

Is Emotion a Natural Kind?*

Paul E. Griffiths,
Department of History and Philosophy of Science,
University of Pittsburgh,
Pittsburgh,
PA 15232,
USA

pauleg@pitt.edu

7280wds plus references

1. Introduction

In *What Emotions Really Are: The problem of psychological categories* (Griffiths, 1997) I argued that it is unlikely that all the psychological states and processes that fall under the vernacular category of emotion are sufficiently similar to one another to allow a unified scientific psychology of the emotions. The psychological, neuroscientific and biological theories that best explain any particular subset of human emotions will not adequately explain all human emotions. In a slogan, emotions are not a natural kind (pp. 14-17; 241-247)¹. I argued that the same is probably true of many specific emotion categories, such as anger and love (p. 17). On some occasions when a person is properly said to be angry, certain psychological, neuroscientific and biological theories will adequately explain what is happening to that person. On other occasions of anger, however, different theories will be needed. I described my position as eliminativism about emotion, because it implies that the term 'emotion' and some specific emotion terms like 'anger' are examples of what philosophers of language have called 'partial reference' (p. 242). The term 'jade' is the classic example of partial reference. The term 'jade' is used as if it referred to a particular kind of mineral, in the same manner as 'malachite' or 'diamond'. In reality, however, the term covers two different stones, jadeite or nephrite. The term 'jade' partially refers to each of these two minerals. Hence, for the purposes of geology or chemistry, jade cannot be treated as a single kind of thing. The properties of the two substances have to be investigated separately, their geological origins explained separately and their abundance in unknown geological deposits predicted separately. Likewise, I argued, the sciences of the mind will have to develop separate theories of the various different kinds of emotion and also of the various different kinds of some particular emotions. In the same sense that there is really no such thing as jade, only jadeite and nephrite, there is no such thing as emotion, only 'affect programs', 'socially sustained pretences' and other more specific categories of psychological state and process.

While the critical response to my book has been generally positiveⁱⁱ, many philosophers remain unconvinced. I suspect that this is to a significant extent because the question of whether emotion is a natural kind has been conflated with the question of whether the concept of emotion can be given a univocal analysis. These two questions have very little to do with one another. The concept of a superlunary object can be analyzed — something is superlunary *iff* it is outside the orbit of the moon — but the consensus since Galileo has been that those things do not form a natural kind. Likewise, my claim was not that the vernacular concept of emotion is unanalyzable — I myself suggested a prototype analysis (pp. 242-245). My claim was that the things that fall under the concept do not constitute a distinct kind for the purposes of scientific enquiry. The concept of ‘vitamin’ is a useful comparison. Vitamins are not, as was once thought, ‘vital amines’ but a diverse group of chemicals with diverse roles in physiology sharing the feature that humans cannot synthesize them, or can synthesize them, as with Vitamin D, only under advantageous environmental conditions. Their absence leads to ‘deficiency diseases’ with diverse etiologies and diverse prognoses. So the concept of a vitamin’ can be analysed, and individual vitamins and even some groups of vitamins are natural kinds, but ‘vitamin’ itself is a superficial descriptive category. It is not a sensible scientific project to investigate the nature of vitamins *in general*. The question ‘what is a vitamin?’ is best answered by describing the main kinds of vitamin and how different they are from one another.

Another important line of reply has been that it is a mistake to ask whether emotion and emotions are natural kinds at all, since they are primarily normative kinds (Doris, 2001). I have addressed this issue at length elsewhere (Griffiths, In Press-b). I wholeheartedly agree that the normative aspects of the semantics of emotion concepts are vital to an adequate account of those concepts, a view I endorsed in my book (pp. 7; 196-201). However, I do not accept that the normative uses of emotion concepts can be cleanly separated from the descriptive uses. Hence, the fact that emotion concepts are normative concepts does not mean that they are insulated from revision in the light of empirical discoveries about emotion

In the remainder of this section I restate what I mean by ‘natural kind’ and my argument for supposing that emotion is not a natural kind in this sense. In the following sections I discuss the two most promising proposals to reunify the emotion category: the revival of the Jamesian theory of emotion associated with the writings of Antonio Damasio and a philosophical approach to the content of emotional representations that draws on ‘multi-level appraisal theory’ in psychology.

