

# The Natural Kind Status of Emotion

Louis C. Charland

---

## ABSTRACT

It has been argued recently that some basic emotions should be considered natural kinds. This is different from the question whether as a class emotions form a natural kind; that is, whether emotion is a natural kind. The consensus on that issue appears to be negative. I argue that this pessimism is unwarranted and that there are in fact good reasons for entertaining the hypothesis that emotion is a natural kind. I interpret this to mean that there exists a distinct natural class of organisms whose behavior and development are governed by emotion. These are emoters. Two arguments for the natural kind status of emotion are considered. Both converge on the existence of emotion as a distinct natural domain governed by its own laws and regularities. There are then some reasons for being optimistic about the prospects for consilience in emotion theory.

- 1 *The mantra*
  - 2 *Griffiths on emotions as natural kinds*
  - 3 *Panksepp on emotions as natural kinds*
  - 4 *Emotion as a neurobiological kind*
  - 5 *Emotion as a psychological kind*
  - 6 *Response to the mantra*
  - 7 *Unification or fragmentation?*
  - 8 *Concluding remarks*
- 

## 1 The mantra

Is there a unified natural scientific domain that corresponds to the term ‘emotion’? As is frequently asked, is emotion a natural kind? Many philosophers who write about emotion seem to think not (Ben-Ze’ev [2000], de Sousa [1987], Elster [1999], Griffiths [1997], Neu [2000], Rorty [1978], Solomon [1995]). Ronald de Sousa, at least, appears uneasy about his place on this list. He notes that ‘many philosophers have repeated, like a mantra, the claim that emotions do not form a natural kind’ (de Sousa [1999], p. 910). He then remarks—with some irony—that they go on ‘to lump them

together anyway' (*ibid.*). So maybe there is something anomalous about this mantra and mass denial. The situation is certainly anomalous in other respects. Typically, no specific citations or arguments are given for the thesis. Apparently, this is something everyone is willing to deny but no one is willing to defend. In addition, the meaning of the thesis is usually taken for granted. This is surprising given ongoing controversies over the 'nature' of natural kinds (Boyd [1991], [1999]; de Sousa [1989]; Griffiths [1997], [1999]; Hacking [1991]; Millikan [1999]; Wilson [1999]). On the whole, the time seems ripe for a clarification and defence of the thesis that emotion is a natural kind.

There is much at stake in the question of whether emotion is a natural kind, and de Sousa is right to be uneasy. At the most basic level, there is the status of emotion as a field of inquiry. Should the diverse phenomena currently grouped together under that rubric be united in that way (Lewis and Haviland-Jones [2000])? Or is the term 'emotion theory' really a misnomer; a mixed amalgam of different strands of inquiry that should probably be disentangled or eliminated (Griffiths [1997])? And what about the relation of emotion to other phenomena like cognition? That boundary may shift or even disappear depending on how one demarcates the domain of emotion theory. Or it may turn out to be a bigger gulf than expected. It is hard to imagine a more central issue in emotion theory than the question whether emotion is a natural kind.

In this discussion, I formulate and defend a version of the thesis that emotion is a natural kind. I will interpret that thesis as an empirical hypothesis about the status of emotion in the natural sciences. Specifically, it is the hypothesis that there exists a distinct natural kind of organism whose behavior is moved and governed by emotion. In other words, there are emoters. We humans are emoters, although we are not the only ones. A neurobiological and a psychological version of the hypothesis that emotion is a natural kind will be examined. Both have precedents in contemporary emotion theory. The central argument of the paper is that there are plausible reasons for entertaining the hypothesis that emotion is a natural kind. Natural kinds will be defined as homeostatic property clusters along the lines suggested by Richard Boyd and Paul Griffiths (Boyd [1991], [1999]; Griffiths [1997], [1999]). So defined, natural kinds do not require precise necessary and sufficient conditions. Neither do their constituent properties have to be fixed for all time. These and other characteristics of Boyd's homeostatic property cluster theory make it ideally suited to the mixed and rugged terrain of emotion theory. To be sure, this is not the only way to frame questions about natural kinds in emotion theory, or indeed elsewhere. However, as Griffiths has convincingly shown, when applied to emotion theory, Boyd's approach leads to an extremely interesting and fecund research program (Griffiths [1997]). In this discussion, I try and take Griffiths' project a step further.

Specifically, I want to extend his claim that *emotions* are natural kinds to the claim that *emotion* is a natural kind.

## 2 Griffiths on emotions as natural kinds

To start, let us distinguish the hypothesis that *emotion* is a natural kind from the hypothesis that *emotions* are natural kinds. For example, Paul Griffiths argues that there are several basic emotions that deserve natural kind status (Griffiths [1997], p. 78). However, he also denies that there is a unified natural scientific domain that corresponds to the term 'emotion' (*ibid.*, p. 1). In fact, he argues that the term 'emotion' should be eliminated (*ibid.*, pp. 15, 247). Nevertheless, despite this negative prognosis for *emotion*, Griffiths' account of the natural kind status of *emotions* provides an ideal starting point for our discussion of the natural kind status of emotion. Indeed, with some modifications and the addition of new evidence, his innovative approach to the natural kind status of emotions can be extended to emotion. Thus, in just the same way that emotions can be argued to be natural kinds, so the same can be argued of emotion. The argument is not that emotion is a natural kind *because* emotions are natural kinds. In fact, in both cases the corroborating empirical evidence is vastly different. In this first section we examine Griffiths' hypothesis that emotions are natural kinds and its supporting evidence.

Griffiths says that natural kinds are 'ways of classifying the world that correspond to some structure inherent in the subject matter being classified' (*ibid.*, p. 6). Originally, natural kinds were fundamental categories in the physical sciences: the planets of astronomy, the atoms of chemistry. Now the concept is also used to refer to any fundamental theoretical posit in a scientific domain: money in economics, depression in psychiatry. Ironically, natural kinds no longer have to be 'natural' in the full sense of the word. But they do have to be categories in a strong sense, epistemologically and ontologically. Epistemologically, natural kinds play a central role in helping state and capture reliable generalizations in the sciences. In that role they are central to explanation and prediction. Ontologically, they serve to specify the ontology of a science. In that role they tell us what a branch of science says there is.

Griffiths' central thesis is that there are certain basic emotions that are natural kinds. The evidence for that thesis is largely derived from the empirical research of Paul Ekman (Ekman [1992]). Ekman identifies six affect programs that he says are universally shared among humans. Those programs 'are adaptive responses to events that have a particular ecological significance for the organism' (Griffiths [1997], p. 89). They are complex, coordinated, and automated responses (*ibid.*, p. 77). More specifically, affect programs involve: (a) facial expressions, (b) musculoskeletal changes, (c)

vocal changes, as well as (d) endocrine and (e) autonomic system changes (p. 77). The six affect programs and their corresponding emotional labels are: fear, surprise, anger, disgust, sadness, and joy (*ibid.*, p. 78). These are the natural kinds of emotion according to Griffiths. Since they do not look anything like the traditional natural kinds of philosophy (gold, water etc.) the claim requires some explanation.

Griffiths tells us that a natural kind is a category that ‘brings together a set of objects with correlated properties’ (*ibid.*, p. 188). An important condition for being a natural kind in his sense is that any such category must exhibit *causal homeostasis*. That occurs when the set of correlated properties lumped together in a category ‘has some underlying explanation that makes it projectible’ (*ibid.*). The reason for this requirement is that for a category to count as a genuine natural kind it must be able to function in causal generalizations and predictions in the explanatory domain to which it belongs. Among other things, the correlations it identifies must ‘hold up in unobserved instances’ (*ibid.*). The basic idea is that natural kinds are theoretically important in helping capture reliable generalizations. An important consequence of this account is that it entails a very broad conception of what a natural kind is. Traditional natural kinds such as those of chemistry are the richest (*ibid.*, p. 190). They have their properties because of their internal microstructure (*ibid.*, p. 189). Next in line are biological kinds. Here it is not so much internal microstructure as more external factors that count. In the case of biological kinds, ‘the causal homeostatic mechanism is descent’ (*ibid.*). Natural kinds of that sort ‘form projectible categories because their members are descended from a common ancestor’ (*ibid.*). Whatever their specific homeostatic mechanism may be, the primary epistemological role of natural kinds is their role in explanation and induction (*ibid.*, p. 198).

