed a disgust–basal ganglia correlation, and Murphy et al. (2003) reported that 70% of them detected a disgust–globus pallidus correlation (the globus pallidus is a part of the basal ganglia), as well as a disgust-insula correlation. The two meta-analyses appear to disagree with respect to anger. Phan et al. did not report any evidence of a neural correlate for anger, whereas Murphy et al. indicated that 80% of the studies report an anger–lateral orbital frontal cortex correlation.

A similar discrepancy occurs with respect to happiness. Whereas Phan et al. (2002) reported that approximately 70% of the studies showed a happiness–basal ganglia correlation, Murphy et al. (2003) reported that about 60% of studies showed a happiness–rostral supracollosal anterior cingulate cortex correlation and a happiness–dorsomedial prefrontal cortex correlation. On the other hand, Cacioppo et al. (2000, 185) reported

"stable individual differences in activation of left and right anterior cortical areas" in correspondence with, respectively, positive and negative emotions.

Barrett (2006a) concludes that the neuroimaging evidence for neural specificity for discrete folk emotions is at best inconclusive. I would add that we have no reason to expect a one-to-one correspondence between types of neural events and folk emotion types if we do not find such one-to-one correspondence at the phenomenological, physiological, and behavioral levels. If the same folk emotion can be associated with different subjective experiences, different patterns of autonomic activation, different instrumental behaviors, and different facial and vocal expressions, neural underpinnings should be expected to mirror such differentiation.

3. From Discrete Emotions to Core Affect. Barrett (2006a) interprets the empirical evidence I have discussed as showing that discrete emotion categories are not natural kinds. By this, Barrett means both that folk discrete emotion categories are not natural kinds and that no discrete emotion categories, say the categories of affect program fear and affect program anger, are natural kinds. "The natural kind view [of discrete emotions]," she concludes, "has outlived its scientific value, and now presents a major obstacle to understanding what emotions are and how they work" (29). The corollary is nothing short of revolutionary: what affective science needs is a new paradigm. At the foundation of the new paradigm, developed most prominently by Russell (2003) and Barrett (2006b), is the idea that affective science should be first and foremost a science of core affect.

Barrett (2006a) defines "core affect" as follows:

> Core affect is characterized as the constant stream of transient alterations in an organism's neurophysiological state that represents its immediate relationship to the flow of changing events. . . . Core affect (i.e., the neurophysiological state) is available to consciousness, and is experienced as feeling good or bad (valence) and to a lesser extent as feeling activated or deactivated (arousal). (48)

Roughly speaking, "core affect" is a barometer that informs the organism of his or her "relationship" to the flow of events. The readings of the barometer are feelings, understood as blends of pleasure-displeasure and activation-deactivation.[2] These readings are represented as points along what is commonly called a circumplex structure, as seen in Figure 1 (Russell 1980).

2. Occasionally, these readings will not give rise to subjective experiences. Even though core affect is commonly felt, it is not necessary for it to be felt.

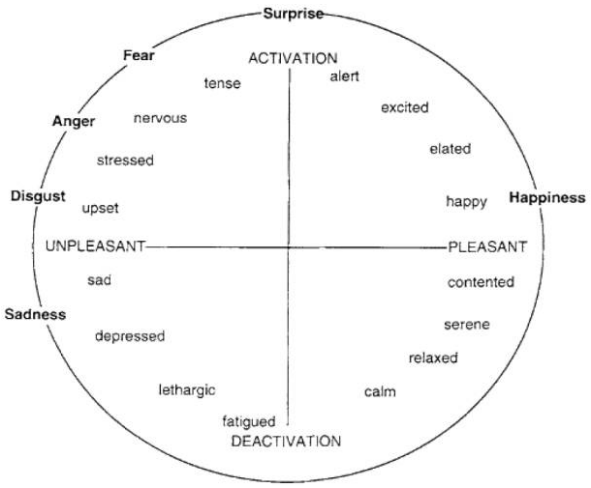

Figure 1. The circumplex.

The horizontal pleasure-displeasure dimension of the circumplex ranges from extreme unpleasantness (e.g., agony) to extreme pleasantness (e.g., ecstasy). The vertical activation-deactivation dimension ranges from extreme deactivation (e.g., sleep) to extreme activation (e.g., frantic excitement). Core affect, understood as the category comprising the set of all possible valence and arousal combinations on the circumplex, differs from discrete emotions in three crucial ways: it is *ubiquitous*, it is *objectless*, and it is *primitive*.