1.2 What do I mean by ‘natural kind’?

I use the traditional term ‘natural kind’ to denote categories which admit of reliable extrapolation from samples of the category to the whole category. In other words, natural kinds are categories about which we can make scientific discoveries. In my book I built on the work of several other philosophers and scientists to construct an

account of natural kinds in psychology and biology, an account further elaborated in (Griffiths, 1999, 2001a) and briefly sketched here.

The fundamental scientific practices of induction and explanation presume that some of the observable correlations between properties are ‘projectable’ (Goodman, 1954). That is, correlations observed in a set of samples can be reliably ‘projected’ to other instances of the category. Scientific classifications of particulars into categories embody our current understanding of where such projectable clusters of properties are to be found. The species category, for instance, classifies particular organisms into sets that represent reliable clusters of morphological, physiological and behavioral properties. Hence, these properties of the species as a whole can be discovered by studying a few members of the species.

The traditional requirement that natural kinds be the subjects of universal, exceptionless ‘laws of nature’ is too strong and would leave few natural kinds in the biological and social sciences where generalizations are often exception-ridden or only locally valid. Fortunately, it is easy to generalize the idea of a law of nature to the broader idea that statements are to varying degrees ‘lawlike’ (have counterfactual force). This broader conception of a lawlike generalization allows a broader definition of a natural kind. A category is (minimally) natural if it is possible to make better than chance predictions about the properties of its instances. This, of course, is a very weak condition. Very many ways of classifying the world are minimally natural. The aim is to find categories that are a great deal more than minimally natural. Ideally, a natural kind should allow *very reliable* predictions in a *large domain* of properties. The classic examples of natural kinds, such as chemical elements and biological species, have these desirable features.

It is important to note that categories are natural only relative to specific domain(s) of properties to which they are connected by background theories. The category of domestic pets is not a natural category for investigating morphology, physiology or behavior, but might be a natural category in some social psychological theory or, of course, in a theory about domestication. Emotion, I argue, is not a natural kind relative to the domains of properties that are the focus of investigation in psychology and the neurosciences. It is not the case that the psychological states and processes encompassed by the vernacular category of emotion form a category which allows extrapolation of psychological and neuroscientific findings about a sample of emotions to other emotions in a large enough domain of properties and with enough reliability to make emotion comparable to categories in other mature areas of the life sciences, such as biological systematics or the more robust parts of nosology.

1.3 Why emotion is not a natural kind in this sense

On some occasions when a person ‘has an emotion’ they are producing an affect program response - a ‘basic emotion’ in Paul Ekman’s sense. The response is short lived, highly automated, triggered in the early stages of processing perceptual

information, and realized in anatomically ancient brain structures that we share with many other vertebrates. It is found in all human cultures and closely related to responses in other primates. Suppose, for example, that you are waiting in line outside a nightclub. After twenty minutes, someone unexpectedly pokes you sharply in the small of the back. You spin around, making a threat expression, probably the 'square-mouthed' variety, your body adjusts physiologically for violent action and your attention is entirely on your assailant. If the situation is rapidly defused (you are male and an attractive young woman has tripped against you and is smiling apologetically) then this will be a pure case of affect program anger. On other occasions, however, a person 'having an emotion' is responding in a more cognitively complex way to more highly analyzed information. The episode may or may not involve the occurrence of one or more affect program responses. Suppose, for example, that you are locked into a dysfunctional pattern of interaction with your spouse involving continual fault finding and put-downs, this pattern emerging without any intention from the particular patterns of relationship management you both bring to the marriage. The pattern has resisted your occasional attempts to consciously improve your behavior and, as you reflect one day on what appears to be the inevitable degeneration of the relationship, you experience a deep sense of guilt and self-loathing. I referred to such states as 'higher cognitive emotions'. Finally, on some occasions, 'having an emotion' may centrally involve an internalized cultural model of appropriate behavior. In my book, I suggested that people who respond to losing their job by 'going postal' - going on a killing spree followed by attempted suicide - could be simultaneously 'out of control' and following a 'script' derived from real and fictional incidents that are culturally salient for them. I presented a tentative analysis of such cases as 'socially sustained pretences'.