The notion of homeostatic property clusters comes from Richard Boyd (Boyd [1991], [1999]). He urges us to give up the idea that natural kinds require precise necessary and sufficient conditions for their definition.<sup>1</sup> That requirement, he says, is ‘a holdover from traditional empiricist conceptions of linguistic precision which must be abandoned once it is agreed that kind definitions must conform to the (sometimes messy and complex) casual structure of the world’ (Boyd [1991], pp. 142–3). If being a natural kind required fixed necessary and sufficient conditions, then many of the

<sup>1</sup> Ian Hacking disputes Boyd’s claim that traditionally natural kinds have required definitions in terms of necessary and sufficient conditions (Hacking [1991b], p. 152). However, the ‘tradition’ Hacking is concerned with starts with Mill (Hacking [1991a], p. 111). Others trace the idea back to Aristotle (Churchland [1982], Solomon [1995]). On that view, it is not implausible to view natural kinds as requiring definition in terms of fixed and precise necessary and sufficient conditions. This seems to be how many philosophers who discuss the natural kind status of emotion have understood the issue.

theoretical posits we might want to count as natural kinds would not qualify. Homeostatic property clusters do not present this problem. Their unity and defining conditions are mainly causal rather than conceptual (*ibid.*, p. 141). Homeostatic property clusters can also be imperfect, meaning that they can have indeterminate extensions. This occurs when ‘some of the properties in the cluster are absent or some of the mechanisms inoperative’ (*ibid.*, p. 142). Boyd’s point is that despite these imperfections, kinds of these sorts may still function successfully in induction and explanation. Griffiths’ natural kinds of emotion are meant to be part of this expanded causal theory of natural kinds. They do not appear to have precise necessary and sufficient conditions. But they do share clusters of correlated properties that permit them to play a productive role in identifying and capturing reliable generalizations about emotional behavior. We posit basic emotions and treat them as natural kinds because without them we could not state and capture the kinds of projectable generalizations about emotion we want to. In circumstances like these, ‘natural kinds reflect a strategy of deferring to nature in the making of projectability judgments’ (*ibid.*, p. 139).

### 3 Panksepp on emotions as natural kinds

Griffiths is not alone in believing that there are some basic emotions that are natural kinds. Recently, the neuroscientist Jaak Panksepp has argued for a similar thesis. He argues that basic emotions are ‘natural kinds within the brain’ (Panksepp [2000]; see also Panksepp [1998], pp. 46, 305). Panksepp’s list of basic emotions is different from Griffiths’. He lists seven basic emotions compared to Griffiths’ six. They are: seeking, rage, joy, distress, care, lust and play (Panksepp [2000], p. 144; [1998], pp. 41–58). His basic emotions are ‘emotional systems in the brain’ (Panksepp [1998], p. 41). Depending on the purposes at hand, he refers to them as command, executive, or operating systems (*ibid.*, p. 49). Speaking of basic emotions as command systems ‘implies that a circuit can instigate a full-blown emotional process’ (*ibid.*). Referring to them as operating systems means that they can ‘coordinate and synchronize the operation of several subsystems’ (*ibid.*). Finally, talk of executive systems is meant to imply that ‘a neural system has a super-ordinate role in a cascade of hierarchical controls’ (*ibid.*). At one point Panksepp even suggests that his basic emotions can be viewed as ‘organ systems’. This he says is meant to conceptualize the fact that ‘each system is composed of an anatomical network of interconnected neurons and endocrine, paracrine, and immune influences’ (*ibid.*).

Griffiths’ affect programs are primarily found in humans. The possibility that they may have homologues in other species is mentioned but not stressed (Griffiths [1997], p. 8). Panksepp is both clearer and bolder about the scope of

his basic emotional circuits. He claims that they are shared by all mammals and correspond to 'specifiable neural substrates within the mammalian brain' (Panksepp [2000], p. 137). His central thesis is that 'basic emotional processes emerge from homologous brain mechanisms in all mammals' (Panksepp [1998], p. 51). Panksepp believes that because of their shared mammalian heritage, humans and other mammals are capable of internal affective experiences, which at times he refers to as emotional feelings (*ibid.*, p. 12). An important aspect of his theory is that those affective experiences 'not only sustain unconditioned behavioral tendencies but also help guide new behaviors by providing simple value-coding mechanisms that provide self-referential salience, thereby allowing organisms to categorize world events efficiently so as to control future behaviors' (*ibid.*, p. 14). The claim that emotional feelings are a key ingredient in the scientific study of emotion is one of the most controversial and interesting aspects of Panksepp's overall theory. Another innovative aspect is the role he ascribes to neuropeptides in the definition and elaboration of his seven primitive emotional circuits (*ibid.*, pp. 97–120). Suffice it to say that this is an enormously complex theory. Nonetheless, philosophically it is very similar to Griffiths' on the matter of the natural kinds of emotion.

Both Griffiths and Panksepp agree that there is a strong case to made for the hypothesis that some basic emotions are natural kinds. Moreover, both agree that specific clusters of physiological and neurobiological properties play a key role in defining those natural kinds. Finally, both agree that the natural kinds of emotion should be classified and individuated by homology rather than analogy. This last point is worth pondering. It is the springboard for our extrapolation from the claim that emotions are natural kinds to the hypothesis that emotion is a natural kind.

In biology, classification by analogy tends to be by function and resemblance, while homology is a matter of history and origins (Griffiths [1997], pp. 12–13; Elster [1999], pp. 7–8; Panksepp [1998], p. 17). Thus 'human arms and bat wings are *homologous* because they both arise from the genetic information that controls forelimb development' (Panksepp [1998], p. 17). On the other hand, 'the wings of birds and bees are *analogous*—serving similar functions—even though they do not share a common genetic inheritance' (*ibid.*). For Panksepp, the concept of homology is especially important because it permits him to use evidence acquired through the study of mammals such as rats and dogs to formulate and test hypotheses about the workings of emotion in humans. For Griffiths, homology is especially important because it provides an alternative to analogy as a method of classification. However, these are really only differences in emphasis. In both cases the main result of adopting homology as a criterion of classification is that emotions get classified in terms of particular sequences of evolutionary

events (Griffiths [1997], p. 13). Thus, in both cases the natural kinds of emotion are identified and classified in terms of shared evolutionary history and origin, and not primarily in terms of similarity of function and analogy. Individual basic emotions of the same kind may at times resemble one another in function or other ways. But what makes them instances of a common category is ultimately their origin and history rather than shared function or resemblance.

To conclude, both Griffiths' affect programs and Panksepp's basic emotional circuits are evolutionary homologies. It is in that sense that the basic emotions they identify are natural kinds. Where the two part ways is on the status of emotion. Griffiths' believes that the term 'emotion' should be eliminated. That makes it impossible to argue that emotion is a natural kind. On his side, Panksepp looks very much like he wants to defend a version of the hypothesis that emotion is a natural kind.

#### **4 Emotion as a neurobiological kind**

In Panksepp's theory, rats and apes and humans are all capable of experiencing and manifesting basic emotions. As a result of their shared mammalian ancestry, they share homologous brain systems that elaborate general states of those sorts. Consider rage, one of Panksepp's basic emotions. The reason why all mammals are capable of rage is that they share homologous brain systems that can generate and sustain rage. In the language of natural kinds, those systems form a 'cluster of properties' defined by a specific neuroanatomical and neurochemical profile (Panksepp [2000], p. 144, Table 9.1). Instances of that general neural property cluster will differ across species, within species, and even within individual organisms over time. But the general patterning of the cluster remains the same. Otherwise these would not be instances of rage. Thus a heart can beat in rage in many ways and at many speeds. But if it beats as a result of the brain system that elaborates rage, then it beats in rage. This, very roughly, is what it means to say that basic emotions are defined by neural property clusters. It provides a way to link Panksepp's work on the neurobiology of emotion with the tradition in natural kinds associated with Boyd and Griffiths.<sup>2</sup>

Panksepp and Griffiths believe that the general patterning of neural property clusters responsible for basic emotions can be traced back to common evolutionary origins. This is why they say that basic emotions are evolutionary homologies. It also why they say that basic emotions are natural kinds. I now want to argue that this conception of a natural kind can be applied to emotion itself; that is, to emotions as a class. More precisely, I will attempt to show that such an argument can be found in the work of

<sup>2</sup> See Wilson [1999] for an interesting precedent to the effect that Boyd's conception of natural kinds can be extended to neural taxa.