First, whereas one can have or fail to have a discrete emotion, one is always in a state of "core affect" because one always occupies a neurophysiological state that represents the organism's relationship to the flow of changing events.

Second, "core affect per se is not about anything . . . that is, core affect can be experienced in relation to no known stimulus—in a free-floating form" (Russell 2003, 148). On the contrary, discrete emotions are commonly about specific objects (e.g., one's girlfriend) or events (e.g., being denied tenure).

Third, "core affect" is assumed to be the primitive ingredient out of which every other affective phenomenon is built. Mood is understood as "prolonged core affect without an object" (Russell 2003, 149). Discrete emotions are understood as emerging from a conceptual act of categorization of core affect (Barrett 2006b).[3] A core affective state of high

3. A similar *two-factor theory* was proposed in the 1960s by Schachter and Singer (1962).

pleasantness and high arousal may be categorized as happiness, a state of low pleasantness and high activation as fear, a state of low pleasantness and low arousal as sadness, and so on. Barrett (2006a, 49) argues that "categorizing the ebb and flow of core affect into a discrete experience of emotion corresponds to the colloquial idea of 'having an emotion.'"

Whereas discrete emotions are unlikely to be "appropriate categories to support a cumulative science" (Barrett 2006a, 46), core affect is considered by Barrett and the core affect movement to be a more suitable scientific category. Does the empirical evidence favor this conclusion?

4. The Empirical Evidence Reassessed.

4.1. Affective Kind Essentialism Rejected. As reported in Section 2, Barrett (2006a) thinks that there are two importantly different ways to cash out the general idea that some kinds are natural rather than conventional. I will consider them in turn, expanding on Barrett's reflections. The first is that a "category constitutes a natural kind if every instance of the kind . . . shares a collection of features or properties that co-occur" (29). Under this view, all members of a natural affective kind need to share the same co-occurring properties, by virtue of which they can be embedded in scientifically interesting generalizations.

I call this understanding of natural kinds in affective science *affective kind essentialism*, to emphasize the analogy between this position and traditional essentialism in biology, the thesis that natural kinds "must possess definitional *essences* that define them in terms of necessary and sufficient . . . properties" (Boyd 1999, 146). This form of essentialism about natural kinds has its roots in Aristotle, was famously defended by Locke, and found its most influential contemporary formulation in Kripke (1972) and Putnam (1975).

The empirical evidence canvassed in Section 2, I argue, supports the following thesis:

> **No Necessity Thesis**. There is no individually necessary criterion for being an instance of a discrete folk emotion category such as fear, anger, sadness, joy, disgust, shame, surprise, guilt, and so on.

The No Necessity Thesis provides the best explanation for centuries of failed attempts to articulate definitions of folk emotions in terms of individually necessary and jointly sufficient criteria. If the No Necessity Thesis is true, there are no such criteria to be found. Importantly, the conjunction of the No Necessity Thesis and of affective kind essentialism leads to the conclusion that discrete folk emotion categories are not natural kinds, as suggested by Barrett (2006a). If all the members of a discrete emotion category E need to share the same set of co-occurring properties

in order for E to be a natural kind, folk emotion categories such as anger, fear, joy, disgust, shame, and so on do not constitute natural kinds.

The trouble is that affective kind essentialism is an unsuitable theory of natural kinds for affective science, just as it is an unsuitable theory of natural kinds for psychological and biological kinds (see Machery 2005; Samuels, forthcoming; Wilson, Barker, and Brigandt, forthcoming). The empirical evidence I discussed in Section 2 shows that there is massive variability among the instances of any discrete folk emotion E. This is true whether we focus on the phenomenological, physiological, behavioral, or neural responses associated with any folk emotion E.

The point is that this variability is in principle compatible with the presence of a scientifically important causal mechanism that is distinctive of E and capable of identifying a natural kind. This explains why Barrett (2006a) focused mostly on a second, more promising formulation of the "natural kind view."

4.2. Affective Kind Causalism and Discrete Emotions. According to a second formulation, natural kinds are defined by "some underlying causal structure or mechanism that makes a set of instances the kind that they are" (Barrett 2006a, 29). I call this understanding of natural kinds *affective kind causalism*, to emphasize that what defines a natural affective kind is in this case an underlying causal mechanism rather than a shared essence. This anti-essentialist doctrine of natural kinds also has a long history in philosophy (see Hacking 1991), and it finds its most influential contemporary formulation in Boyd's (1991, 1999) theory of homeostatic property cluster (HPC) kinds.