My argument that emotion is not a natural kind rested mainly on the first two cases, the affect programs, or 'basic emotions' and the more cognitively complex emotions, which I termed 'higher cognitive emotions'. I regret using the latter phrase, as it gives the impression that I have a substantive theory of those emotional responses. In fact, I discussed the views of leading theorists like Antonio Damasio (pp. 102-106) and Robert Frank (pp. 117-122) whose ideas clearly do something to illuminate these more complex emotions, but my only firm conclusion was that these states and processes are unlikely to be reduced to the basic emotions or understood as blends or elaborations of them. I accepted, however, that these other emotions may involve basic emotions as parts, depend on basic emotions for their development in the child, and interact with basic emotions in typical ways in real-life situations. To better indicate how little we know about non-basic emotions, I now prefer to call them 'complex emotions' (Griffiths, In Press-a, In Press-c). Finally, I rejected the view that the basic emotions are not emotions or that they are merely proto-emotions (pp. 26,164). This sort of procrustean treatment is inevitable as long as we insist that emotions are a single kind of thing. Instead, I suggested we should accept that there are two or more different kinds of psychological process involved in emotion, and if research into complex emotions suggests that, like basic

emotions, they can be classified into emotion types, then there are two or more different kinds of emotions.ⁱⁱⁱ

The other major element of my case for ‘eliminating emotion’ was my specific account of natural kinds in biology. I have defended that account in numerous places (Griffiths, 1994, 1996a, 1996b, 1997, 1999, 2001a, 2001b) and can only briefly sketch it here. Evolution leads to the existence of two fundamental sets of biological categories - homologies and analogies. A homologue is “The same organ in different animals under every variety of form and function.” (Owen, 1843: 374), a definition interpreted since Darwin to mean that these organs are descended from a common ancestral form. Analogies are cases where two unrelated structures resemble one another because natural selection has adapted them for the same ecological role. My hips are homologous to those of a horse, but they are analogous to the articulation of an arthropod limb-pair with the rest of the arthropod segment. These two concepts have been routinely applied to behavior since the rise of ethology in the 1920s.

I argued that some basic emotions are analogous to complex emotions that fall under the same vernacular category. The different kinds of fear, for example, are all responses to danger. Any psychobiological theory of emotions *in general*, I suggested, would have to be a theory of psychological analogies — traits that fulfill the same functions in relation to the environment. The categories that would be generated by such a theory, although they might enter into useful ecological generalizations, would be systematically unsuited to the distinctive purposes of psychology and neuroscience. They would support induction and explanation on the wrong domain of properties (pp. 230-241). Suppose that two animals have homologous psychological traits, such as the basic emotion of fear in humans and fear in chimpanzees. We can predict that, even if the function of fear has been subtly altered by the different meaning of ‘danger’ for humans and for chimps, the computational methods used to process danger-related information will be very similar and the neural structures that implement them will be very similar indeed. After all, Joseph LeDoux’s widely accepted account of fear processing in the human brain is largely, and legitimately, based on the study of far more distantly homologous processes in the rat (LeDoux, 1996). Now suppose that two animals have psychological traits that are analogous - fear in the rat and fear in the octopus, for example. It is a truism in comparative biology that similarities due to analogy (shared adaptive function) are ‘shallow’. The same problem can be solved in different ways and so the deeper you dig the more likely it is that mechanisms will diverge. Bat wings and bird wings, for example, have similar aerodynamic properties but their internal structure is radically different, reflecting their different ancestries. In contrast, similarities due to homology (shared ancestry) are notoriously ‘deep’: even when function has been transformed, the deeper you dig the more similarity there is in the underlying mechanisms. Threat displays in chimps look very different from anger in humans, but when their superficial appearance is analyzed to reveal the specific muscles whose movement produces the expression and the order in which those muscles move, it becomes clear that

they are homologues of one another. The same is almost certainly true of the neural mechanisms that control those movements.

Now, psychology is in the business of uncovering the mechanisms that produce behavior. This is even more evident in the case of neuropsychology. Hence these disciplines seek categories that are heuristically valuable for the study of underlying mechanisms. Inferences to shared mechanism based on homology are not 100% reliable, but they are reliable enough to build good science with - the criterion I outlined above - and they are more reliable than inferences to shared mechanism based on analogy. I concluded that replacing the categories of basic emotion - which are explicitly categories of psychobiological homology - with more general categories that included any mechanism that performs the same broad adaptive function - would reduce the inductive and explanatory power of cognitive psychology and the neurosciences. It would be a move from a more natural category to a less natural category in the sense defined above. The aims of these sciences are better served by recognizing that there is more than one kind of emotion and by investigating each on its own terms.