Panksepp. My strategy will be to argue that emotion is a neurobiological natural kind for the very same reasons that emotions are natural kinds.

Both Griffiths and Panksepp trace the origins of their claim that emotions are natural kinds back to Darwin (Griffiths [1997], pp. 44–77; Panksepp [2000], p. 138). Darwin was concerned with the *expression* of emotion, which is different from the ‘state of mind’ that constitutes the core of the emotion itself (Darwin [1872/1998], p. 84). As Griffiths tells us, he held ‘a version of the feeling theory of emotion’ (Griffiths [1997], p. 45). Panksepp makes the same point, noting that for Darwin ‘the key feature of emotions was a feeling tone’ (Panksepp [2000], p. 138). The view that emotions were largely mental feeling states created a problem for Darwin, since feelings cannot be objectively scientifically studied. He tried to circumvent the problem by focusing his efforts on how emotional feelings get expressed in behavior. Nevertheless, it would be a mistake to think that his theory was concerned with the expression of emotion alone. His view appears to have been that emotions were feelings *with* characteristic types of behavioral manifestations. On that interpretation, the basic principles of emotional expression are also basic principles of emotion.

Griffiths provides an excellent account of the details of Darwin’s theory and a fair assessment of its strengths and weaknesses (Griffiths [1997], pp. 44–76). However, unlike Panksepp, he does not appear to be impressed with Darwin’s vision of emotion theory as a unified domain. In fact, according to him, emotion theory fractures into two, even maybe three, distinct explanatory domains. There exists a basic distinction between the basic emotions that correspond to affect programs and higher emotions (*ibid.*, p. 229). At times, he also speaks of a three-way fracturing that also includes what he calls socially sustained pretenses (*ibid.*, p. 17). The details of this proposed division are fascinating but need not concern us here. The relevant point is that for Griffiths the term ‘emotion’ does not designate a unified explanatory domain. He reserves that term for affect programs only. These are what emotions ‘really’ are (*ibid.*, p. 230). Panksepp, on the other hand, does believe in a unified view of emotion theory. He also ascribes a unified view of emotion to Darwin and interprets him as viewing emotion as a relatively independent scientific domain of inquiry, defined by its own special explanatory laws and principles. Because he could not study the underlying neural substrates of emotions in sufficient detail, Darwin was limited to charting the laws that govern emotional expression and behavior from the outside. With the tools of modern neural science, Panksepp thinks he can improve on this attempt to tackle the problem from the inside. His aim is to characterize the basic properties of emotions in neural terms. In Panksepp’s view, the reason Darwin is so important is because he ‘realized that emotions had certain basic properties’ (Panksepp [2000], p. 138). How exactly this

claim should be understood in relation to Darwin is problematic.<sup>3</sup> It is in any case central to Panksepp's account, which is our main concern, and in that respect its interpretation is clear enough.

In *Affective Neuroscience*, Panksepp tackles the 'problem of defining emotions' (Panksepp [1998], pp. 47–50). Specifically, he attempts to provide a 'neurally based definition of emotion' (*ibid.*, p. 47). His definition is an extremely complex description of 'the various neural interactions that characterize all major emotional systems of the brain' (*ibid.*, p. 48). He describes seven neural criteria—clusters of neurobiological and physiological properties, if you will—that are shared by all his basic emotional systems. The details are too elaborate to summarize here (*ibid.*, pp. 48–9; see also Fig. 3.3). Nonetheless, the following quotation should convey the basic idea:

(1) The underlying circuits are genetically predetermined and designed to respond unconditionally to stimuli arising from major life challenging circumstances; (2) These circuits organize diverse behaviors by activating or inhibiting motor subroutines and concurrent autonomic-hormonal changes that have proved adaptive in the face of life-challenging circumstances during the evolutionary history of the species; (3) Emotive circuits change the sensitivities of sensory systems that are relevant for the behavioral sequences that have been aroused; (4) Neural activity of emotive systems outlasts the precipitating circumstances; (5) Emotive circuits can come under the conditional control of emotionally neutral environmental stimuli; (6) Emotive circuits have reciprocal interactions with the brain mechanisms that elaborate higher decision making processes and consciousness. (Panksepp [1998], pp. 48–9)

In addition to the above six criteria, there is a seventh criterion that Panksepp deems very important. It is that 'emotive circuits must be able to generate affective feelings' (*ibid.*, p. 49).

<sup>3</sup> Panksepp suggests that what Darwin meant by this is that 'each basic emotional system of the brain (i.e. his "principle of action", due to the constitution of the nervous system) interacts with other systems (his "principle of anti-thesis") and is also accompanied by the vast baggage of accumulated learning (his "principle of associated serviceable habits")' (Panksepp [2000], p. 138). The above claim and its purported clarification are ambiguous. They can both be interpreted to mean that Darwin believed that *individual emotions* each have their own basic properties. Or they can both be interpreted to mean that *as a class* emotions share properties on a more abstract level. On the one hand, Panksepp seems to be saying that individually taken each emotion is elaborated by a specific patterning of Darwin's three principles of emotion. However, on the other hand, he appears to be saying that as a class emotions share properties that arise from being the product of the three principles. In the former case, we have a precedent for the claim that emotions are natural kinds. In the later, we have a precedent for the hypothesis that emotion is a natural kind. The first interpretation supports Panksepp's claim that 'emotions are natural kinds within the brain' (Panksepp [2000], pp. 137, 143; see also Griffiths [1997], pp. 44–76). The second interpretation is more controversial. It supports the view that emotion is a natural kind. In the end, only Panksepp can tell whether he intended the first or second interpretation. However, it is clear that the second interpretation is consistent with his own theory. In fact, he makes the exact same point himself.

Now according to Panksepp, all basic emotional systems share these higher order neurobiological and physiological principles of functioning and organization. Although basic emotions are themselves defined by clusters of neurobiological and physiological properties, what makes them all cases of *emotion* is the fact that normally they also share most or all of these more abstract general defining properties. Those criteria also form a cluster of neurobiological and physiological properties, but on a higher order of neural function and organization. Note that Panksepp's seven criteria are not precise necessary and sufficient conditions. They are criteria that permit some flexibility and imperfection, and so some indeterminacy, as to if and when a neural circuit elaborates a basic emotion. Most important for our purposes is the fact that the property clusters captured by the seven criteria are evolutionary homologies. They are inherited neural structures and principles of brain organization shared by all creatures that have the relevant mammalian ancestry. This supports the view that, as a class, Panksepp's basic emotional systems should count as a natural kind. Although he does not put it this way, he is in fact a defender of the hypothesis that emotion is a natural kind.

To sum up, any creature with a brain that has evolved to the point where it functions according to Panksepp's seven criteria of emotional organization is an *emoter*. In this sense, emoters and their basic emotions can be said to form a natural kind. They constitute a distinct class of natural being that encounters and responds to the world through brain systems that reflect distinct biological values. It is worth mentioning that on Panksepp's view the link between values and emotions is intimate: basic emotions reflect and elaborate innate biological values (Panksepp [1998], pp. 303–14). This and the suggestion that many psychiatric illnesses arise from malfunctions in basic emotional circuitry are among the most fascinating aspects of his wide ranging psychobiological synthesis. Those aspects of his work would not have the significance he ascribes to them if emotion did not form a natural kind. Emotions are not simply scattered patterns of arousal and response with no underlying unity or purpose. They are the shared evolutionary heritage of a distinct kind of organism that comes to the world designed to appraise it according to specific biological values and priorities. It is probably not an exaggeration to say that for Panksepp the idea of emotions without emotion is senseless. Only emoters have emotions.