As Boyd (1991, 141) puts it, "the natural definition of . . . *homeostatic property cluster kinds* is determined by the members of a cluster of often co-occurring properties and by the ('homeostatic') mechanisms that bring about their co-occurrence." What makes HPC kinds natural is that the imperfect co-occurrence of the properties brought about by causal homeostatic mechanisms is important relative to the inductive and explanatory purposes of some scientific discipline. On this view, kinds are natural not *simpliciter*, but relative to a scientific discipline at a time.

Importantly, none of the cluster properties needs to be possessed by all members of an HPC kind. This is what allows HPC kinds to accommodate the variability of kinds in biology and the special sciences while preserving their suitability for induction and explanation. What makes them useful to predict and explain is their ability to track relevant causal mechanisms.

Once this more sophisticated theory of natural kinds is on board, however, Barrett's (2006a) argument against discrete emotions no longer goes through. What Barrett aimed to prove is in effect two theses (often confusedly mixed): (*a*) discrete folk emotions are not natural kinds, and (*b*)

no discrete emotions (e.g., discrete emotions as defined by affect program theory) are natural kinds.

Barrett's core idea is that if there were causal mechanisms specific to discrete emotions, then "all observable events, such as voluntary action or action tendencies, facial muscle movements, vocalizations, subjective experiences, and peripheral nervous system responses, [would] encode specific information that derives from the causal mechanism that produced them" (2006a, 35). The specificity of this information would have to translate into a one-to-one correspondence between discrete emotion types (e.g., fear) and types of phenomenological, physiological, behavioral, and neural events (e.g., amygdala activation).

This assumption explains why the bulk of Barrett's (2006a) article is spent mustering empirical support for the No Necessity Thesis. The problem is that the conjunction of the No Necessity Thesis and affective kind causalism does *not* lead either to the conclusion that discrete folk emotion categories are not natural kinds or to the conclusion that no discrete emotion categories are natural kinds. The reason is that the same causal mechanism for emotion E may in principle bring about different actions, facial expressions, subjective experiences, and peripheral nervous system responses in different circumstances, while still defining the identity of the kind. This is precisely what differentiates affective kind essentialism from affective kind causalism.[4]

This being said, I am convinced that the empirical evidence supports one-half of Barrett's case: discrete folk emotion categories are not HPC natural kinds. In the absence of room for a full discussion, the following sketched remarks will have to suffice (see Griffiths 1997, 2004 for a more developed defense of the thesis). In a nutshell, the reason why discrete folk emotion categories are not HPC kinds is that the sorts of properties that instances of the same discrete folk emotion E tend to share lack inductive and explanatory import relative to the purposes of scientific psychology.

The problem is not that instances of E do not all share the same set of *perfectly co-occurring* affective properties (this is a problem only for an essentialist theory of natural kinds), but that they do not share any set of *imperfectly co-occurring* affective properties by virtue of which they can be embedded in the same inductions and explanations. If this is the

4. In other places, Barrett seems fully aware of the problem I raised. For instance, she writes that on the causal view of natural kinds, "the anger mechanism also serves to identify an instance as anger when some factor . . . intervenes between emotion elicitation and expression and disrupts the signature profile of response" (2006a, 31). But if this is the case, evidence that there exists no 'signature profile of response' is insufficient to conclude that a given discrete emotion is not a natural kind.

case, different instances of the same discrete folk emotion E will have to be subject to different inductions and different explanations, defying the emergence of a unified scientific psychology of E. To put it otherwise, what we discover about, say, affect program fear may shed no light at all, for the purposes of scientific psychology, on unconscious fear of failing, even though they both qualify as bona fide instances of folk fear. I conclude that discrete folk emotions such as fear, anger, sadness, joy, disgust, shame, surprise, guilt, and so on are not natural HPC kinds.

