emotion review

Emotion Review
Vol. 4, No. 4 (October 2012) 358–368

ISSN 1754-0739
DOI: 10.1177/1754073912445810
er.sagepub.com

How to Define Emotions Scientifically

Andrea Scarantino
Department of Philosophy and Neuroscience Institute, Georgia State University, USA

Abstract

The central contention of this article is that the classificatory scheme of contemporary affective science, with its traditional categories of *emotion, anger, fear*, and so on, is no longer suitable to the needs of affective science. Unlike psychological constructionists, who have urged the transition from a discrete to a dimensional approach in the study of affective phenomena, I argue that we can stick to a discrete approach as long as we accept that traditional emotion categories will have to be transformed in order to do any scientific work. I conclude by articulating some general rules for turning traditional emotion categories into suitable scientific tools.

Keywords
affect, basic emotions, core affect, emotion, natural kinds, psychological constructionism

The central contention of this article is that the classificatory scheme of contemporary affective science, with its traditional discrete categories of *emotion*, *anger*, *fear*, and so on, is no longer suitable to the needs of affective science. Elements of this position have been anticipated in the literature, both in emotion science (e.g., Barrett, 2006; Duffy, 1934, 1941; Kagan, 2007, 2010; Mandler, 1975; Russell, 2003; Zachar, 2006) and in philosophy (e.g., Griffiths, 1997, 2004; Rorty 1980).

However, these calls for conceptual revision have yet to produce significant changes in the way emotions are studied. This is in part because the alternative classificatory schemes proposed have lacked details or have been considered unappealing on other grounds. In this paper, I want to once again defend the view that affective science needs a new classificatory scheme and provide some details on how to build it.

Unlike psychological constructionists, who have urged emotion scientists to transition from a discrete to a dimensional approach in the study of affective phenomena (Barrett, 2006; Russell, 2003), I argue that we can stick to a discrete approach as long as we accept that traditional emotion categories will have to be heavily revised in order to do scientific work.

The core idea I defend is that the job of an emotion scientist should be to search for *natural kinds* in affective science, namely categories that are theoretically homogenous for purposes of scientific extrapolation. Although I agree with psychological constructionists that traditional emotion categories do not designate natural kinds, I disagree that natural kinds can only be found among primitive building blocks of affective phenomena, and will explain how we can search for natural kinds of discrete emotions.

Since the notion of a natural kind plays a key role in what follows, I first clarify how it will be used. I then summarize previous critiques of the thesis that emotions are natural kinds, and introduce my own. Finally, I introduce some ground rules for transforming traditional emotion categories into scientific tools.

What Are Natural Kinds?

The notion of a natural *kind*, namely a *grouping* or *set* or *class* of items that is natural, originated in philosophy (cf. Hacking, 1991), but it is at this point commonly invoked in psychology as well (Barrett, 2006; Barrett et al., 2007; Izard, 2007; Panksepp, 2007). Two intuitions have been central in anchoring the distinction between natural and nonnatural kinds. The first is an intuition of *ontological independence*, according to which natural kinds are nonarbitrary or nonconventional categories that exist in nature independently of our acknowledgment of them. Echoing this intuition, Barrett (2006, p. 29) states that a natural kind is a "category … given to us by nature … discovered, not created, by the human mind."

Corresponding author: Andrea Scarantino, Department of Philosophy, Georgia State University, P.O Box 4089, Atlanta, GA 30302-4089, USA. *Email*: ascarantino@gsu.edu

The second is an intuition of *epistemic usefulness*, according to which natural kinds are categories that are maximally suitable for the epistemic aims of scientists, most importantly predicting and explaining natural phenomena. According to Boyd, for instance, "the naturalness of natural kinds consists in their aptness for induction and explanation" (Boyd, 2007, p. 410).

Paradigmatic examples of natural kinds include chemical kinds such as *water*, physical kinds such as *mass*, and biological kinds such as *tiger*. These kinds do not result merely from an arbitrary human convention, and they are maximally suitable for the epistemic aims of, respectively, chemists, physicists, and biologists. This is because they participate in a large body of empirically discovered reliable generalizations on account of the many scientifically interesting properties they have in common.

For example, chemists have discovered lots of reliable generalizations about *water*, concerning its boiling point, its melting point, its valence, its viscosity, its surface tension, its solvent properties, and so on. And once they make a new discovery about a sample of water, they can reliably project it—with an attached likelihood—to the whole kind, thus accumulating more and more scientific knowledge about *water*.

In contrast, nonnatural kinds such as "things taller than 1 meter," "Libra constellation," or "not water" have arbitrary boundaries and are unsuitable for the epistemic aims of scientists. No scientifically reliable generalizations useful for the prediction and explanation of natural phenomena seem to be available with respect to such groupings. And once a new discovery is made about a member of a nonnatural class, say the class of "things taller than 1 meter," that discovery *cannot* be reliably projected to the rest of the category, which contains widely heterogeneous items. Thus, nonnatural kinds are not suitable categories for the accumulation of scientific knowledge.

Before turning to emotions, I note that the notion of natural kinds I have sketched differs from other notions sometimes invoked in psychology, sociology, and anthropology. In particular, natural kinds are often understood in such disciplines as kinds that are *not socially constructed*, either in the material sense that they are *not* artefacts like *tables* and *chairs* or in the metaphorical sense that they are not the result of contingent social factors like *money* or the *monarchy*. This is a different notion of natural kind from the one I will be relying on in what follows.

The Natural Kind Assumption in Emotion Theory

When scientists produce anger in experimental subjects, and notice that in a statistically significant number of experiments a particular brain area differentially lights up, they generally conclude that when a subject is in anger, the differential activation of that brain region is likely to result. This inference presupposes that scientific discoveries made about samples of anger can be extrapolated to the rest of the anger category.[1] This amounts to assuming that *anger* is a natural kind.

For most of its history, affective science has relied on the *natural kind assumption* (NKA)—the assumption that either emotion or the categories subordinate to it, or both, are natural kinds. NKA has provided unity of purpose to an otherwise fragmented research field. Competing scientific theories of emotion/anger/fear/etcetera vehemently disagree on content, but they generally agree there is nothing wrong with emotion, anger, or fear as scientific concepts.

This being said, NKA has periodically been challenged, both in emotion science and in philosophy (e.g., Barrett, 2006; Duffy, 1934, 1941; Griffiths, 1997, 2004; Kagan, 2007, 2010; Mandler, 1975; Rorty, 1980; Russell, 2003; Zachar, 2006). Here, I consider two especially influential recent challenges that will constitute the starting point of my own critique.

Griffiths on Why Emotion Is Not a Natural Kind

Philosophers are responsible for bringing the language of natural kinds to bear on the question of whether *emotion* is a theoretically homogenous category for purposes of scientific extrapolation. Rorty (1980, p. 1), for instance, argued that "[e]motions do not form a natural class." This idea has been most forcefully articulated by Griffiths (1997, 2004) in recent years. According to Griffiths, *emotion* is not a natural kind because it contains instances of at least three different types: *basic emotions*, *higher cognitive emotions*, and *socially sustained pretenses*. Let us consider them in turn.