## 5 Emotion as a psychological kind

Panksepp's hypothesis that emotion is a natural kind is a neurobiological hypothesis. He is concerned with uncovering the neural substrates of emotion. The homologies he identifies are primarily subcortical in origin;

vestiges of the older parts of the mammalian brain (Panksepp [1998], pp. 70–9). But like a tree, in many species the subcortical systems he describes branch out into the cortex with multiple influences and feedback loops (Panksepp [1998], p. 302). Panksepp's neuroscience is meant to provide a foundation for emotion theory. However, his overall goal for emotion theory as a whole is a 'psychobiological synthesis' in which psychology and other sciences will continue to play an important role (*ibid.*, p. vii). He continually stresses the fact that 'the existence of basic emotional circuits in the brain does not diminish the need for complementary perspectives' (Panksepp [2000], p. 140). So there is an important sense in which Panksepp is not a reductionist, even though he sometimes says he is (*ibid.*, p. 20). His affective neuroscience cannot do without the laws and posits of the various other sciences that investigate emotion.

In keeping with this pluralistic vision, Panksepp expresses the hope that eventually the findings of the various branches of emotion theory will exhibit consilience regarding its fundamental laws and posits (*ibid.*, p. 140). William Whewell defined consilience as a situation where 'an induction, obtained from one class of facts, coincides with an induction, obtained from another different class' (Butts [1968], p. 139). The 'inductions collected from one class of facts, supply an unexpected explanation of a new class' (Butts [1968], p. 159). An example of consilience in emotion theory would be if both neurobiological and psychological theory and evidence converged on the existence of basic emotions of the sort Panksepp identifies. The affinities between Panksepp's and Griffiths' accounts of the natural kinds of emotion provide some basis for optimism on this matter. There are therefore reasons for being optimistic about consilience regarding the natural kind status of *emotions*. I now want to argue that there are also reasons for being optimistic about the prospects for consilience regarding the natural kind status of *emotion*. For in addition to being a distinct neurobiological kind in the manner described above, it also appears that emotion may be a psychological kind; not just any psychological kind, but just the sort of kind that reflects the same inductions on which the neurobiological kind status of emotion is based. Recall that Griffiths believes that the term 'emotion' should be eliminated from psychology. He argues that 'emotion' is not a real psychological category (Griffiths [1997], pp. 1–2, 228–9). But what would it be like for emotion to be a genuine psychological category? There are in fact numerous precedents for that hypothesis. They constitute an important source of evidence for the hypothesis that emotion is a distinct psychological kind. They show that such a hypothesis is not only possible, but also plausible.

In a significant number of philosophical analyses of the emotions, it is argued that these should be treated as a distinct kind of psychological state,

irreducible to beliefs and desires, and irreducible to physiological or other physical states (de Sousa [1987], Gordon [1987], Lyons [1980], Solomon [1976]). Many psychological accounts of emotions also treat these as a distinct kind of psychological state (Arnold [1960], Lazarus [1991]). In both cases, it is possible to view emotions as states that involve some form of *representation*. In philosophical terminology, they are intentional states about objects and events in the world. Generally, the idea is that emotions involve their own distinct mode of representation. In psychology, this is sometimes expressed by saying that emotions involve evaluation or appraisal (Lazarus [1991]). In philosophy, it is expressed by saying that emotions are normative or evaluative judgments (Solomon [1976]). Whatever the case may be, the general idea is that emotions are a distinct kind of representational state. Panksepp captures this point by distinguishing *cognitive* from *affective* representation (Panksepp [1998], pp. 30–1).

To the claim that emotions form a distinct kind of representational state, we can add the claim that emotions are governed by special laws and principles. In this respect, philosophers tend to speak of the logic or structure of the various emotions (Gordon [1987], Solomon [1976]). Some, like Aaron Ben Ze'ev ([2000]), prefer to speak of generalizations. On their side, psychologists sometimes speak of laws of emotion (Frijda [1988]). The point here is that there are reliable generalizations and principles of inference that govern emotional behavior. To the extent that emotions are a distinct kind of representational state, governed by their own special laws and regularities, it is plausible to view them as a distinct psychological kind. Vision might be argued to be a psychological kind in this sense. It is a distinct psychological domain of inquiry, defined by its own special representational kinds and governed by its own special laws and regularities (Marr [1982]). Language might also be argued to be a psychological kind in this sense (Fodor [1975]). An even better example is cognition.

In his attempt to articulate the theoretical foundations of cognitive science, Zenon Pylyshyn argues that *cognizer* is a natural kind (Pylyshyn [1984], pp. 35, 113).<sup>4</sup> He describes the domain of cognitive science as ‘knowing things’ and argues that cognition has its own laws and regularities (*ibid.*, p. ix). What is special about cognizers is the fact that their behavior is representation governed (*ibid.*, pp. 1–28). Cognizers act and reason on the basis of internally encoded mental rules and representations. Pylyshyn’s main point is that there are many regularities in human behavior that can only be

<sup>4</sup> In answer to the question ‘Is Thinker a Natural Kind?’, Paul Churchland answers ‘quite possibly, bordering on probably’ (Churchland [1982], p. 237). He defines a natural kind as ‘a kind that figures in some natural laws’ (*ibid.*, p. 223). This is a typical characterization of what it means to be a natural kind and is compatible with homeostatic view of natural kinds espoused here.

captured if we posit mental rules and representations. An analogous case can be made for the category *emoter*. The argument is that there exist regularities in human and animal behavior that can only be captured if we suppose organisms represent the world with affective categories and act according to special affective rules and principles. Different versions of this general approach can be found in the literature. For example, Robert Gordon argues that emotions are distinct propositional states with their own particular cognitive structure and special ties to action (Gordon [1987]). This can be interpreted to mean that emotions are a distinct species of representational state, governed by their own special affective principles and regularities (Charland [1995a], [1995b]). The basic argument can be traced back to Donald Hebb, who challenged psychologists to try and explain animal behavior without resorting to affective terms and categories (Hebb [1946]). He concluded that this was impossible. Hebb's fundamental insight is that psychological explanation requires emotion principles and categories. It is a short step from this to the view that emotion constitutes a special psychological *system* (Ben Ze'ev [2000], p.175).<sup>5</sup> One reason emotion is a distinct system is because it has its own special regularities (*ibid.*, pp. 531–2). Another is that it has its own special affective representational posits; namely, appraisals (*ibid.*, pp. 72–3). Taken together, these two premises provide good grounds for concluding that emotion is a natural kind. To be sure, the evidence I have presented for this hypothesis has been very general and schematic. Nonetheless, hopefully it is sufficient to establish the *plausibility* of the hypothesis that emotion is a distinct psychological kind, which is the main goal of this paper.

Now that we have established the plausibility of the hypothesis that emotion is a neurobiological and a psychological kind, we can inquire into the question of their consilience. The key is folk psychology. Griffiths argues that folk psychology is no basis for emotion theory (Griffiths [1997], pp. 1–2). Panksepp takes the opposite stance. He attempts to build his theory from the terms and notions of folk psychology. His position is that 'some of the old emotional words used in everyday folk psychology can still serve our purposes well, since they approximate the realities that exist, as genetic birthrights, within mammalian brains' (Panksepp [1998], pp. 12, 306). Clearly, Panksepp's purpose is not to vindicate the posits and principles of folk psychology at all costs. Rather, his aim is to test and refine them in order to uncover their neural basis. He fully expects this will result in some conceptual regimentation of those basic terms and notions. Indeed, his own struggles with what to call his basic emotions reflect this 'bottom-up'

<sup>5</sup> Ironically, Ben Ze'ev ([2000]) denies that emotion is a natural kind. I argue in the concluding section of this paper that, like many others, he has been misled by an overly restrictive conception of natural kinds.

influence of neuroscience on folk psychology (*ibid.*, p. 51). In the end, Panksepp hopes that some of the posits and generalizations of folk psychology will be corroborated by a mature neuroscience (*ibid.*, p. 304). This is where consilience comes in.

The prospects for consilience in emotion theory currently seem to lie in two sets of mutually reinforcing inductions. First, both psychology and neuroscience appear to be converging on a small set of basic affective representational posits that appear to be universally shared among mammals. Second, both psychology and neuroscience appear to be converging on the existence of a small set of regularities that govern the working of those basic posits and tie them to distinct classes of behaviors. Taken together, these claims mean that psychology and neurobiology appear to be converging on the autonomy of emotion as a distinct natural explanatory domain. This is the version of the hypothesis that emotion is a natural kind we set out to clarify.