The other half of Barrett's case, however, does not go through: the empirical evidence discussed in Section 2 does not support the thesis that *no* discrete emotions (e.g., discrete emotions as defined by affect program theory) are natural kinds. As long as the instances of, say, *affect program fear*—fear as defined by Ekman (1999a)—share a set of imperfectly co-occurring affective properties by virtue of which they can be embedded in the same inductions and explanations, affect program fear still qualifies as a natural kind for scientific psychology. Granted, such a natural kind should not be called *fear* simpliciter, because this labeling wrongly suggests that the kind is coextensive with the folk category of fear.[5]

Semantic issues aside, the point is that it may well be the case that folk fear is not a natural kind, that folk fear is not an affect program, but that affect program fear is a natural kind. The evidence in Section 2 in no way shows that the subset of instances of folk fear that qualify as affect program fear do not tend to share scientifically important properties, which is what would have to be demonstrated to argue that affect program fear is not an HPC kind. The central lesson we should glean from the impressive empirical evidence collected by Barrett, I conclude, is simply that no natural discrete emotion categories will be coextensive with discrete folk emotion categories.

4.3. Affective Kind Causalism and Core Affect. The real trouble for the core affect program emerges when we ask whether core affect itself is a natural kind (in the causal sense, the only one viable for affective science). Remember that the reason why core affect theorists recommended that we focus on core affect instead of discrete emotions is that the latter are supposed not to be natural kinds. Barrett explicitly claims that "the em-

5. Ekman (1999a) speaks of anger, fear, sadness, joy, disgust, etc. as if all instances of the category satisfied his definition of an affect program. But this is certainly false, in the sense that, as shown by the empirical evidence considered by Barrett, many bona fide instances of each discrete folk category are not affect programs (e.g., they lack the appropriate facial expressions, neural underpinnings, etc.). Ekman's terminological choice nurtures the misunderstanding I highlighted in the text, according to which if folk emotions are not natural kinds, affect programs that share their names with folk emotions are also not natural kinds.

pirical case supporting the hypothesis that core affect is a natural kind is suggestive" (2006a, 48), even though she admits that more evidence is needed to provide strong support for the hypothesis.

In order for core affect to be a natural kind, it must be the case that there exists an "underlying causal structure or mechanism that makes a set of instances [of core affect] the kind that they are" (Barrett 2006a, 29). This is the criterion used by Barrett to claim that discrete emotion categories are unsuitable for the accumulation of scientific knowledge. It is therefore the criterion we need to use to evaluate whether core affect is, as claimed, a more suitable scientific category.

I want to argue, contra Barrett (2006a), that the case against the natural kind status of core affect is stronger than the case against the natural kind status of discrete folk emotions. In other words, core affect is less likely to be an HPC kind than any discrete folk emotion. I emphasize that this is not to say that there is no good reason to study core affect scientifically. It is simply to say that the reason why we should study core affect had better not be that discrete folk emotions are not natural kinds, because core affect is not a natural kind either.

The trouble with core affect is that it is a significantly more heterogeneous category than any discrete folk emotion category could ever be. It is easy to see why: it contains all instances of all discrete folk emotions, and much more. Organisms are at all times in a state of core affect, by definition of core affect. From birth to death, they are located *somewhere* along the dimensions of pleasure and arousal.

This is to say that core affect does not designate any specific combination of hedonic and arousal values, but rather the entire collection of possible combinations of pleasure and arousal. It consequently includes states of folk fear, folk anger, folk joy, folk disgust, folk anxiety, folk depression, and so on. But it also includes innumerable nonemotional states that combine degrees of pleasure and arousal, for example, fatigue, sleep, serenity, focused concentration, and so on.

The argument I sketched at the end of the previous section now applies with manifold force. There exist no frequently co-occurring inductively and explanatorily important properties that instances of core affect tend to share by virtue of causal homeostatic mechanisms. This is not to say that instances of core affect do not share any properties. For example, all instances of core affect can be characterized as having some degree of valence and having some degree of arousal. The point is that no inductions or explanations of interest to scientific psychologists are licensed by this mere fact.

Different instances of core affect will have to be subject to different inductions and different explanations, defying the emergence of a unified scientific psychology of core affect. What we discover about, say, the

instances of core affect involved in affect program fear will shed no light, for the purposes of scientific psychology, about the instance of core affect involved in fatigue, even though they both qualify as bona fide instances of core affect.

A core affect theorist may reply that along several of the dimensions we have considered in this paper—phenomenological, autonomic, behavioral, and neural—there is more specificity with respect to positive and negative emotions than there is specificity with respect to any discrete folk emotion. Here, I will not take a position on whether this is actually true. The point is that even if it were true, it would not follow that core affect is an HPC kind.