According to Ekman (1999, p. 46), the most prominent defender of basic emotion theory, "emotions evolved for their adaptive value in dealing with *fundamental life tasks*" such as avoiding predators, progressing towards goals, coping with losses, finding mates, fighting, and so on (Ekman, 1999; Tooby & Cosmides, 2000). Paradigmatic examples of basic emotions comprise fear, anger, disgust, happiness, sadness, and surprise. Each of these basic emotions, Ekman (1999) emphasized, forms a *family*. Instances of irritation, rage, and fury, for instance, all belong to the *anger family*.

Basic emotions are defined by markers such as distinctive universal signals, distinctive physiology, autonomic appraisal tuned to distinctive universals in antecedent events, distinctive developmental appearance, distinctive thoughts, memories and images, distinctive subjective experience, quick onset, brief duration, unbidden occurrence, and presence in other primates (Ekman, 1999).

Griffiths argued that this account does not fit emotions such as envy, moral indignation, resentment and others, which seem to lack quick duration, are not present in other primates, and require "responding in a more cognitively complex way to more highly analyzed information" (2004, p. 236) than basic emotions do. Griffiths proposed to call such emotions *higher cognitive emotions*.

Finally, Ekman's account does not fit emotions that "involve an internalized cultural model of appropriate behaviour." This is the case for emotions such as "going postal," which follow a script "derived from real or fictional incidents that are culturally salient" (Griffiths, 2004, p. 236). He proposed to call such emotions *socially sustained pretenses*.

Griffiths' key point is that higher cognitive emotions and socially sustained pretenses differ from basic emotions so

significantly that they cannot be comprised within one and the same scientific theory of emotions.[2] Thus, *emotion* is not a natural kind. Griffiths' argument has been criticized by a number of philosophers (Charland, 2002; Nussbaum, 2001; Prinz, 2004a, 2004b, 2004c), but their objections have, in my opinion, already been successfully rebutted by Griffiths himself (Griffiths, 2004).

In what follows, I want to develop a theme hinted at, but not fully developed, by Griffiths (1997, 2004), namely that what is true of emotion "is probably true of many specific emotion categories, such as anger and love" (Griffiths, 2004, p. 233). To be clear, the thesis that anger, love, fear, disgust, etcetera are not natural kinds does not follow from the thesis that emotion is not a natural kind. It may well be that, even though *emotion* is not sufficiently homogeneous to allow for extrapolation of discoveries about emotion to the whole emotion category, the traditional categories subordinate to it *are* sufficiently homogenous for scientific extrapolation. In this case, the heterogeneity of the *emotion* category would result from the fact that there are important differences *between* subordinate categories such as, say, anger and love—instances of anger are not sufficiently similar to instances of love to belong to one and the same scientific theory.

Barrett on Why Discrete Emotions Are Not Natural Kinds

My view is that subordinate categories such as anger, love, fear, disgust and so on are themselves not natural kinds. This thesis *does* entail that *emotion* is not a natural kind. If categories subordinate to emotions such as love and anger are each not sufficiently homogenous for scientific extrapolation, emotion will also not be sufficiently homogenous for scientific extrapolation. The heterogeneity of *emotion* in this case would result from the fact there are important differences both *within* and *between* subordinate categories—instances of anger are not sufficiently similar to other instances of anger or to instances of love to belong to one and the same scientific theory.

I begin from Barrett's (2006) attack of what she calls the *natural kind view* of discrete emotions such as anger, fear, shame, and so on. Barrett focused her critique primarily on basic emotion theory and appraisal theory. "Unlike the basic-emotion approach," Barrett (2006, p. 30) writes, "the appraisal approach does not assume … that objects or situations trigger prescribed emotional responses in an unmediated or reflexive way … [positing instead] that cognitive processes mediate emotion elicitation." Generally speaking, appraisal models give pride of place to the complex evaluation of stimuli that initiates the cascade of emotional responses (see Scherer, Schorr, & Johnstone, 2001).

What basic emotion theory and appraisal theory have in common, Barrett argued, is that they are both "natural kind models":

> Natural kind models of emotion not only assume that there are distinct profiles of responses to characterize each kind of emotion, but they also assume that these responses are caused by distinct emotion mechanisms. The causal mechanism for *anger* is presumed responsible for the coordinated package or correlated set of features that constitute an anger response. (Barrett, 2006, p. 31)

More precisely, "natural kind models" take traditional discrete emotion categories such as anger, fear, disgust, etcetera to be associated with an emotion-specific *response profile* and/or an emotion-specific *mechanism* causally responsible for the response profile. A *response profile*, in turn, may comprise subjective experience, facial and vocal signals, peripheral nervous system responses, voluntary behaviors, and neural circuitry. These are the most common components that have appeared in proposed definitions of discrete emotions over the history of affective science.

Barrett's (2006) key point is that there is no *one-to-one* correspondence between anger, fear, shame, etcetera and any response profile and/or emotion-specific mechanism. On the one hand, neither subjective experience, nor facial and vocal signals, nor peripheral nervous system responses, nor voluntary behaviors, nor neural circuitry appear to be distinctive of anger, or fear, or disgust and so on in the sense that they apply to *all and only* instances of what we call in English "anger," "fear," "shame," etcetera. Whatever signature may be proposed for, say, anger at the phenomenological, physiological, expressive, behavioral, or neural level, we seem to be able to find instances of anger without that signature.

Furthermore, if we collect *all* instances of what we call in English "anger," "fear," "shame," etcetera, we will not find among such instances any specific set of components that correlate to a high degree. If so, not only will there be no bodily signatures for anger, fear, shame, etcetera at the level of *any single component*, but neither will there be any signatures at the level of *any single set of correlating components*.

I have argued elsewhere that Barrett's (2006) empirical data on lack of componential specificity and on lack of high correlation among components are convincing with respect to traditional emotion categories such as anger, fear, shame, etcetera (Scarantino, 2009; Scarantino & Griffiths, 2011). Before considering some possible responses to such data, I discuss a central disagreement with Barrett's analysis of the failures of basic emotion theory and appraisal theory.

The Problem of Scope

According to Barrett, the trouble with basic emotion theory and appraisal theory lies in the assumption that "emotions have definable boundaries in the brain or body" (Barrett, 2006, p. 29). In Lindquist, Wager, Kober, Bliss-Moreau, & Barrett (in press, p. 5), it is explicitly argued that "[t]he natural kinds approach assumes that emotion categories like *anger, sadness, fear,* et cetera, map on to biological categories that are given by the brain and body."