Having summarized my arguments or the view that emotions do not form a natural kind, I now go on to consider the most promising proposals for the opposite view.

2. Affective Neuroscience and the Category of Emotion

2.1. Are complex emotions blends of basic emotions?

Innumerable theorists have suggested that complex emotions are blends of basic emotions. In my book I argued that some of the features that characterize complex emotions cannot be accounted for merely by supposing that several basic emotions occur simultaneously or in sequence. First, it is characteristic of some complex emotions that they occur in response to complex properties of the stimulus situation and so need a more sophisticated appraisal of the environment than would be obtained by adding together the appraisal criteria for the basic emotions: “Situations that elicit sexual jealousy or moral indignation do not differ from each other merely in the proportions of anger, conspecific challenge, noxiousness and loss that they involve” (Griffiths, 1997: 102). Secondly, some complex emotions endure much longer than the basic emotions. What is more, they endure as real psychological processes, not mere dispositions. When a woman’s feeling of guilt explains her behavior through a long session of negotiation with her husband and their lawyers, it does more than dispose her to intermittently display affect-program sadness and affect program fear. Thirdly, basic emotions have behavioral consequences of the sort detected by the facial affect coding system. I denied that all complex emotions result in blends of the facial action associated with the known basic emotions. Finally, while basic emotions have reciprocal interactions with more complex cognitive processes, some complex emotions are more directly involved in the control of long-term, planned action. The woman’s guilt in the

example just given is arguably an intimate part of the thought processes by which she arrives at a decision on which demands to concede and which to resist.

The proposal that complex emotions involve blends of basic emotions is more promising when conjoined to the idea that complex emotions involve additional cognitive activity. The idea that complex emotions are elaborations of basic emotions resulting from the integration of activity in phylogenetically ancient brain structures with activity in the neocortex is currently the most popular proposal to reintegrate the domain of emotion. This is largely because of the work of the neuroscientist Antonio Damasio (Damasio, 1994; Damasio, 1999). Jesse Prinz has captured the spirit of this approach with the analogy that basic emotions are shots of hard liquor and complex emotions are cocktails in which specific hard liquors are mixed with specific non-alcoholic ingredients (Prinz, 2000). The basic emotion is the motivational ‘kick’ in each complex emotional cocktail.

2.2. *Emotion and affective neuroscience*

The ‘cognitive elaboration’ view is a promising approach to complex emotions, provided it is borne in mind that basic emotions are themselves emotions and not *only* constituents of complex emotion, and that basic emotions and complex emotions are very different from one another. It is all too easy to gloss over these differences. Simon Blackburn, for example, has argued that Damasio’s work refutes my claim that emotion is not a natural kind because,

“Empirically, the suggestion that we split the operation of the affect program from ‘higher cognitive emotion’ seems to ignore the most fascinating result of Damasio’s work, which is the extent to which ‘higher-order’ decision making has to harness the limbic system in order to work at all.” (Blackburn, 1998, 129).

But, empirically, the operation of two kinds of emotions *can* be split, whether this occurs in pathologies like phobias or in the phenomena reported in the literature on ‘affective primacy’ (Öhman, 1999, 2002; Zajonc, 1980). Basic emotions can occur without the accompanying activity in the neo-cortex and basic emotion appraisal processes can reach conclusions contradictory to those reached by full-blown cognitive appraisal of the stimulus situation. Furthermore, there is a longstanding consensus in the literature on the evolution of emotion that this is why humans still have basic emotions. Basic emotions are rapid acting, failsafe devices that produce evolved behavioral, physiological *and cognitive* responses tailored to certain critical features of the environment. They are faster and more reliable than the slower, but arguably more accurate responses that we make using our recently evolved neocortical resources. This, of course, is entirely consistent with Damasio’s claims about the role of the emotions in rational decision-making. That role makes use of one of the outputs of the basic emotional response, which, Damasio argues, is a representation of the physiological changes that have been produced by that initial emotional response. Other outputs to cognition from the basic emotion process

include orienting sensory systems to the emotional stimulus and allocating attentional resources to that stimulus.