Homology is the key to understanding the hypothesis that emotion is a natural kind. Because of their common ancestry, mammals come to the world with certain genetically-ingrained neural capacities that permit them to appraise their environment and respond accordingly. Their encounter with the world is mediated by a few primitive affective representational categories, which are tied to specific preset action tendencies and responses. Variants of those basic individual appraisal and response patterns can be found across mammals on account of shared homology in relevant brain structures. These are the natural kinds of emotion—the emotions that are natural kinds. If Panksepp is right, then what makes basic *emotions* possible is the fact that *emotion* is a natural kind. For it is only in virtue of the higher and more abstract principles of brain organization that define emotion that the various basic emotions are tied together in the manner they are, and develop in the manner they do. Emotions, then, are natural kinds. But this only makes empirical sense on the assumption that emotion is a natural kind.

## 6 Response to the mantra

We started this discussion with the observation that philosophers who write about emotion typically deny that it forms a natural kind. It is now time to examine and assess those arguments. Many of them appear to rely on an anachronistic or inappropriate understanding of what it means to be a natural kind. All of the arguments fail in different ways and for different reasons that are interesting to consider.

In a wide-ranging study of emotion and addiction, Jon Elster declares that the natural kind status of emotion is an ‘unresolved issue’ (Elster [1999], p. 12). He contrasts the case of emotion with addictions, which he says do

form a natural kind (*ibid.*, p. 71). What he means by this is that addictions form a relatively homogenous category, while emotions do not. Elster interprets the natural kind status of emotion as an issue that has to do with whether they 'are unified by a common causal mechanism or whether they are simply lumped together on the basis of phenomenological similarities' (Elster [1998], p. 239, note 2). Like Griffiths and Panksepp, he is impressed with the notion of homology as a mode of classification (Elster [1999], p. 7; [1998], pp. 239–42). However, he does not apply that notion to individual emotions as they do. Instead, he applies it to emotion. In doing so he turns the issue of the natural kind status of emotion into an empirical one. That is the sense in which it is unresolved for him.

We have seen that there are in fact plausible empirical reasons for believing that emotion may be a natural kind in precisely the sense Elster is concerned with. Panksepp's neurobiological hypothesis about the common properties of emotion is such a hypothesis. To be sure, Panksepp's research does not definitively resolve the debate over whether emotion is a natural kind. But it does show that there is a plausible hypothesis of that sort to be entertained. Therefore, Elster's ambivalence about the natural kind status of emotion needs to be reassessed. Until then, there is no reason to view the natural kind status of emotion as an unresolved issue for the reasons he does. In fact, the evidence provided here clearly indicates that there are plausible empirical reasons for leaning toward ascribing natural kind status to emotion. The same considerations apply to Griffiths' pessimistic assessment of the natural kind status of emotion and the category 'emotion'. Therefore, neither Elster nor Griffiths provides convincing reasons to doubt or reject the hypothesis that emotion is a natural kind.

There is another empirical source of doubt that has been raised about the natural kind status of emotion. It concerns the natural kind status of emotion on a strictly neuroscientific level. Joseph LeDoux has argued that 'there is no such thing as the "emotion" faculty' (LeDoux [1996], p. 16). His argument is that 'there is no single brain system dedicated to this phantom function' (*ibid.*). A similar argument is advanced by Griffiths, who appears to believe that in order for a neurobiological entity to be a natural kind there must be 'one kind of process' (Griffiths [1997], p. 14). Speaking of the general concept of emotion, he writes: 'it is meant to be a kind of psychological process that underlies a certain range of behaviors. But there is no one kind of process that underlies enough of this behavior to be identified with emotion' (*ibid.*).<sup>6</sup>

<sup>6</sup> Elsewhere Griffiths expands on this point. He writes: 'although I have argued that some emotions are affect programs, I do not believe that the category emotion in general should be identified with the category of affect program phenomena' (Griffiths [1997], p. 241). Note that this appears to leave open the possibility that the general category affect program (which comprises the various affect programs) is a natural kind.

The irony with this argument is that Panksepp himself says that there is no one brain system that corresponds to emotion. He writes: ‘there is no single motivational circuit in the brain, and there is no unitary emotional circuit there’ (Panksepp [2000], p. 136). However, this does not prevent him from postulating that there are ‘shared psychoneurological properties of basic emotional systems’ (*ibid.*). And on that basis it makes sense to believe that emotion may be a neurobiological kind. So the idea that in order to count as a natural kind ‘emotion’ must consist of a single brain system is both implausible and too simplistic.

None of the empirical objections we have considered provide sufficient grounds for doubting the plausibility of the hypothesis that emotion is a natural kind in the neurobiological sense. Neither do they provide any evidence it should be rejected. Unless there are more convincing reasons, the hypothesis stands firm. There remain the more philosophical objections to the natural kind status of emotion. These tend to focus on the psychological version of the hypothesis that emotion is a natural kind. In general they aim to show that emotions do not form a distinct psychological class.

To start, consider the following philosophical argument proposed by Ronald de Sousa. He writes:

The formal object of belief—its criterion of success—is truth; the formal object of wanting is goodness. If they formed a congeneric species with beliefs and wanting, one would expect to discover a single criterion of success, corresponding to truth and goodness, for emotions as a class. I shall argue that there is no such formal object of emotions: in that precise sense it is each emotion type that is congeneric with belief and wanting, not emotions as a class. (de Sousa [1987], p. 20)

The conclusion of this argument is that emotions do not form a distinct class of psychological entities. The reason is that they do not have their own formal object. There are two problems with the argument. First, it is based on an analysis of emotions in terms of formal objects. This particular philosophical approach to the analysis of emotions is not shared by many researchers who believe that emotion is a distinct psychological kind. To that extent, the argument begs the question at issue. Second, the argument is almost totally a priori. This is unlikely to impress those who want to argue that emotions form a distinct psychological class on the basis of empirical evidence. Therefore, even though the argument may be sound enough on its own terms, it is likely to be considered a red herring from the point of view of anyone concerned with the psychological status of emotion from an empirical point of view. It must be said that de Sousa has written about the problem of natural kinds in biology (de Sousa [1989]). Unfortunately, he does not appear to have explicitly confronted the question of the natural kind status of emotion from that perspective.

Some philosophers object to the natural kind status of emotion on ostensibly more empirical grounds. A good example is Amelie Rorty. Like de Sousa, she argues that ‘emotions do not form a natural class’ (Rorty [1978], p. 141). However, her argument is not as abstract or aprioristic as de Sousa’s. It consists of a rich description of how different philosophical, literary, and scientific conceptions of passion, emotion and sentiment have co-evolved and intermingled in the history of Western thought (Rorty [1978], [1982]). Because it is so rich in details, her argument is hard to summarize. Perhaps the best way to sum up its import is to say that the attempt to define emotions in terms of necessary and sufficient conditions is futile. Emotions cannot form a distinct class of mental state, because there are no common properties shared by all of them in virtue of which they can be assembled into such a class. Emotions then are simply too heterogeneous to form a unified psychological domain. They do not form a natural class.

It is interesting that both de Sousa and Rorty use the terminology of ‘classes’ in their respective discussions. Neither explicitly refers to ‘natural kinds’. Nevertheless, what they mean seems clear enough. Both of their arguments hinge on a common premise that invokes the relationship between emotions and their properties. De Sousa says that emotions do not have the property of sharing a single formal object. Rorty despairs of finding any one fixed set of properties common to all emotions. In both cases, the argument is that because emotions do not have the right type or number of properties, they do not form a natural class. This, roughly, is Russell’s sense of a natural kind. He defined a natural kind as ‘a class of objects all of which possess a number of properties that are not known to be logically connected’ (Hacking [1991a], p. 112). So we need not be concerned with the interchangeable terminology of natural kinds and natural classes here, although distinguishing these can sometimes be important for other purposes (*ibid.*, pp. 115–16). The relevant point here is that some philosophers of emotion may be denying that emotion is a natural kind based on something like the Russellian account of natural kinds. In that set-theoretical sense, natural kinds might easily be thought to require a fixed set of properties. They might also be thought to require precise necessary and sufficient conditions. All of this is very different from the view that natural kinds are homeostatic property clusters. In sum, Rorty and De Sousa appear to be relying on an overly restrictive sense of what counts as a natural kind. For that reason, they are unable to do justice to the natural kind status of emotion. There are other problems of this sort in the literature on the natural kind status of emotion.