At best, the evidence may support the case that "positive emotion" and "negative emotion" are natural kinds, namely, that their instances tend to share inductively and explanatorily important properties by virtue of causal mechanisms specific to, respectively, positive and negative emotions. Core affect is a different scientific category, one that includes "positive emotion" and "negative emotion" as subcategories in the same sense in which "fear" includes "affect program fear" and "unconscious fear" as subcategories.

I conclude that core affect, just like discrete folk emotions, is not a natural HPC kind.

5. Conclusion. Core affect theorists have recently proposed a shift of focus from discrete to dimensional models of emotion. The rationale for this methodological recommendation is that, according to core affect theorists, discrete emotion categories are not natural kinds, whereas core affect is more likely to be a natural kind. I have argued that the case Barrett (2006a) mounts against discrete folk emotions applies with manifold force to core affect itself. Neither core affect nor discrete folk emotion categories, I concluded, are natural kinds for the purposes of scientific psychology.

What emotion scientists should do is to keep looking for natural affective kinds. The strategy I advocate is to split folk emotion categories and look for discrete, nonfolk subcategories that are homogeneous for purposes of induction and explanation. The fact that all things we ordinarily call anger, disgust, fear, joy, sadness, and so on do not share scientifically important homeostatic causal mechanisms in no way licenses the conclusion that there are no subsets of such categories that do (e.g., the subcategory "affect program fear"). By the same token, the fact that all things we call core affect do not share scientifically important homeostatic causal mechanisms, as I have argued, in no way licenses the conclusion that there are no subsets of core affective states that do (e.g., the subcategory "positive emotion").

REFERENCES

Adolphs, R., D. Tranel, and T. W. Buchanan (2005), "Amygdala Damage Impairs Emotional Memory for Gist but Not Details of Complex Stimuli", *Nature Neuroscience* 8 (4): 512–518.

Aggleton, J., ed. (1992), *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction*. New York: Wiley.

Barrett, L. F. (2006a), "Are Emotions Natural Kinds?", *Perspectives on Psychological Science* 1: 28–58.

——— (2006b), "Solving the Emotion Paradox: Categorization and the Experience of Emotion", *Personality and Social Psychology Review* 10: 20–46.

Barrett, L. F., et al. (2007a), "The Amygdala and the Experience of Affect", *Social, Cognitive, and Affective Neuroscience* 2: 73–83.

Barrett, L. F., K. Lindquist, E. Bliss-Moreau, S. Duncan, M. Gendron, J. Mize, and L. Brennan (2007b), "Of Mice and Men: Natural Kinds of Emotion in the Mammalian Brain?", *Perspectives on Psychological Science* 2: 297–312.

Bedford, E. (1957), "Emotions", *Proceedings of the Aristotelian Society* 57: 281–304.

Bouton, M. E. (2005), "Behavior Systems and the Contextual Control of Anxiety, Fear, and Panic", in L. F. Barrett et al. (eds.), *Emotions: Conscious and Unconscious*. New York: Guilford, 205–227.

Boyd, R. (1991), "Realism, Anti-foundationalism and the Enthusiasm for Natural Kinds", *Philosophical Studies* 61: 127–148.

——— (1999), "Homeostasis, Species, and Higher Taxa", in R. A. Wilson (ed.), *Species: New Interdisciplinary Essays*. Cambridge, MA: MIT Press, 141–185.

Cacioppo, J. T., G. G. Berntson, J. T. Larsen, K. M. Poehlmann, and T. A. Ito (2000), "The Psychophysiology of Emotion", in R. Lewis and J. M. Haviland-Jones (eds.), *Handbook of Emotions*, 2nd ed. New York: Guilford, 173–191.

Charland, L. C. (2002), "The Natural Kind Status of Emotion", *British Journal for the Philosophy of Science* 53: 511–537.

Christie, I. C., and B. H. Friedman (2004), "Autonomic Specificity of Discrete Emotion and Dimensions of Affective Space: A Multivariate Approach", *International Journal of Psychophysiology* 51: 143–153.

Clore, G. (1994), "Why Emotions Are Never Unconscious", in P. Ekman and R. J. Davidson (eds.), *The Nature of Emotion: Fundamental Questions*. New York: Oxford University Press, 285–290.

Damasio, A. R. (1999), *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace.

Darwin, C. ([1872] 1998), *The Expression of the Emotions in Man and Animals*. 3rd ed. Edited by P. Ekman. New York: Oxford University Press.