I suspect Blackburn is less interested in whether basic emotions can occur without complex emotions, than in whether complex emotions can occur without basic emotions and thus whether all emotions share a set of core processes. In fact, according to philosopher Louis Charland, a fair-minded reading of contemporary affective neuroscience makes it quite clear that all emotions do, indeed, share a single kind of core process. Affective neuroscience, he argues, provides a general theory of the nature of 'emoters', or 'affective systems'. These are, "a distinct class of biological systems whose behavior is largely governed by emotion and only explainable in those terms. This is the neurobiological version of the hypothesis that emotion is a natural kind term." (Charland, 2001: 151-2). Charland is referring in part to the idea elaborated by the neuroscientist Paul D. Maclean in a series of publications from the 1950s to the 1980s that emotion is a distinctive form of information processing which we have inherited from our distant evolutionary ancestors, and which manifests itself to us as what we call emotion 'feelings'. These processes are realized in phylogenetically ancient anatomical regions surrounding the brainstem, regions that Maclean termed the 'limbic brain'. Charland notes that the leading contemporary neuroscientist of emotion Joseph LeDoux regards the limbic brain concept as more or less anatomically and functionally meaningless, but he sets against this the fact that LeDoux's experimental work has dealt only with one emotion - fear - and that his views are not shared by all other neuroscientists. The views of another leading neuroscientist of emotion, Jaak Panksepp, are more congenial to Charland. Panksepp accepts that MacLean's concept of an 'emotional brain' is oversimplified, but defends the underlying concept that emotion represents an ancient form of information processing that we share with many other species. In Charland's view, the work of Panksepp, Damasio and others suggests very strongly that there is a single, natural kind of psychological process that generates affect. He takes it to be obvious that we should identify the category of emotion with this class of 'affective' processes.

I am sympathetic to the view that the basic emotions represent a distinctive form of information processing that humans share with many other animals (Griffiths, 1997: 91-97) but I have argued that identifying the emotions with that form of processing would amount to a substantive and stipulative revision of the vernacular concept of emotion (pp. 230-234). This revision would both exclude some things that are in the vernacular category and include some things that are not in the vernacular category. In my book I was concerned to leave room for the possibility that some complex emotions may not involve basic emotions, or may involve them only peripherally^{iv}. I still think it is important not to foreclose that issue, but here I will concentrate on the opposite problem, namely that the new category is too inclusive to be simply identified with the vernacular category. There is, of course, nothing wrong with scientists deciding to use the term 'emotion' in a revised sense, and most leading emotion neuroscientists seem to be aware that they are in the

business of productive stipulation, not conceptual analysis. In his own response to my claim that emotion is not a natural kind, Damasio remarks:

“At this point, my preference is to retain the traditional nomenclature, clarify the use of the terms, and wait until further evidence dictates a new classification, my hope being that by maintaining some continuity we will facilitate communication at this transitional stage. I will talk about three levels of emotion - background, primary and secondary. This is revolutionary enough, given that background emotions are not part of the usual roster of emotions.” (Damasio, 1999, 341)

Thus, rather than vindicating the vernacular category of emotion Damasio is recruiting the old term for a new and more general category. This category covers phylogenetically ancient ‘primary emotions’ (basic emotions), ‘secondary emotions’ that are elaborations of these ancient responses into complex and variegated responses that involve substantial areas of the neo-cortex and uniquely human cognitive abilities (similar to my ‘complex emotions’), and background emotions, which are the ever-present awareness of our own body.

This perspective on how affective neuroscience uses the emotion category is reinforced by the way in which the discipline treats basic drives like hunger and thirst, and hedonic states like pain and pleasure. Panksepp remarks that he will follow tradition in distinguishing between emotions proper and drives which regulate some specific state of the body, and thus that he will not initially consider hunger, thirst or disgust to be emotions. But this distinction, he remarks, will become ‘less defensible’ in later chapters of his book (Panksepp, 1998, 47, 342 note 17). In the same vein, Damasio remarks that, “I will refer to drives and motivations and pain and pleasure as triggers or constituents of emotions, but not as emotions in the proper sense. No doubt all these devices are intended to regulate life, but it is arguable that emotions are more complex...” (p. 341). These remarks exemplify a research strategy that I take to be central to much recent work on the emotions. Panksepp and Damasio take the domain of affective neuroscience to be all mental processes that involve affective feelings. Panksepp has described this as ‘the basic psychological criteria that emotional systems should be capable of elaborating subjective feeling states that are affectively valenced’ (Panksepp, 1998, 48). But this domain is both much broader than the vernacular category of emotion and somewhat difficult to work with in practice, because subjective feeling states ‘have so far defied neural specification’ (p. 48). In practice, then, emotions are defined as that class of affective processes that have certain distinctive performance features:

“(1) Various sensory stimuli can unconditionally access emotional systems; (2) Emotional systems can generate instinctual motor outputs, as well as; (3) modulate sensory inputs. (4) Emotional systems have positive feedback components that can sustain emotional arousal after precipitating events have passed. (5) These systems can be modulated by cognitive inputs and (6) can modify and channel cognitive activities.” (Panksepp, 1998: 48)

A comparison of these criteria with descriptions of the distinctive features of affect programs or basic emotions in the Tomkins-Izard-Ekman tradition reveals striking similarities (Griffiths, 1997, 77-99). In both research traditions, emotion is being conceived as information processing that is somehow simpler than paradigm examples of cognition, perhaps because it involves bodily feelings rather than explicit representations of external states of affairs, but which is nevertheless more complex than tropisms, reflexes and homeostatic drives. In my book I used Paul Ekman's argument that the startle reaction is not an emotion as a paradigm of what it takes to establish that some reaction is a basic emotion (Ekman, Friesen, & Simons, 1985). Ekman and his collaborators argue that the startle reaction is too reflex-like to count as an affect program and I commented that, "This suggests that the information processing arrangements underlying startle are not of the same sort as those underlying affect programs. ... Extending the concept of an affect program state to cover it would not be a positive step in theory construction, since findings about startle may not be true of the affect program states and vice-versa." (Griffiths, 1997, 241; see also Robinson, 1995).

In my view, what Charland fails to appreciate is that the category of felt affective states is very large, and certainly larger than the existing vernacular category of emotion. In addition to classic emotions like anger and joy, it will include drive states like hunger and thirst, hedonic states like pain and pleasure, and the ubiquitous awareness of bodily states that Damasio calls 'background emotion'. Furthermore, given the strong connection between motivation and affect made by writers like Damasio, the category of felt affective states probably includes desires and preferences. Blackburn, for example, has argued that the motivational power of our long-term goals reflects an emotional commitment to those goals and cites with approval Damasio's idea that our awareness of the body is the motive power of practical reason (Blackburn, 1998, 129). Because the category of felt affective states is so broad, it is natural within this framework to seek distinctive *kinds* of processes involving affective feelings. Hence, I suggest that rather than demonstrating that emotions are a natural kind, the empirical success of current approaches to affective neuroscience would establish that there is a very large domain of affective and motivational phenomena, within which we could distinguish categories of state and process such as Damasio's background emotions, primary emotions and secondary emotions, homeostatic drives, and pure hedonic states like pain and pleasure. The scientific domain of affective neuroscience would be like the scientific domain of chemistry, where atoms and their bonds are at the basis of everything, but where lumping together mixtures, compounds, alloys, pure elements and pure isotopes on the grounds that they are all 'chemical substances' is not very helpful. Making and exploring those distinctions was essential to the development of modern chemistry and I suggest that the same will be true of affective neuroscience.

3. Multi-level Appraisal Theory and the Content of Emotional Representations

3.1 Appraisal and the philosophy of emotion

Appraisal theories are the closest scientific equivalents of the theories that have dominated philosophy of emotion since the 1960s. Like appraisal theorists, philosophers have sought to understand emotions in terms of the situations that elicit them. Emotions have been analyzed as states with specific ‘formal objects’ (Kenny, 1963), as evaluative judgments (Solomon, 1976), as evaluative judgments that cause bodily arousal (Lyons, 1980), as feelings of comfort or discomfort directed towards an evaluative thought (Greenspan, 1988) and as the results of either true belief or uncertainty about emotion-inducing situations (Gordon, 1987). For these and many other authors, the central aim of a philosophical theory of emotion is to identify the *content* of an emotion: the actual or imagined state of affairs in the world that corresponds to that emotion.