Amelie Rorty’s claim that emotions do not form a natural class is based on historical and cultural evidence, even though much of it is very general and anecdotal. A more concrete empirical argument of that sort is advanced by Robert Solomon, who also argues that emotion is not a natural kind. He

writes: ‘in the context of cross-cultural comparison, emotion is not an ultimately defensible category, or what Aristotle called a ‘natural kind’ (Solomon [1995], p. 177). The allusion to Aristotelian natural kinds is significant here, since these were supposed to be fixed properties of just the sort that are definable by necessary and sufficient conditions. The problem is that the Aristotelian conception of natural kinds is incompatible with the fact that there are things we want to call natural kinds that change and evolve. This is why the Darwinian account of species posed such a problem for biology when it appeared (de Sousa [1989]). It is also why philosophers like Boyd and Griffiths urge us to adopt a more liberal and flexible definition of that notion based on property clusters and descent. Solomon’s rejection of the natural kind status of emotion can therefore be challenged on the grounds that it is based on a conception of natural kinds that is not well-suited to the present context. However, the empirical evidence he offers in defense of his position still requires comment.

Solomon cites numerous anthropological and cross cultural studies and argues that these show that many cultures and languages do not share our category ‘emotion’ (Solomon [1995], p. 175). He concludes that ‘at best, it would seem that we can say that “emotion” delineates a more or less agreed upon set of phenomena in English, at the present time, with a considerable gray area around the margins and little agreement about what it is that qualifies a phenomenon as an emotion’ (*ibid.*).

But why should ‘emotion’ be a fixed and perfect folk psychological kind? There is absolutely no reason to expect or insist on this. Of course, there are variations in the incidence of emotion terms and even maybe the category ‘emotion’ across cultures. However, from the vantage point of Panksepp’s *Affective Neuroscience* and projects like it, this is to be expected and does not in itself constitute a problem. In fact, Solomon grants the underlying assumptions of Griffiths’ and Panksepp’s hypotheses that some basic emotions are natural kinds. For example, he observes that ‘our human condition or our neurology might be such that we rather routinely develop much the same emotional repertoire, no matter how embellished or how complicated’ (*ibid.*, p. 183). He goes on to say that ‘there may be some emotions that are so “basic”, whether by virtue of our condition or our neurology, that they appear, virtually unblemished, in every society’ (*ibid.*). This, he says, is why ‘some emotions—anger and fear, for example—seem to be more or less regulars’ (*ibid.*, p. 177). Thus, although he does not apply the natural kind terminology to basic emotions, Solomon provides just the sort of evidence that Panksepp and Griffiths use to argue that some basic emotions are natural kinds. At the same time, he provides corroborating reasons for Panksepp’s hypothesis that emotion is a neurobiological kind. For it is partly in virtue of their neurology that humans are held to be capable

of shared basic emotions. Solomon's argument that emotion is not a natural kind therefore fails. He relies on an outdated notion of natural kinds and actually provides just the sort of empirical evidence that leads others to the conclusion that emotion and its emotions are natural kinds.

Another author who denies that emotion is a natural kind is Aaron Ben-Ze'ev. He proposes a model of emotions according to which they are prototypes. His main thesis is that 'emotions in general, as well as each particular emotion separately, constitute prototypical categories' (Ben-Ze'ev [2000], p. 6). Membership in a prototypical category 'is determined by an item's degree of similarity to the best example in the category' (*ibid.*). An important consequence of viewing emotions in this way is that 'there is no single essence which is a necessary and sufficient condition for all emotions' (*ibid.*). Instead, membership in a general category of emotion is a matter of similarity and degree. Prototypical categories do not have clear-cut boundaries.

The problem with Ben-Ze'ev's denial of the natural kind status of emotion is, again, the assumption that in order to be a natural kind an entity must have clear-cut boundaries and precise necessary and sufficient conditions. Once we dispense with that assumption, there is nothing incompatible between Ben-Ze'ev's prototype analysis of 'emotion' and the hypothesis that emotion is a natural kind. Indeed, much of what he says supports the hypothesis that emotion is a psychological kind in the sense described above. Thus, he views emotions as a distinct class of psychological phenomena with distinct characteristics and components that distinguish them from other psychological states and processes (*ibid.*, pp. 113–14). And unlike Griffiths, he believes that the general concept of 'emotion' has explanatory value (*ibid.*, p. 5). There are, he says, 'plausible generalizations about emotions' and 'general regularities typical of emotions' (*ibid.*, pp. 5, 531–2). So although he denies that emotion is a natural kind, Ben-Ze'ev can actually be classified as a defender of the hypothesis that emotion is a psychological kind.

There is one last example to consider. Like so many others, Jerome Neu argues that 'emotions are not natural kinds' (Neu [2000], p. 19). The argument he gives is that 'they have conventional boundaries' (*ibid.*). The reasoning here appears to be that natural kinds are defined by classes that have the boundaries they do by virtue of nature, not convention. What sets emotions apart from natural kinds in this sense is the fact that they are classed according to boundaries that are determined by convention. Neu argues that he does not believe it is possible to set precise boundaries for distinguishing emotions from other kinds of psychological states. The reason is that he does not think that 'those boundaries are precise' (*ibid.*, p. 298, note 3). Here again we seem to have a version of the anachronistic assumption that in order to count as a natural kind an entity must have fixed and precise

boundaries. To the extent that the assumption is misguided, the objection fails. The argument also fails simply because it is too general. There is nothing wrong with saying that many distinctions between emotions and other psychological states are a product of convention. But why generalize hastily to all aspects of all emotions? As Solomon points out, emotions ‘may be “constructed” but they are constructed out of something, from raw material that is, first of all, to be found in human experience, in the human body and in the human condition’ (Solomon [1995], p. 183). This leaves ample room for a legitimate empirical hypothesis that some basic emotions are natural kinds.

The above considerations apply to the natural kind status of individual emotions. However, Neu is also concerned with the natural kind status of emotion. Indeed, the above arguments are presented in the context of a discussion that is also concerned with the plausibility of ‘setting precise boundaries to the concept of emotion’ (*ibid.*, p. 298, note 3). The argument here appears to be that since individual emotions do not admit of sufficiently precise boundaries, so the concept of emotion as a whole is perniciously imprecise. Thus, because emotions do not form a precise enough kind, emotion cannot be a natural kind. The problem with this argument is the same as with the one above. Emotions do not need precise boundaries in order to count as natural kinds, neither does emotion.

Neu’s argument relies on a contrast between boundaries set by convention and boundaries set by nature. This sounds very much like the view that emotions are socially constructed. Social constructionists about emotion are unlikely to be happy with the suggestion that emotion and emotions are natural kinds. Suffice it to say that neither Panksepp nor Griffiths denies that there is much in emotion that is the product of society and convention. Moreover, there is nothing incompatible between their projects and inquiry into the social and political determinants of emotion. As researchers interested in the biological and evolutionary dimensions of emotion, they have simply chosen to explore emotions in those terms. The extent to which emotions and emotion might be socially constructed is important, even if it is often difficult to specify what exactly it means to say that something is socially constructed (Hacking [2000], pp. 18–19). Perhaps an even more interesting question is whether the natural kinds of emotion might be *interactive* kinds; that is, ‘kinds that can influence what is classified’ (*ibid.*, p. 103). Interestingly, Boyd’s account of natural kinds allows for what he calls relational kinds (Boyd [1999], pp. 153–4). So perhaps his account can accommodate interactive kinds. But these are difficult and complex issues that are best left for another time. There is a more urgent question that needs to be addressed before we can conclude this discussion.