The implication is that scientific models of anger, sadness, and fear that do not put emphasis on brain or body would not encounter the problems faced by basic emotion theory and appraisal theory. This implication naturally leads to social constructionist theories, proposed as an antidote to the "tendency to treat emotions as biologically primitive, instinctive response patterns" (Averill, 1980b, p. 57). Barrett's own conceptual act theory, the details of which I cannot consider in

this article, proposes that "emotion categories are … socially constructed" (as cited in Lindquist et al., in press, p. 5).

I will now argue that the problems encountered by basic emotion theorists and appraisal theorists have nothing to do with the fact that they map traditional emotion categories to biological categories, and all to do with the fact that they map traditional emotion categories to theoretically homogeneous categories for purposes of scientific extrapolation. This can be shown by demonstrating that theories which search for the distinctive social signature of traditional emotion categories have as much trouble finding it as theories which search for the distinctive biological signature of traditional emotion categories.

Consider Averill's social constructionist theory. It is worth considering not only because it is highly influential, but because it clearly aims for universal scope. In Averill's (1980a) view, "[a]n emotion is a transitory social role (a socially constituted syndrome) that includes an individual's appraisal of the situation, and is interpreted as a passion rather than as an action" (1980, p. 312). The interpretation of passivity, Averill suggested, results from a limited understanding of the social role played by emoting, which is "resolving conflicts which exist within the social system" (1980b, p. 37).

For example, Averill proposed that anger is a social role people take on in order to be justified in the exercise of aggression. This account seems to capture some forms of anger, for example the "wild pig" syndrome characteristic of the Gururumba tribe of New Guinea (Newman, 1964). The syndrome consists of a ritualized sequence of aggressive behaviors the locals interpret as resulting from being bitten by the spirit of someone who died.

As pointed out by Averill, this attribution manifests a limited insight into the social function played by the syndrome, namely giving the "victims" a means to communicate how overwhelmed they are by their social responsibilities. The syndrome generally occurs to newly married males who have just begun sharing communal responsibilities, and it leads the community of elders to reduce their share of social responsibilities.

What Averill's theory certainly does not do is to apply to *all* instances of what we call "anger" in English. Consider the sort of anger in which infants engage when they are suddenly restrained. It has been detected in infants as early as their 5th month of life in a variety of cultures (Camras, Campos, Oster, Miyake, & Bradshaw, 1992), it is associated with characteristic facial expressions, it is automatic, it has quick onset, brief duration, unbidden occurrence, it exists in homologous form across species, and so on.

In other words, it fits Ekman's (1999) definition of a basic emotion. In contrast, it does not fit Averill's (1980a, b) own definition of anger, because there is no theoretically interesting sense in which the infant is playing a social role, let alone the specific social role Averill ascribes to anger.[3] The case of infant anger due to restraint is one of many to which Averill's theory does not apply.

The inability to provide definitions that fit all instances of the target emotion is not specific to Averill's formulation of social constructionism, nor to Ekman's formulation of basic emotion theory. I have argued in Scarantino (2005) that all scientific theories which try to tell us what all emotions are like, or what all angers are like, end up facing the following:

> *Problem of scope*: For every scientific theory T that tells us that an emotion/anger/fear/etc. is X, we can find counterexamples consisting of things called "emotion"/"anger"/"fear"/etcetera in English that are not X, and/or things not called "emotion"/"anger"/"fear"/etcetera in English that are X.

The problem of scope is what frustrates models that endorse the natural kind assumption (NKA). To reiterate, these are not models that map traditional emotion categories to biological categories (Barrett's [2006] suggestion), but, more broadly, models that map traditional emotion categories to theoretically homogenous categories, namely natural kinds in the sense I discussed earlier. I conclude that social constructionist models are as guilty of the natural kind assumption as basic emotion and appraisal models.

The Trouble with Traditional Emotion Theory

Griffiths (1997, 2004) and Barrett (2006) have focused primarily on explaining why traditional emotion categories are not natural kinds. What they have not done is provide compelling evidence that the NKA is widespread. Why should we think that emotion scientists assume that traditional emotion categories map on to theoretically homogeneous categories? The relevant evidence comes from the following:

> *Core dialectic*: When someone alleges that there are empirical data about what we call in English "emotion"/"anger"/"fear"/etcetera that your favorite scientific definition cannot account for, reject the data as being nonpertinent or demonstrate that, when properly interpreted, your definition can account for such data. When someone proposes an alternative scientific definition, find empirical data about what we call in English "emotion"/"anger"/"fear"/etcetera that the rival definition, unlike your own, cannot account for.

The argumentative strategies at the heart of this core dialectic only make sense on the assumption that a single scientific definition, whether biologically or socially based, should apply to all instances of the traditional emotion categories. But this can only be the case if all emotions/angers/fears/etcetera belong to a single natural kind—respectively, the emotion kind, the anger kind, the fear kind, etcetera—whose condition of membership is captured by one's favorite definition, but not captured by any of its rivals.

Traditional emotion theorists are indeed convinced that their own definition of emotion/anger/fear/etcetera achieves two objectives at the same time: it individuates a theoretically homogeneous category for purposes of scientific extrapolation, and it accounts for all empirical data about what we call in English "emotion"/"anger"/"fear"/etcetera. If that were true, the problem of scope would not hold, and one scientific theory would be able to achieve universal scope.

In what follows, I discuss the two main strategies that have emerged to preserve the illusion of universal scope: the *ex cathedra* strategy and the hidden unity strategy. I call it an illusion because, as I will now argue, neither strategy succeeds. I present the two strategies with respect to basic emotion theory, but I emphasize that they are used by theorists who "biologize" emotions just as much as they are used by theorists who "socialize" emotions.

Ex Cathedra Strategy

One possible response to the charge that one's favorite scientific definition of emotion/anger/fear/etcetera does not apply to all instances of the relevant traditional emotion categories is to stipulate that the contradictory data are not "really" about emotions/angers/fears/etcetera, but about some other affective phenomenon. I call this the *ex cathedra strategy*, because it consists of stipulating from a position of authority that what does not fit the proposed definition not only should not count as an emotion for the purposes of the theory (this would be a legitimate move), but is not an emotion *simpliciter*.

A conspicuous example of the *ex cathedra* strategy can be found in Ekman's edict that "I do not allow for 'non-basic' emotions" (Ekman, 1999, p. 57). If we accept this edict, we can reject all empirical data in contradiction with basic emotion theory by stipulating that they do not concern emotions. Consider Barrett's (2006) data calling into question the assumption that there are facial expressions distinctive of all instances of what we call "anger" in English, such as furrowing of the brow and tightening of the lips. If distinctive facial expressions are an essential part of the definition of a basic emotion, as Ekman has often suggested, and if basic emotions are the only emotions that there are, anything that does not have the requisite facial expressions cannot be a genuine instance of the emotion of anger. Therefore, it cannot be a counterexample to a scientific theory of anger. We may therefore dismiss the data on alleged instances of anger that do not involve furrowing of the brow and tightening of the lips as being about affective phenomena other than anger.