Appraisal theories aim to identify the features of an emotion-eliciting situation that lead to the production of one emotion rather than another (Scherer, 1999). Typical appraisal theories include a set of dimensions against which potentially emotion-eliciting situations are assessed. The dimensions of the emotion hyperspace might include, for example, whether the eliciting situation fulfills or frustrates the subject’s goals or whether an actor in the eliciting situation has violated a norm. Richard Lazarus’s well-known model of emotional appraisal has six dimensions, and the regions of the resulting hyperspace that correspond to particular emotions are summarized by Lazarus as the ‘core relational themes’ of those emotions. Anger, for example, is elicited by the core relational theme ‘a demeaning offence against me and mine’, sadness by ‘having experienced an irrevocable loss’, and guilt by ‘having transgressed a moral imperative’ (Lazarus, 1991). These themes correspond to each emotion’s ‘content’ in philosophical theories of emotion.

The ongoing effort to test appraisal theories has produced a consensus that actual emotional responses do not walk in step with subjects’ evaluation of stimulus situations unless the notion of ‘cognitive evaluation’ is broadened to include sub-personal processes (Teasdale, 1999). Many appraisal theorists have come to accept that even such apparently conceptually complex appraisals as Lazarus’s core relational themes can be assessed: 1. Without the information evaluated being available to other cognitive processes, 2. Before perceptual processing of the stimulus has been completed, and 3. Using only simple, sensory cues to define where the eliciting situation falls on the dimensions. These conclusions are congruent with both the older literature on ‘affective primacy’ (Öhman, 1999, 2002; Zajonc, 1980) and with the recent mapping of multiple neural pathways to the same emotional response (LeDoux, 1996; LeDoux, 1993).

One of the reasons philosophers remain convinced that emotion is a natural kind is their belief that, however psychologically different the various instances of anger or any other emotion may be, every instance shares the same or similar *content*. Moreover, the general category of emotion is unified by the fact that all emotions

have a certain broad kind of *evaluative content*. Thus, to take a prominent recent example, Martha Nussbaum argues that emotions are all of one kind because they are all ‘intelligent responses to the perception of value’ (Nussbaum, 2001: 1). In defending this view she draws on psychologist Richard Lazarus’s classical presentation of appraisal theory, which she describes as ‘in all essentials the view of emotions I have defended in Chapter 1’ (Nussbaum, 2001: 109). Nussbaum’s treatment of the emotions in children and animals also makes use of something like multi-level appraisal. Animals as well as humans make the evaluative judgments that constitute emotions according to Nussbaum’s theory, but they do so without self-conscious awareness and in such a way that the content of their judgments cannot be rendered in language without distortion. Nevertheless, she argues, emotion remains primarily an intentional phenomenon. Despite the existence of low-level appraisal that cannot be expressed in language, ‘...emotions include in their content judgments that can be true or false, and good or bad guides to ethical choice’ (Nussbaum, 2001: 1).

“What we need, in short, is a multifaceted notion of cognitive interpretation or seeing-as, accompanied by a flexible notion of intentionality that allows us to ascribe to a creature more or less precise, vaguer or more demarcated, ways of intending an object and marking it as salient.” (Nussbaum, 2001: 129)

Nussbaum’s description of what is needed for an adequate account of emotional cognition is compelling, but levels of appraisal do not just differ *between* organisms. They also differ *within* a single organism. Multi-level models of emotional appraisal suggest that the same stimulus can be represented in several places in the human brain by different representations. Hence it is vital to understand not only what these multiple appraisals have in common, but also how they *differ* and how they *interact*. The existence of multiple representations in a ‘hierarchical’ emotional architecture (Delancey, 2001) violates a key assumption of most philosophical reasoning about emotion, which is that emotional cognition manipulates emotional representations on the basis of their content, and thus that emotional processes can be explored via the semantic ‘logic’ of emotions:

‘all emotions presuppose or have as their preconditions, certain sorts of cognitions - an awareness of danger in fear, recognition of an offense in anger, appreciation of someone or something as lovable in love. Even the most hard-headed neurological or behavioral theory must take account of the fact that no matter what the neurology or the behavior, if a person is demonstrably ignorant of a certain state of affairs or facts, he or she cannot have certain emotions.’ (Solomon, 1993: 11).