## 7 Unification or fragmentation?

The argument for the natural kind status of emotion just proposed applies only to so-called 'basic' emotions. These are emotions such as fear, anger, joy, sadness, and so on. Where does this leave the more cognitive emotions such as shame, resentment, pride and envy? Griffiths distinguishes basic from higher cognitive emotions (Griffiths [1997], pp. 14, 77–99, 100–22). He argues that these two classes of emotion demarcate, respectively, two different explanatory projects. In the end, he opts for the view that only basic emotions are really 'emotions'. Thus, we should stop using the term 'emotion' as a general term of art to cover all 'emotions' (*ibid.*, pp. 14, 228–47). Only affect programs are emotions and only the study of affect programs is properly called 'emotion theory'. The rest should be called something else. Griffiths is certainly not alone in believing that what is traditionally called 'emotion theory' ranges over two separate classes of emotions. James, for instance, distinguishes coarse from subtle emotions along much the same lines (James [1890/1950], pp. 442–72). More recently, the neuroscientist Antonio Damasio has argued for a very similar distinction between primary and secondary emotions (Damasio [1994], pp. 131–4). However, unlike Griffiths, neither James nor Damasio suggests that subtle or secondary emotions are not 'emotions'. Both appear to view emotion theory as a unified domain. Are they justified in that assumption?

The answer to the above question is that it is probably too early to tell conclusively. What is clear is that there exists sufficient empirical evidence to argue that the basic emotions form a natural kind. That hypothesis at least is empirically plausible. On the matter of whether that kind will form a part of an expanded natural kind emotion category that includes cognitive emotions, it is important to be cautious. That said, there are some good reasons for believing that basic and cognitive emotions may be part of a wider natural kind category of emotion. For example, Damasio argues that secondary emotions depend on the same brain structures and mechanisms as basic emotions (Damasio [1994], p. 137). On his view, cognitive emotions are indeed separate and different from basic emotions in important ways. One difference is that they require supplementary cortical mechanisms for their operation. However, because of the manner in which cognitive emotions depend on basic emotions for their operation, it is conceivable that they may be homeostatic extensions and partial homologues of those basic emotional kinds. According to this interpretation, cognitive and basic emotions share a common causal ancestry, which they retain despite their different neuro-evolutionary trajectories.

Panksepp is more tentative than Damasio about the relationship between cognitive and basic emotions. Like Damasio, he is open to the fact that 'in

some yet undetermined manner, these secondary, cognitive-type emotions may [...] be linked critically to the primitive affective substrates we have discussed so far' (Panksepp [1998], p. 321). However, he also mentions the possibility that 'they may reflect newly evolved neural functions that have evolved within the higher areas of the brain' (*ibid.*). In the former case, cognitive emotions may turn out to be homeostatic extensions of basic emotions in a general overarching emotion natural category, as suggested above. In the latter, they may form a distinct kind of their own which does not figure in a generalized emotion category.

In the end, Panksepp seems to take the same route as Damasio. He appears to favor a unified view of emotion theory that encompasses both basic and cognitive emotions. There are many aspects of his affective neuroscience and his proposed psychobiological synthesis that allude to or incorporate the cognitive dimension of emotion. It is just that neuroscience now cannot tell us much about these higher cognitive dimensions. The key to Panksepp's unified approach lies in the fact that *affective representation* is central to emotion (*ibid.*, pp. 7, 24–40). In that respect, the higher cognitive emotions are no less affective in nature than their basic primitive counterparts (*ibid.*, pp. 315–23). In short, although there may be important differences between basic and cognitive emotions and the neurobiological mechanisms that generate them, both are fundamentally affective representational phenomena. In philosophical terminology, both involve evaluation (Solomon [1976], Lyons [1980]). In psychological terminology, both involve appraisal (Ekman [1992], Lazarus [1991]). Affective representation, then, is the common denominator between basic and cognitive emotions (Charland [1997]). It provides the primary rationale for classifying them both as emotions.<sup>7</sup>

There is another interesting way to address our problem of the status of cognitive emotions and their relation to the natural kind status of basic emotions. A central element of Boyd's theory of natural kinds is the notion of a disciplinary matrix. This Boyd defines as a 'family of inductive and inferential practices united by common conceptual resources, whether or not these correspond to academic or practical disciplines otherwise understood' (Boyd [1999], p. 148). Now according to Boyd, we posit natural kinds relative to a discipline or disciplinary matrix. Natural kinds are required to help carry out the explanatory and inductive projects relative to a specific domain of inquiry. What I want to propose is that emotion theory as it is currently conceived is a disciplinary matrix and that, on a conceptual level, the kind

<sup>7</sup> To be precise, Panksepp argues that the affective realm should be divided into three conceptual categories, which he says reflect three different kinds of processes. These are: 'reflexive affects', 'blue-ribbon, Grade A emotions' and 'higher sentiments' (Panksepp [2000], p. 143). However, he does not suggest that all of these processes are natural kinds. Only emotional processes are granted that status. Ben Ze'ev also provides an insightful discussion of the affective realm that, among other things, distinguishes sentiments from emotions (Ben Ze'ev [2000]).

'emotion' is required for carrying out the explanatory projects and activities in that domain. In other words, the conceptual kind 'emotion' is a necessary and productive conceptual element of emotion theory. We need that term to state hypotheses and capture generalizations, and even to debate the current status of the term itself (Charland [2001]). Under these circumstances, and in light of the above, it is best to provisionally assume that both basic and cognitive emotions will somehow find their place within the general natural kind classification 'emotion'. Griffiths' proposed fragmentation of emotion theory is simply too radical and premature.

## **8 Concluding remarks**

We started this discussion with the observation that many contemporary philosophers deny that emotions form a natural kind. In general, the reason appears to be that emotions are thought to have imprecise boundaries. They do not appear to have a fixed essence that can be captured by necessary and sufficient conditions. Recent developments in the theory of natural kinds offer a new way to understand the natural kind status of emotions. In particular, developments in the causal evolutionary theory of natural kinds allows for kinds that have indistinct boundaries and that do not admit of definition in terms of precise necessary and sufficient conditions. What is central about natural kinds in that sense is the role they play in stating reliable generalizations, and the causal mechanisms in virtue of which they exercise that role. Some emotions can plausibly be argued to be natural kinds in this causal sense. The same is true of 'emotion', the kind formed by those basic emotions. Only there, both the generalizations and causal mechanisms involved function at a higher level of abstraction. Using this causal sense of a natural kind it is possible to argue that emotion is a neurobiological kind. It also possible to argue that, on a different level, emotions also form a psychological kind. The two domains are not exactly coextensive, since many complex social emotions in the psychological domain have no counterparts in the basic mammalian emotions of the neurobiological arena. However, at least in the case of basic emotions, there is some reason to be optimistic about the prospects for consilience. This is an area where both neurobiology and psychology appear to converge on the existence of a few central affective posits and the emoters whose lives they enrich.

The plausibility of the hypothesis that emotion is a natural kind has at least two important consequences. First, it means that there may be a legitimate and very real empirical basis for the distinction between emotion and reason. In a recent essay on the philosophy of emotions, Robert Solomon describes the historical origins of that distinction in Western thought. He says it is 'as if we were dealing with two different natural kinds, two conflicting and

antagonistic aspects of the soul' (Solomon [2000], p. 3). It should now be clear that there are good reasons to think that *both* emotion and cognition may be natural kinds. But if so, how are they related? The issues now appear to be largely empirical. Some researchers argue that there are respects in which emotion and reason complement each other nicely (Damasio [1994]). That certainly leaves room for tension and conflict in other matters. Whatever the case may be, it appears that there are two distinct systems of brain organization involved. Panksepp refers to these as the 'somatic-cognitive nervous system', which he locates on the thalamic neocortical axis, and the 'visceral-emotional nervous system', which he locates on the hypothalamic-limbic axis (Panksepp [1998], p. 62, Fig. 4.1). He says that the former system corresponds to the 'stream of thought', while the latter corresponds to the 'stream of feeling' (*ibid.*, p. 62). Thus he appears to believe that affect and cognition are distinct natural kinds within the brain (*ibid.*, pp. 318–19).<sup>8</sup> If so, then we are divided selves, as folk psychology suggests. But how we are divided is a long and complicated story that only science can tell.