This is the strategy used by Ekman to account for traditional emotion categories of which no members fit the profile of a basic emotion, for instance grief and romantic love (in the case of anger, some members do fit such profile). According to Ekman, grief and romantic love are not emotions, but *emotional plots*. Emotional plots are "more specific" and "more enduring" than emotions (Ekman, 1999). The emotional plot of grief, for instance, lasts longer than the emotion of sadness, and it is "much more specific than sadness. We know that in grief a death has occurred, in sadness we only know that the person has suffered an important loss, but not what kind of loss" (Ekman, 1992, p. 194).

This characterization of emotional plots strikes me as unable to capture any relevant distinction between grief and sadness, but this is not the key issue. The key issue is that the *ex cathedra* strategy is a bad way to make a good point, namely that there are important theoretical differences between members of the same traditional emotion category such as *emotion* or *anger*. If we assume that a good theory of emotion/anger/fear/etcetera must apply to all emotions/all angers/all fears/etcetera, we are forced to respond to seeming members of such categories that do not fit one's definition by expelling them *ex cathedra* from membership.

The reason why this is a bad way to acknowledge the presence of theoretical differences is that it insulates the theory from the possibility of falsification. Empirical data contradicting one's theory of emotion/anger/fear/etcetera should not be rejected because they do not fit the theory, lest the theory becomes true by stipulation. As we shall see, a better way to acknowledge theoretical differences is to give up on the assumption that a good theory of emotion/anger/fear/etcetera should apply to all emotions/all angers/all fears/etcetera.

Hidden Unity Strategy

A second possible response to the charge that one's favorite scientific definition of emotion/anger/fear/shame/etcetera does not apply to all instances of the relevant traditional emotion categories is to acknowledge the presence of differences, but to argue that, over and above such differences, emotion/anger/fear/shame/etcetera are still theoretically homogeneous categories for purpose of scientific extrapolation.

I call this the *hidden unity strategy*, because it consists of finding some nonobvious source of unity among instances of the same traditional emotion category, while acknowledging that there are important theoretical differences among them. Here, I consider one example of this strategy at work, namely Izard's (2007) defense of basic emotion theory from Barrett's (2006) attack.

Izard proposed that there are two main kinds of emotions: *basic emotions* and *emotion schemas*. Basic emotions are "affective processes generated by evolutionarily old brain systems upon the sensing of an ecologically valid stimulus" (Izard, 2009, p. 7). Emotion schemas, on the other hand, consist of "emotion interacting dynamically with perceptual and cognitive processes to influence mind and behavior" (Izard, 2009, p. 8).

A key difference with Ekman's account is that emotion schemas, unlike emotional plots, are acknowledged to be emotions. In other words, Izard allows for nonbasic emotions. His view is that "[t]he cumulative evidence suggests that the following basic emotions meet criteria for classification as natural kinds: interest, joy/happiness, sadness, anger, disgust, and fear" (Izard, 2007, p. 261). At the same time, he states that "emotion schemas are not natural kinds because they have properties that differ across individuals and cultures" (Izard, 2007, p. 261).

This last passage makes clear that the notion of natural kind Izard presupposes differs from the one I relied on in this article. In Izard's view, natural kinds are categories whose instances do not have properties that differ across individuals and cultures (call this the *social invariance* sense of "natural kinds"). In my view, natural kinds are theoretically homogenous categories for purposes of scientific extrapolation (call this the *inductive* sense of "natural kinds").

These two notions of natural kinds are orthogonal to one another (Scarantino, in press). For instance, a category may be

nonnatural in the social invariance sense because its members have properties that differ across cultures, but natural in the inductive sense because its members participate in the same body of empirically discovered reliable generalizations (e.g., generalizations about cultural variation). Since this article focuses on whether traditional emotion categories are theoretically homogeneous, I will disregard the social invariance sense of natural kinds in what follows.

Relative to the inductive sense of natural kinds, both basic emotions and emotion schemas *are* natural kinds by Izard's own lights. This is because they both map on to theoretically homogeneous categories for purposes of scientific extrapolation. Izard himself proposes a number of empirically-based generalizations that are meant to apply to, respectively, basic emotions and emotion schemas.

The problem is that some of the instances of Izard's basic emotions—interest, joy/happiness, sadness, anger, disgust, and fear—satisfy the generalizations characteristic of emotion schemas. Izard argues that there is evidence for the "universality of expressions of a limited set of basic emotions," for "their recognition via processing in evolutionarily old brain stem and amygdala systems," and for the fact that basic emotions do "not depend on or include complex appraisals or higher order cognition" (Izard, 2007, p. 262).

This would not do as a response to Barrett (2006). There are instances of interest, joy/happiness, sadness, anger, disgust, and fear—the alleged basic emotions—that lack distinctive expressions, are not recognized via amygdala systems, and depend on higher order cognition (Barrett, 2006; Lindquist et al., in press). These instances fully qualify as emotion schemas.

Izard admits that when we speak of, say, "anger" in English we can refer either to a basic instance—for example, "anger to explain the frustration of blocked goal responses" (2007, p. 267)—or to a schema instance, for example, "a pattern of … anger schemas to explain reactions to the terrorist disaster of 9/11/2001" (2007, p. 267). This would seem to force Izard into accepting the conclusion that traditional emotion categories are not theoretically homogeneous categories because they contain emotions of different types.

It is at this juncture that Izard engages in the *hidden unity strategy*. His proposal to preserve unity is to say that basic emotions and emotion schemas, different though they may be, have *evolved feelings* at their motivational core. On this view, the trademark thesis that all emotions are biological adaptations can be upheld, because "all emotion feelings, whether they are a component of a basic or nonbasic emotion, are products of evolution" (Izard, 2007, p. 265). "Although there are important differences at the cognitive and action levels between an anger schema and a basic anger episode," Izard concludes, "the quality of the anger feelings is the same" (2007, p. 265).

But why should we believe that? The quality of the feeling associated with a given anger episode is determined by what it is like to experience such an episode. This in turn depends on which specific emotion components are involved in it. Since basic anger and schema anger are instantiated by different components (e.g., basic anger involves facial changes, schema anger doesn't; basic anger involves automatic appraisal, schema anger doesn't; etc.), there is no reason why the quality of the anger feeling should be the same, except in the trivial sense that it will be a feeling of anger in both cases.

But let us assume for the sake of argument that emotion feelings have the exact same experiential quality in all cases in which a discrete emotion exists in both basic and schema forms. Still, there are emotions—for example, *shadenfreude* (joy at someone else's misfortune)—which only seem to exist in the form of emotion schemas. How would evolved feelings manage to get to their motivational core? One could argue that all emotion schemas lacking a basic version come from a combination of different basic emotions, the way cocktails result from a combination of spirits (Prinz, 2004b). If that were true, a combination of evolved feelings may still be at the motivational core of *shadenfreude*. This combinatorial project, however, faces a number of serious difficulties, and it is far from clear that they can be overcome (Scarantino & Griffiths, 2011).