Davis, M., and P. J. Whalen (2001), "The Amygdala: Vigilance and Emotion", *Molecular Psychiatry* 6: 13–34.

Ekman, P. (1994), "Strong Evidence for Universals in Facial Expressions: A Reply to Russell's Mistaken Critique", *Psychological Bulletin* 115: 268–287.

——— (1999a), "Basic Emotions", in T. Dalgleish and M. Power (eds.), *Handbook of Cognition and Emotion*. Chichester, UK: Wiley, 45–60.

——— (1999b), "Facial Expressions", in T. Dalgleish and M. Power (eds.), *Handbook of Cognition and Emotion*. Chichester, UK: Wiley, 301–320.

Ekman, P., R. W. Levenson, and W. V. Friesen (1983), "Autonomic Nervous System Activity Distinguishes between Emotions", *Science* 221: 1208–1210.

Fanselow, M. S. (1994), "Neural Organization of the Defensive Behavior System Responsible for Fear", *Psychonomic Bulletin and Review* 1: 429–438.

Feldman, L. A. (1995), "Variations in the Circumplex Structure of Emotion", *Personality and Social Psychology Bulletin* 21: 806–817.

Fridlund, A. (1997), "The New Ethology of Human Facial Expressions", in J. A. Russell and J.-M. Fernández-Dols (eds.), *The Psychology of Facial Expression*. Cambridge: Cambridge University Press, 103–129.

Fridlund, A. J., and B. Duchaine (1996), "'Facial Expressions of Emotions' and the Delusion

of the Hermetic Self", in R. Harré and W. G. Parrott (eds.), *The Emotions: Social, Cultural and Biological Dimensions*. Thousand Oaks, CA: Sage, 259–284.
Frijda, N. H. (1986), *The Emotions*. Cambridge: Cambridge University Press.
——— (2007), *The Laws of Emotion*. Mahwah, NJ: Erlbaum.
Griffiths, P. E. (1997), *What Emotions Really Are: The Problem of Psychological Categories*. Chicago: University of Chicago Press.
——— (2004), "Is Emotion a Natural Kind?", in R. C. Solomon (ed.), *Thinking about Feeling: Contemporary Philosophers on Emotions*. Oxford: Oxford University Press, 233–249.
Griffiths, P. E., and A. Scarantino (forthcoming), "Emotions in the Wild", in Murat Aydede and Philip Robbinsin (eds.), *Cambridge Handbook of Situated Cognition*. Cambridge: Cambridge University Press.
Hacking, I. (1991), "A Tradition of Natural Kinds", *Philosophical Studies* 61: 109–126
Hatzimoysis, A. (2007), "The Case against Unconscious Emotions", *Analysis* 67 (October): 292–299.
Izard, C. E. (1977), *Human Emotions*. New York: Plenum.
——— (1994), "Innate and Universal Facial Expressions: Evidence from Developmental and Cross-Cultural Research", *Psychological Bulletin* 115: 288–299.
——— (2007) "Basic Emotions, Natural Kinds, Emotion Schemas, and a New Paradigm", *Perspectives on Psychological Science* 2 (3): 260–275
James, W. (1884), "What Is an Emotion?", *Mind* 9: 188–205.
Keltner, D., and P. Ekman (2000), "Facial Expression of Emotion", in R. Lewis and J. M. Haviland-Jones (eds.), *Handbook of Emotions*, 2nd ed. New York: Guilford, 236–264.
Kripke, S. (1972), "Naming and Necessity", in D. Davidson and G. Harman (eds.), *Semantics of Natural Language*. Dordrecht: Reidel, 253–355.
Levenson, R. W., P. Ekman, and W. V. Friesen (1990), "Voluntary Facial Action Generates Emotion Specific Autonomic Nervous System Activity", *Psychophysiology* 27: 363–384.
Machery, E. (2005), "Concepts Are Not a Natural Kind", *Philosophy of Science* 72: 444–467
Murphy, F. C., I. Nimmo-Smith, and A. D. Lawrence (2003), "Functional Neuroanatomy of Emotion: A Meta-analysis", *Cognitive, Affective, and Behavioral Neuroscience* 3: 207–233.
Panksepp, J. (2007), "Neurologizing the Psychology of Affects", *Perspectives on Psychological Science* 2: 281–296.
Phan, K. L., T. D. Wager, S. F. Taylor, and I. Liberzon (2002), "Functional Neuroanatomy of Emotion: A Meta-analysis of Emotion Activation Studies in PET and fMRI", *NeuroImage* 16: 331–348.
Prinz, J. (2004a), *Gut Reactions: A Perceptual Theory of Emotion*. Oxford: Oxford University Press.
——— (2004b), "Which Emotions Are Basic?", in P. Cruise and D. Evans (eds.), *Emotion, Evolution and Rationality*. Oxford: Oxford University Press, 69–87.
Putnam, Hilary (1975), "The Meaning of 'Meaning'", in *Mind, Language and Reality*. Cambridge: Cambridge University Press, 215–271.
Russell, J. A. (1980), "A Circumplex Model of Affect", *Journal of Personality and Social Psychology* 39: 1161–1178.
——— (1994), "Is There Universal Recognition of Emotion from Facial Expression? A Review of the Cross-Cultural Studies", *Psychological Bulletin* 115: 102–141.
——— (1995), "Facial Expressions of Emotion: What Lies beyond Minimal Universality?", *Psychological Bulletin* 118: 379–391.
——— (1997), "Reading Emotion from and into Faces: Resurrecting a Dimensional-Contextual Perspective", in J. A. Russell and J.-M. Fernández-Dols (eds.), *The Psychology of Facial Expression*. Cambridge: Cambridge University Press, 295–320.
——— (2003), "Core Affect and the Psychological Construction of Emotion", *Psychological Review* 110: 145–172.
Russell, J. A., J.-A. Bachorowski, and J.-M. Fernández-Dols (2003), "Facial and Vocal Expressions of Emotion", *Annual Review of Psychology* 54: 329–349.
Russell, J. A., and L. F. Barrett (1999), "Core Affect, Prototypical Emotional Episodes, and