A second important consequence of the plausibility of the hypotheses that emotion is a natural kind for the study of mind is that it lends unity and direction to the study of individual emotions. There is now a legitimate reason for lumping emotions together, as evidently so many do. In a famous passage in the *Principles of Psychology*, William James describes his frustration and lassitude with the study of emotions of his day. He writes:

[A]s far as the 'scientific psychology' of the emotions goes [...] I should as lief read verbal descriptions of the shapes of the rocks on a New Hampshire farm as toil through them again. They give one nowhere a central point of view, or a deductive general principle. They distinguish and refine and specify ad infinitum without ever getting to another logical level. Whereas the beauty of all truly scientific work is to get to ever deeper levels. (James [1890/1950], pp. 448–9)

James's hope was that the study of emotion would move beyond the view of them as 'immutable species' and 'eternal and sacred entities' (*ibid.*, p. 449). He urged his contemporaries to look for 'more general causes' and vigorously argued for a *rapprochement* between the psychology of emotion and Darwin's evolutionary ideas (*ibid.*). He believed that the fundamental nature of emotions should be sought in their physiology.<sup>9</sup> Much of modern day philosophy and psychology of emotion consists in the sort of cataloguing of

<sup>8</sup> In a similar vein, Ben Ze'ev argues that emotion and intellect are two different psychological systems (Ben Ze'ev [2000], pp. 175–81). Joseph LeDoux ([1989]) also argues for the view that cognition and emotion are separate systems.

<sup>9</sup> James also argued that there are 'no special brain centres for emotion' (James [1890], pp. 472–4). This has not prevented numerous emotion theorists who believe that there are specific brain mechanisms devoted to emotion from making extensive use of his work (Damasio [1994], LeDoux [1996]).

emotions James describes, and sometimes it does seem as if ‘all that can be done with them is reverently to catalogue their separate characters, points, and effects’ (*ibid.*). However, we do not have to be as dismissive of these efforts as James was. The good news is that now that approach can flourish in a much richer, more theoretically unified climate. There is a new and vigorous wind to fill and guide the sails of emotion theory. This is the hypothesis that emotion is a natural kind.

### Acknowledgements

I would like to thank Evan Thompson and Andrew Wayne for helpful comments on earlier drafts of this paper.

*Department of Philosophy and Faculty of Health Sciences  
University of Western Ontario  
London, Ontario N6A 3K7  
Canada  
charland@uwo.ca*

### References

- Arnold, M. [1960]: *Emotion and Personality*, New York: Columbia University Press.
- Ben-Ze'ev, A. [2000]: *The Subtlety of Emotions*, Cambridge, MA: MIT Press.
- Boyd, R. [1991]: ‘Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds’, *Philosophical Studies*, **61**, pp. 127–48.
- Boyd, R. [1999]: ‘Homeostasis, Species, and Higher Taxa’, in R. A. Wilson (ed.), *Species: New Interdisciplinary Essays*, Cambridge, MA: MIT Press, pp. 141–87.
- Butts, R. E. (ed.) [1968]: *William Whewell’s Theory of Scientific Method*, Pittsburgh: University of Pittsburgh Press.
- Charland, L. C. [1995a]: ‘Emotion as a Natural Kind: Towards a Computational Foundation for Emotion Theory’, *Philosophical Psychology*, **8** (1), pp. 59–84.
- Charland, L. C. [1995b]: ‘Feeling and Representing: Computational Theory and the Modularity of Affect’, *Synthese*, **105**, pp. 273–301.
- Charland, L. C. [1997]: ‘Reconciling Cognitive and Perceptual Theories of Emotion: A Representational Proposal’, *Philosophy of Science*, **64**, pp. 555–79.
- Charland, L. C. [2001]: ‘In Defence of “Emotion”’, Critical Notice of Paul E. Griffiths, ‘What Emotions Really Are: The Problem of Psychological Categories’, *Canadian Journal of Philosophy*, **31** (1), pp. 133–54.
- Churchland, P. M. [1982]: ‘Is “Thinker” a Natural Kind?’, *Dialogue*, **21**, pp. 223–8.
- Damasio, A. [1994]: *Descartes’ Error*, New York: Putnam & Grosset.
- Darwin C. [1872/1998]: *The Expression of Emotion in Man and Animal*, Oxford: Oxford University Press.
- de Sousa, R. [1987]: *The Rationality of Emotion*, Cambridge, MA: MIT Press.

- de Sousa, R. [1989]: 'Kinds of Kinds', *International Studies in the Philosophy of Science*, **3** (2), pp. 119–35.
- de Sousa, R. [1999]: 'Review of Paul E. Griffiths, 'What Emotions Really Are: The Problem of Psychological Categories'', *Dialogue*, **38** (4), pp. 908–11.
- Ekman, P. [1992]: 'An Argument for Basic Emotions', *Cognition and Emotion*, **6**, pp. 169–200.
- Elster, J. [1999]: *Strong Feelings*, Cambridge, MA: MIT Press.
- Elster, J. [1998]: *Strong Feelings: Emotion, Addiction, and Human Behavior*, Cambridge, MA: MIT Press.
- Fodor, J. [1975]: *The Language of Thought*, New York: Thomas Y. Crowell.
- Frijda, N. [1988]: 'The Laws of Emotion', *American Psychologist*, **43**, pp. 349–58.
- Gordon, R. M. [1987]: *The Cognitive Structure of Emotion*, Cambridge, MA: Cambridge University Press.
- Griffiths, P. E. [1997]: *What Emotions Really Are: The Problem of Psychological Categories*, Chicago: University of Chicago Press.
- Griffiths, P. E. [1999]: 'Squaring the Circle: Natural Kinds With Historical Essences', in R. A. Wilson (ed.), *Species: New Interdisciplinary Essays*, Cambridge, MA: MIT Press, pp. 209–28.
- Hacking, I. [1991a]: 'A Tradition in Natural Kinds', *Philosophical Studies*, **61**, pp. 109–26.
- Hacking, I. [1991b]: 'On Boyd', *Philosophical Studies*, **61**, pp. 149–54.
- Hacking, I. [2000]: *The Social Construction of What*, Cambridge, MA: Harvard University Press.
- Hebb, D. [1946]: 'Emotion in Man and Animal: An Analysis of the Intuitive Process of Recognition', *Psychological Review*, **53**, pp. 88–106.
- James, W. [1890/1950]: *Principles of Psychology*, New York: Dover Press.
- Lazarus, R. [1991]: *Emotion and Adaptation*, Oxford: Oxford University Press.
- LeDoux, J. [1996]: *The Emotional Brain*, New York: Simon & Shuster.
- LeDoux, J. [1989]: 'Cognitive-Emotional Interactions in the Brain', *Cognition and Emotion*, **3** (4), pp. 267–89.
- Lewis, M. and Haviland-Jones, J. M. (eds) [2000]: *Handbook of Emotions*, 2nd edn, New York: Guilford Press.
- Lyons, W. [1980]: *Emotion*, Cambridge: Cambridge University Press.
- Marr, D. [1982]: *Vision*, San Francisco, CA: W. H. Freeman.
- Millikan, R. G. [1999]: 'Historical Kinds and the "Special Sciences"', *Philosophical Studies*, **95** (1–2), pp. 45–65.
- Neu, J. [2000]: *A Tear is an Intellectual Thing: The Meanings of Emotion*, Oxford: Oxford University Press.
- Panksepp, J. [1998]: *Affective Neuroscience*, Oxford: Oxford University Press.
- Panksepp, J. [2000]: 'Emotions as Natural Kinds Within the Brain', in M. Lewis and J. M. Haviland-Jones (eds), *Handbook of Emotions*, 2nd edn, New York: Guilford Press, pp. 137–55.
- Pylyshyn, Z. [1984]: *Computation and Cognition*, Cambridge, MA: MIT Press.
- Rorty, A. [1978]: 'Explaining Emotions', *The Journal of Philosophy*, **75**, 139–61.

- Rorty, A. [1982]: 'From Passions to Emotions and Sentiments', *Philosophy*, **57**, 159–72.
- Solomon, R. [1976]: *The Passions*, New York: Doubleday.
- Solomon, R. [1995]: 'Some Notes on Emotion, "East and West"', *Philosophy East and West*, **45** (2), pp. 171–202.
- Solomon, R. [2000]: 'The Philosophy of Emotions', in M. Lewis and J. M. Haviland-Jones (eds.), *Handbook of Emotions*, 2nd edn, New York: Guilford Press, pp. 3–15.
- Wilson, R. A. [1999]: 'Realism, Essence, and Kind: Resuscitating Species Essentialism', in R. A. Wilson (ed.), *Species: New Interdisciplinary Essays*, Cambridge, MA: MIT Press, pp. 187–209.