I will mention a final problem, namely that there is at least preliminary evidence that some emotions do not involve feelings at all because they are unconscious (Berridge & Winkielman, 2003). These results are admittedly controversial, but they put further pressure on the thesis that all emotions have an evolved feeling at their core.

I conclude that even Izard's (2007) sophisticated version of basic emotions theory faces the problem of scope: There are things we call "emotion" or "anger" in English which are not at their core biological adaptations, not even in the weak sense that they have evolved feelings at their motivational core.

One case study is clearly not enough to establish that the hidden unity strategy does not work. In Scarantino (2005), I have considered many more attempts to find hidden unity in traditional emotion categories, and argued that they all fail in one of two ways. They either do not detect a genuine way in which all emotions/angers/fears are alike (Izard's case), or they detect a genuine way in which they are alike but one that fails to be relevant to the predictive and explanatory purposes of affective science.

My interpretation of this collective failure is that there is no unity to be found within traditional emotion categories. Although we cannot be positively sure that traditional emotion categories are not theoretically homogeneous for the purpose of scientific extrapolation, we have good reason to interpret the long history of failure in finding a single scientific theory that fits all cases as a sign that no single scientific theory can fit all cases. In other words, we have good reason to infer from the failures of traditional emotion theory that emotion, anger, fear and so on are not natural kinds (Scarantino, 2005).

Changing the Paradigm of Affective Science

Primitivism versus Pluralism

If the classificatory scheme of traditional emotion theory is no longer suitable to the needs of affective science, what should replace it? Two main proposals for conceptual renewal should be considered:

Primitivism: Infer from the failures of traditional emotion theory that there are no natural kinds of emotion/anger/fear/etcetera, and search for natural kinds at the level of primitive components of discrete emotions.

Pluralism: Do not infer from the failures of traditional emotion theory that there are no natural kinds of emotion/anger/fear/etcetera, and search for natural kinds at the level of discrete, but nontraditional, emotion categories.

Primitivism is the position advocated by psychological constructionists, and pluralism is the position I wish to defend. According to primitivism, there are no natural kinds of emotion/anger/fear/etcetera and the search for natural kinds should focus exclusively on emotion components that are *psychological primitives* "elemental—but still psychological—building blocks" (Russell, 2003, p. 146). Psychological constructionists are convinced that "[t]he goal of psychology is to identify … the most basic psychological descriptions that cannot be further reduced to anything else mental (because at that point they would describe biological mechanisms)" (Lindquist et al., in press, p. 4).

A prime candidate for psychological primitivity is core affect, a blend of hedonic and arousal values (Russell, 2003). Core affect need not be directed at anything, but it is a building block of psychological events such as discrete emotions, which are instead directed at something (at least generally). For example, a fear episode caused by a bear in the woods will include a feeling of high arousal and high displeasure directed at the bear. Barrett states that "the empirical case supporting the hypothesis that core affect is a natural kind is suggestive" (2006, p. 48), even though she admits that further empirical investigation is needed to close the case.

I support the search for natural kinds at the level of emotion components. What I reject is the inference that, since emotion, anger, fear and so on are not natural kinds, there must not be any natural kinds of emotion/anger/fear/etcetera. We can appreciate that this is a bad inference by considering the case of memory. What we call "memory" in English is not a theoretically homogeneous category for purposes of scientific extrapolation. Yet it clearly does not follow that there are no natural kinds of memory.

Long-term memory and short-term memory, just to pick two examples, are promising candidates for natural-kind status. They each participate in a large body of empirically discovered reliable generalizations on account of the many scientifically interesting properties they have in common. For example, lots of reliable generalizations have been discovered about *short-term memory*, concerning its dissociation from long-term memory and other forms of memory, its neural and chemical underpinnings, its maximum capacity, its information storage modality, and so on.

By the same token, the fact that the traditional categories of emotion/anger/etcetera are not natural kinds is perfectly compatible with the existence of natural kinds of emotion/anger/etcetera. As I argue below, these will have to be categories that bear some similarity with traditional emotion categories, but that differ from them in that they constitute homogenous groupings of emotion instances that participate in a large body of empirically discovered reliable generalizations.

Another reason why psychological constructionists infer that there are no natural kinds of emotion/anger/etcetera may be the implicit conviction that natural kinds exist only at an elemental level of analysis. If so, we won't be able to individuate natural kinds of emotions, not because traditional emotion categories are not natural kinds, but because all emotions have more primitive building blocks (e.g., core affect).

I see no reason why there should not be natural kinds at both lower and higher levels of analysis. An analogy may help. Water is a natural kind for chemistry. Yet water is a chemical compound constituted by hydrogen and oxygen molecules, so it has other chemical substances as components. Furthermore, the atoms of the basic chemical elements are themselves composed of more basic parts, namely protons, neutrons, and electrons. So neither *chemical compounds* like water nor *chemical elements* like oxygen and hydrogen are without components, but they both fully qualify as natural kinds for chemistry.

The take-home message here is that not having constitutive components is not a necessary condition for being a natural kind. Consequently, although I consider the project of psychological constructionists well worth pursuing, I reject the idea that only psychological primitives can be natural kinds, and conclude that affective science should engage in an open search for natural kinds, at both elemental and nonelemental levels of analysis.

Folk Emotion Project versus Scientific Emotion Project

It is time to provide some details on how the pluralistic search for natural kinds at the level of nontraditional discrete emotion categories should be carried out. The first order of business is to distinguish between two projects that have been systematically confused in the history of emotion science.

One is the *Folk Emotion Project*, which aims to offer a *descriptive definition* of the conditions of membership of traditional emotion categories such as emotion, anger, and so on. The other is the *Scientific Emotion Project*, which aims to offer a *prescriptive definition* of the conditions of membership of natural kinds of emotion, natural kinds of anger, and so on.[4] Whereas the Folk Emotion Project has the accurate reconstruction of the boundaries of traditional emotion categories as its primary objective, the Scientific Emotion Project has the transformation of such categories into useful scientific tools as its primary objective.

Emotion theorists have not distinguished between these two projects because they have implicitly assumed that one and the same scientific definition can simultaneously individuate a natural kind of emotion/anger/fear/etcetera and account for all empirical data about what we call in English "emotion," "anger," "fear," etcetera. To be pulled out, this achievement would require that traditional emotion categories designate natural kinds. But I have argued that we have good reason to conclude that they do not.

If so, two important consequences follow. First, no single scientific definition can be both descriptively and prescriptively

adequate. Second, affective scientists must choose whether to engage in the Scientific Emotion Project—formulating good prescriptive definitions—or whether to engage in the Folk Emotion Project, formulating good descriptive definitions. If a definition aims to be descriptively adequate, we should ask: Does it account for all empirical data about what we call in English "emotion," "anger," etcetera? If the definition aims to be prescriptively adequate, we should ask: Does it individuate a natural kind of emotion, or a natural kind of anger? What we should stop asking of any scientific definition is the achievement of both objectives at the same time.