Other Things Called Emotion: Dissecting the Elephant", *Journal of Personality and Social Psychology* 76: 805–819.
Ryle, G. (1949), *The Concept of Mind*. London: Hutchinson.
Samuels, R. (forthcoming), "Delusion as a Natural Kind", in M. Broome and L. Bortolotti (eds.), *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*. Oxford: Oxford University Press.
Scarantino, A. (n.d.), "The Case for Unconscious Emotions", unpublished manuscript.
Schachter, S., and J. E. Singer (1962), "Cognitive, Social, and Physiological Determinants of Emotional State", *Psychological Review* 69: 379–399.
Scherer, K. (1994), "Emotion Serves to Decouple Stimulus and Response", in P. Ekman and R. J. Davidson (eds.), *The Nature of Emotion: Fundamental Questions*. New York: Oxford University Press, 127–130.
Schwartz, C. E., C. I. Wright, L. M. Shin, J. Kagan, P. J. Whalen, K. G. McMullin, and S. L. Rauch (2003), "Differential Amygdalar Response to Novel vs. Newly Familiar Neutral Faces: A Functional MRI Probe Developed for Studying Inhibited Temperament", *Biological Psychiatry* 53: 854–862.
Skinner, B. F. (1953), *Science and Human Behavior*. New York, Macmillan.
Stemmler, G., M. Heldmann, C. A. Pauls, and T. Scherer (2001), "Constraints for Emotion Specificity in Fear and Anger: The Context Counts", *Psychophysiology* 38: 275–291.
Tomkins, S. S. (1995), *Exploring Affect: The Selected Writings of Silvan S. Tomkins*. Edited by E. V. Demos. Cambridge: Cambridge University Press.
Whalen, P. J. (1998), "Fear, Vigilance, and Ambiguity: Initial Neuroimaging Studies of the Human Amygdala", *Current Directions in Psychological Science* 7: 177–188.
Wilson, F. A. W., and E. T. Rolls (1993), "The Effects of Stimulus Novelty and Familiarity on Neuronal Activity in the Amygdala of Monkeys Performing Recognition Memory Tasks", *Experimental Brain Research* 93: 367–382.
Wilson, R. A., M. J. Barker, and I. Brigandt (forthcoming), "When Traditional Essentialism Fails: Biological Natural Kinds", *Philosophical Topics*.
Wright, C. I., B. Martis, C. E. Schwartz, L. M. Shin, H. H. Fischer, K. McMullin, and S. L. Rauch (2003), "Novelty Responses and Differential Effects of Order in the Amygdala, Substantia and Inferior Temporal Cortex", *NeuroImage* 18: 660–669.