I won't say much about the Folk Emotion Project, except to note that Fehr and Russell's (1984) prototype proposal goes in the right direction. Since traditional emotion categories are unlikely to belong to the restricted class of classically definable categories, family resemblance accounts in the style of Fehr and Russell's strike me as promising candidates to account for all empirical data about what we call in English "emotion"/"anger"/"fear"/ etcetera.

But, as pointed out by Widen and Russell (2010, p. 378), "a [good] descriptive definition does not provide a good prescriptive definition." The reason should now be clear: If traditional emotion categories are not natural kinds, what tells us what makes something a member of a traditional emotion category does not tell us what makes something a member of a natural kind.

What I want to focus on in the rest of my article is the Scientific Emotion Project, namely the project of discovering natural kinds of emotions.

Essentialism versus Antiessentialism

So far I have worked with a thin notion of natural kinds, understood simply as categories that are homogeneous for purposes of scientific extrapolation. This general idea has been developed in two main ways in philosophy. According to *essentialist accounts*, the members of a natural kind share an essence—a set of individually necessary and jointly sufficient properties. For example, all samples of water have the same chemical constitution (H_2O), and this is what makes them samples of water. In this view, natural kinds have sharp edges and feature in exceptionless laws of nature that hold uniformly across time and space.

This essentialist notion applies to natural kinds in chemistry and physics, but it is an unsuitable theory of natural kinds for the biological and social sciences (Boyd, 1999; Machery, 2005; Samuels, 2009; Wilson, Barker, & Brigandt, 2007). In such disciplines, variability among kind members is the norm, borderline cases often emerge, and generalizations tend to be exception-ridden and only locally valid.

Biological species such as *tiger* are a case in point. There is a great deal of variation among species members. Furthermore, species are products of evolution and as such generalizations about them cannot be unrestrictedly valid across space and time. But despite this, classification of organisms into species is of proven utility in biology, and reflects a real division in nature, rather than being solely an expression of human interests.

To accommodate variation, borderline cases, and the lack of exceptionless and universal generalizations in the special sciences, *antiessentialist accounts* of natural kinds have been proposed. The most influential is Boyd's *homeostatic property cluster* (HPC) theory (Boyd, 1991), according to which "[t]he natural definition of … *homeostatic property clusters 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" (Boyd, 1991, p. 141).

In the HPC view, *tiger* would be defined by a cluster of genetic, morphological, physiological, and behavioral properties that species members tend to reliably share on account of causal mechanisms such as interbreeding, being exposed to common selection pressures, and sharing ancestors. This definition is compatible with the absence of individually necessary and jointly sufficient properties for *tigerhood*, and with the presence of individuals that have just enough of the cluster properties to place them on the borderline for membership.

A final requirement for being a natural kind, on both the essentialist and the antiessentialist approach, is being a *maximal class* of items that share an essence or a cluster of causally correlated properties. For instance, consider the proposal that *tiger living in the Western hemisphere* is a natural kind. The reason why we would reject the proposal is not that tigers living in the Western hemisphere do not share a cluster of properties on account of causal mechanisms, but that there is a larger class—the class *tiger*—whose members tend to share the exact same cluster of causally correlated properties. The problem with *tiger living in the Western hemisphere* is that it is a subset of a larger class of items that share the same cluster of causally sustained properties.

In what follows, I presuppose Boyd's HPC account of natural kinds. I have argued elsewhere that affective scientists should explicitly embrace an antiessentialist theory of natural kinds (Scarantino, in press; Scarantino & Griffiths, 2011). This would allow them to acknowledge the presence of variability among members of affective kinds, of borderline cases, and of exceptions to generalizations, while preserving the idea that some kinds are maximally suitable for induction and explanation in affective science.

What Prescriptive Definitions of Emotions Should Achieve

One of the central activities of a scientist is the transformation of ordinary categories into scientifically suitable categories. Examples include the transformation of *probability* into *relative frequency*, the transformation of *memory* into *short-term memory*, and the transformation of *ability to think* into *ability to pass the Turing test*. These scientifically motivated transformations are governed by two main objectives: increasing precision and testability, and individuating a category more suitable for induction and explanation than the original one.

Affective scientists interested in the Scientific Emotion Project should engage in this sort of transformative activity with

respect to traditional emotion categories such as emotion, anger, fear, and so on, and formulate prescriptive definitions for them. But what would make a prescriptive definition adequate? My proposal goes as follows:

> A good prescriptive definition of emotion/anger/fear/etc. should specify the condition of membership of a natural kind of emotion/anger/fear/etc., namely a transformed category provisionally called "K" such that (a) K's members are the maximal class of items that tend to reliably share inductively and explanatorily important properties on account of one or more causal mechanisms (naturalness condition), (b) most or all of K's members are members of the traditional emotion categories of emotion/anger/fear/etcetera (similarity condition).

A good prescriptive definition, unlike a good descriptive definition, does not aim to tell us what makes something a member of a traditional emotion category. Rather, it aims to tell us what makes something a member of a transformed category K that achieves two objectives: It is a natural kind and it preserves some similarity to the nonnatural kind it transforms.

The Scientific Emotion Project should be pluralistic. Insofar as the members of a given traditional category E belong to multiple natural kinds K, K', K'', there will be a plurality of good prescriptive definitions of the same E. Gone is the assumption that one and only one scientific definition of emotion, or anger, or fear can be scientifically adequate, and that the objective of affective science is to reach consensus around such a definition. As we allow for memory to be split into a plurality of natural kinds of memory (e.g., short-term memory and long-term memory), we should allow for emotion, anger, fear, etcetera to be split into a plurality of natural kinds of emotion, natural kinds of anger, natural kinds of fear, etcetera.

The naturalness condition spells out what it is to be a natural kind of emotion in light of Boyd's theory of HPC kinds. According to this, a natural kind K is a maximal class of items that tend to reliably share a cluster of properties on account of one or more causal mechanisms (e.g., neural mechanisms), where the co-occurrence of these properties is important for purposes of prediction and explanation in affective science. No single property needs to be shared by all kind members, and some individuals will be on the borderline for membership. The key requirement for an affective HPC kind will be that the properties kind members tend to exhibit do not correlate accidentally, but by virtue of causal mechanisms that preserve their imperfect co-occurrence. This is what grounds affective natural kinds in the causal structure of the world.

The similarity condition spells out what it takes for a new category K to count as a natural kind *of* emotion/anger/fear/etcetera, as opposed to a natural kind having nothing to do with emotion/anger/fear/etcetera. The latter case is a distinct possibility. For instance, if *no* member of a natural kind K were a member of the traditional category of emotion/anger/fear/etcetera, it would make no sense to count K as a natural kind of emotion/anger/fear/etcetera. This would not detract in any way from the scientific importance of K. It would only detract from its ability to qualify as a natural kind of emotion/anger/fear/etcetera.

My suggestion is that if *all* or *most* members of a natural kind K are members of the traditional category of emotion/anger/fear/etcetera, it makes sense to count K as a natural kind of emotion/anger/fear/etcetera. Now, if *all* K's members are members of a traditional category E, K is a subset of such category and so it straightforwardly qualifies as a kind of E. If most of K's members are members of a traditional category E, on the other hand, K is not a subset of E, but an overwhelming portion of K overlaps with E. In this case as well, I suggest that we should think of K as a kind of E.

Let us illustrate how the two conditions are supposed to work in the case of anger. My proposal is that a good prescriptive definition of anger is one that individuates a natural kind of anger, understood as a kind K such that its members are the maximal class of items that tend to reliably share a cluster of causally sustained and inductively and explanatorily important properties, and such that all or most of K's members are called "anger" in English.

This is compatible with there being other good prescriptive definitions of anger that individuate different natural kinds K', K'', K''' such that all or most of their members are called "anger" in English. It is also compatible with the fact that *most* members of the category called "anger" in English are not members of K. In other words, whereas we should require that all or most members of a natural kind K of anger are members of *anger*, we should not require that all or most members of *anger* are members of K.[5]

A difficult question I cannot settle in this article concerns the labeling of natural kinds of emotions: If K is a natural kind of anger, how should we call it? Two options present themselves, each with advantages and disadvantages. One option is to use neologisms. For instance, a natural kind of anger may be referred to as "WS34." The advantage of this nomenclature is that it would minimize the risk of mistaking K for the traditional emotion category of *anger*. The disadvantage is that a neologism would fail to signal that there is an important similarity between *anger* and K, namely that all or most of K's members are called "anger" in English.

The other option would be to modify the traditional emotion category being transformed by capitalizing it (e.g., ANGER), adding subscripts and superscripts ($anger_1$, anger*), or introducing qualifiers (basic anger, schema anger). The advantage of a modifier is that it would make clear that there is a relation of similarity and difference within the anger category: similarity because all or most of K's members are called "anger" in English; and difference because K, unlike anger, is a natural kind. The shortcoming would be that a modifier may still wrongly suggest that what is being defined is anger *simpliciter*. I leave the discussion of which of these two options—neologisms or modifiers—is to be preferred to another article.

A New Dialectic for Affective Science

Which of the many available scientific definitions of emotion, or anger, or fear is prescriptively good? The one offered by

basic emotion theory? The one offered by appraisal theory? The one offered by social constructionism? All of them? None of them? Answering these substantive questions is a challenging task better left to other articles. What I want to do in conclusion is to explain how the dialectic of the Scientific Emotion Project ought to differ from the core dialectic of traditional emotion theory.

According to the latter, the job of an emotion scientist is to either find empirical data about what we call "emotion," "anger," "fear," and so on that a rival theory cannot accommodate, or to accommodate empirical data presented by a rival theory by either stipulating that such data are not pertinent (e.g., *ex cathedra* strategy) or by showing that one's own definition accommodates them (e.g., hidden unity strategy).

This dialectic has no place in the context of the Scientific Emotion Project. This is because neither proponents nor opponents of a given prescriptive definition would expect it to simultaneously account for all empirical data about the psychological events referred to by the English terms "emotion," "anger," "fear," etcetera and to individuate natural kinds. Debates would rather focus on whether a proposed prescriptive definition satisfies the naturalness condition and the similarity condition.

Consider by way of illustration Ekman's (1999) account of basic emotions such as anger, fear, disgust, and so on. As we have seen, this account has been criticized because there is no bodily signature for what we call in English "anger," "fear," "disgust," etcetera at the level of any single component (e.g., expressions, physiology, phenomenology, behavior, or neural circuitry), or at the level of any single cluster of correlating components (Barrett, 2006).

Basic emotion theorists have responded to this charge by either stipulating that what does not fit the definition of a basic emotion is not really an emotion (e.g, Ekman's *ex cathedra* strategy), or by trying to demonstrate that there is a hidden evolutionary signature shared by all instances of traditional emotion categories (e.g., Izard's hidden unity strategy). I have argued that both strategies fail.

A more powerful reply is now available to basic emotion theorists. They could simply point out that their definitions aim for prescriptive adequacy, not for descriptive adequacy. If so, it becomes irrelevant that there is no one-to-one correspondence between, say, *anger* and any single emotional component or any single cluster of correlating components. This would only demonstrate that *anger* is not a natural kind.

A prescriptive definition does not aim to define *anger in general*, but a *natural kind of anger*. On this interpretation, basic emotion theory is perfectly compatible with the presence of a one-to-many correspondence between anger and distinctive sets of correlated properties and/or causal mechanisms. The sets of correlated properties and/or causal mechanisms characteristic of *basic anger* would be distinctive of some, but not all, instances of anger.

This is not to say that the definition proposed by basic emotion theorists is prescriptively adequate. We could still reject it if we realized that it does not individuate a maximal class of items that tend to share inductively and explanatorily important properties on account of one or more causal mechanisms, or if we realized that it is not sufficiently similar to what we call "anger" in English to qualify as a natural kind of anger.

What would be removed is the threat constituted by a handful of empirical data about what we call "anger" in English that the definition cannot account for. This limitation will be shared by every good prescriptive definition of anger, and it is nothing to worry about. Empirical data about what we call "anger" in English can still be grounds for dismissing a prescriptive definition of anger, but only if they call into question that the similarity condition has been fulfilled.

Conclusion

The first goal of this article has been to argue that there is something fundamentally wrong with the way discrete emotions have been studied so far. What is wrong is the natural kind assumption—the assumption that traditional emotion categories map onto theoretically homogenous categories for purposes of scientific extrapolation. I have provided evidence that this assumption is widespread and I have argued that it is false.

The second goal of this article has been to move beyond the natural kind assumption. One option is primitivism, advocated by psychological constructionists. In their view, natural kinds must be searched among building blocks of emotion episodes such as core affect, a blend of hedonic and arousal values. The option I have defended instead is pluralism. According to this, traditional emotion categories should be revised in the search for a multiplicity of fundamental, yet still discrete, natural kinds.

This is where the difference between my position and the position of psychological constructionists lies. Whereas we both agree that there cannot be a good scientific theory of emotion *as such* or anger *as such*, I have suggested that there can be multiple good scientific theories of *natural kinds of emotion* or *natural kinds of anger*. Each of them will account for a portion of the empirical data, but no model will account for all of the empirical data on, respectively, what we call "emotion" or "anger" in English.

Notes

1 By "scientific extrapolation" I do not mean that what is true of a sample must be true of the whole, but simply that there is a significant likelihood of that being the case.

2 Griffiths' argument does not require accepting the scientific viability of the notions of higher cognitive emotions or socially sustained pretenses, both of which I find problematic (Griffiths [2004] himself takes a distance from them). It only requires accepting the scientific viability of the notion of a basic emotion (for a defense of it, see Scarantino and Griffiths [2011]) and the additional thesis that there are emotions fundamentally different from basic emotions. From this it follows that *emotion* is not a single kind of thing from a scientific point of view.

3 In Averill's full picture, by being "overcome" by anger, people manage to get away with violating social norms against aggression for the sake of social norms that entitle them to the protection of their rights (Averill, 1980a, b). This conflict of norms does not play any role in explaining the infant's anger at being restrained.

4 This distinction between descriptive and prescriptive definitions was originally introduced by Russell (1991). In Widen and Russell (2010,

p. 377), we read that a prescriptive definition "is used to pick out the set of events that a scientific theory of emotion [or fear or of anger, etc.] purports to explain," whereas a descriptive definition is "a definition of the word emotion [or the word fear or the word anger, etc.] as it is used in everyday life." Although I use this distinction for different purposes, what I wish to preserve of it is the idea that whereas descriptive definitions aim to capture conditions of memberships of traditional categories as they are, prescriptive definitions aim to determine what the conditions of membership of suitably transformed emotion categories ought to be.

5 Consider the case of short-term memory. Most members of the category we call "memory" in English are *not* members of the short-term memory category, because the ordinary "memory" category contains a great many instances of other types of memory. On the other hand, *all* members of the short-term memory category are members of the "memory" category, which substantiates labeling it as a natural kind *of* memory.

References

Averill, J. R. (1980a). A constructivist view of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion*: *Theory, research, and experience* (pp. 305–339). New York, NY: Academic Press.

Averill, J. R. (1980b). Emotion and anxiety: Sociocultural, biological and psychological determinants. In A. O. Rorty (Ed.), *Explaining emotions* (pp. 37–72). Berkeley, CA: University of California Press.

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

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

Berridge, K.C., & Winkielman, P. (2003). What is an unconscious emotion? (The case for unconscious "liking"). *Cognition & Emotion*, *17*, 181–211.

Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies*, *61*, 127–148.

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

Boyd, R. (2007). Homeostasis, species, and higher taxa. In M. Lange (Ed.), *Philosophy of Science: An Anthology* (pp. 406–426). Oxford, UK: Blackwell Publishing.

Camras, L. A., Campos, J. J., Oster, H., Miyake, K., & Bradshaw, D. (1992). Japanese and American infants' responses to arm restraint. *Developmental Psychology*, *28*, 578–583.

Charland, L. (2002). The natural kind status of emotion. *The British Journal for the Philosophy of Science*, *53*, 511–537.

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

Duffy, E. (1941). An explanation of "emotional" phenomena without the use of the concept "emotion." *Journal of General Psychology*, *25*, 383–293.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*, 169–200.

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

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

Griffiths, P. E. (1997). *What emotions really are: The problem of psychological categories*. Chicago, IL: University of Chicago Press.

Griffiths, P. E. (2004). Is emotion a natural kind? In R. C. Solomon (Ed.), *Thinking about feeling: Contemporary philosophers on emotions* (pp. 233–249). Oxford, UK: Oxford University Press.

Hacking, I. (1991). A tradition of natural kinds. *Philosophical Studies*, *61*, 109–126.

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science*, *2*, 260–275.

Izard, C. E. (2009). Emotion theory and research: Highlights, unanswered questions, and emerging issues. *Annual Review of Psychology*, *60*, 1–25.

Kagan, J. (2007). *What is emotion?* New Haven, CT: Yale University Press.

Kagan, J. (2010). Once more into the breach. *Emotion Review*, *2*, 91–99.

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

Machery, E. (2005). Concepts are not a natural kind. *Philosophy of Science*, *72*, 444–467.

Mandler, G. (1975). *Mind and emotion.* New York, NY: Wiley.

Newman, P. L. (1964). "Wild man" behaviour in a New Guinea Highlands community. *American Anthropologist*, *66*, 1–19.

Nussbaum, M. C. (2001). *Upheavals of thought: The intelligence of emotions*. Cambridge, UK: Cambridge University Press.

Panksepp, J. (2007). Neurologizing the psychology of affects. *Perspectives on Psychological Science*, *2*, 281–296.

Prinz, J. (2004a). Embodied emotions. In R. C. Solomon (Ed.), *Thinking about feeling: Contemporary philosophers on emotions* (pp. 89–105). Oxford, UK: Oxford University Press.

Prinz, J. (2004b). *Gut reactions: A perceptual theory of emotion*. Oxford, UK: Oxford University Press.

Prinz, J. (2004c). Which emotions are basic? In P. Cruise & D. Evans (Eds.), *Emotion, evolution and rationality* (pp. 69–87). Oxford, UK: Oxford University Press.

Rorty, A. (Ed.). (1980). *Explaining emotions*. Berkeley, CA: University of California Press.

Russell, J. A. (1991). Natural language concepts of emotion. In D. J. Ozer, J. M. Healy, Jr. & A. J. Stewart (Eds.), *Perspectives in personality: Self and emotion* (pp. 119–137). London, UK: Jessica Kingsley.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*, 145–172.

Samuels, R. (2009). Delusion as a natural kind. In M. Broome & L. Bortolotti (Eds.), *Psychiatry as cognitive neuroscience: Philosophical perspectives* (pp. 49–82). Oxford, UK: Oxford University Press.

Scarantino, A. (2005). *Explicating emotions*. Unpublished doctoral dissertation, University of Pittsburgh.

Scarantino, A. (2009). Core affect and natural affective kinds. *Philosophy of Science*, *76*, 940–957.

Scarantino, A. (in press). *Discrete emotions: From folk psychology to causal mechanisms*. In P. Zachar & R. Ellis (Eds.), *Categorical versus dimensional models of affect: A seminar of the theories of Panksepp and Russell* (Consciousness and Emotion book series). Amsterdam, The Netherlands: John Benjamins.

Scarantino, A., & Griffiths, P. (2011). Don't give up on basic emotions. *Emotion Review*, *3*, 444–454.

Scherer, A., Schorr, A., & Johnstone, T. (Eds.). (2001). *Appraisal processes in emotion: Theory, methods, research*. New York, NY: Oxford 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, NY: Guilford Press.

Widen, S., & Russell, J. (2010). Descriptive and prescriptive definitions of emotion. *Emotion Review*, *2*, 377–378.

Wilson, R., Barker, M., & Brigandt, I. (2007). When traditional essentialism fails: Biological natural kinds. *Philosophical Topics*, *35*, 189–215.

Zachar, P. (2006). The classification of emotion and scientific realism. *Journal of Theoretical and Philosophical Psychology*, *26*, 120–138.