Multi-level models imply that this picture is too simple. How emotional and other representations interact, if they interact at all, depends on details of cognitive architecture as well as on the content of the representations. This architecture, of course, cannot be determined by studying the conceptual relations between the

contents of emotional representations. Phobias and affective primacy phenomena provide insight into the architecture of the emotion system by revealing that certain information, such as partially analyzed visual data, is available to low-level appraisal but not to high-level appraisal^v. A complimentary insight is provided by people with ‘flattened affect’, who are apparently able to carry out high-level appraisal but not low-level appraisal and who do not experience the physiological components of normal emotional response. The possibility of flattened affect without intellectual impairment reveals that only low-level appraisal has direct connections to the effector systems for the automated components of rapid emotional response.

Normal human emotion involves several subsystems that interact, and interact with other cognitive subsystems, in ways that reflect the particular cognitive architecture in which they are embedded. Nussbaum suggests that we can cope with this phenomenon using a ‘flexible notion of intentionality’. This is supposed to allow us to identify what is in common between animal and human emotion. The suggestion is presumably that there is some degree of isomorphism between the way in which high-level representations relate to one another on the basis of their content and the way in which low-level representations relate to one another on the basis of their ‘content’. Thus, the ‘logic’ of the emotion will be the same in the two cases. But in my view, there are likely to be more radical differences between the representational states involved in low-level and high-level appraisal. Ruth Millikan has suggested that mental representations in simple organisms may unite the functions of beliefs and desires (Millikan, 1996). Low-level appraisal in humans seems to manifest the same ‘collapse of the attitudes’. Consider the low-level appraisal of the core relational theme ‘a demeaning offence to me and mine’ that presumably occurs when a soccer player is dribbling the ball down the field, another player grabs his jersey causing him to lose the ball, and the first player turns *angrily* towards the second. It is misleading to say that the relevant brain region *believes* that the core relational theme has been instantiated. Beliefs are mental states that represent how things are and which produce action in conjunction with desires - representations of how the world should be. But in low-level appraisal for anger there is no question as to what action will be taken. The frustrated player in our example will orient to the stimulus, produce the pan-cultural facial expression of anger and undergo physiological changes to prepare them for aggressive action. The ‘affective computation’ in this example is simultaneously the belief that the world is a particular way and the intention to act in a particular way. Likewise for the better understood case of affective computing of fear in the amygdala, and, presumably, for any emotion that has a clear behavioral signature and can be induced to exhibit affective primacy. I suggest that it is simply misleading to describe low-level appraisal as evaluative judgment, or using any other locution derived from a psychology that presumes a fundamental distinction between data and goals. Instead, low-level emotional appraisal seems to involve action-oriented representation (Griffiths, In Press-c; Scarantino, In Press).

Another way in which low-level emotional appraisal may differ from high-level is in terms of the narrow inferential role imposed on low-level representations by the task-specific architecture in which they occur. The inferential role of these representations is impoverished in three ways. First, low-level appraisal processes do not have access to most of what is represented elsewhere in the brain, which is why knowledge that the cockroach in my drink has been completely sterilized does not eliminate the disgust response. Hence, many inferences that would seem to follow from the content we ascribe to this low-level appraisal — ‘I am taking in or being too close to an indigestible object’ — are not actually made by subjects because they cannot recombine that content appropriately with their other contentful states. Secondly, the processes of affective computing, as opposed to their final output, are not available for inspection by other cognitive sub-systems. Once again, architectural barriers to information flow block inferences that follow from what would otherwise seem the natural content to ascribe to those states. Finally, the inferential principles used in affective computing are not truth-preserving, but heuristically survival-enhancing. It does not follow by any reasonable deduction that if I have been poked hard and unexpectedly in the small of my back then I have suffered ‘a demeaning offence to me and mine’ but the automatic appraisal mechanism for anger will reliably draw that conclusion^{vi}.

If the concepts that figure in the content ascribed to a representation do not have their usual inferential role, then what is meant by attributing that content? The differences between the role of the representations involved in low-level appraisal and the inferential role of the content-sentences with which we describe those appraisals strongly suggests that, at least when applied to these low level processes, appraisal theories are not theories of cognitive content. I have suggested elsewhere that they are theories of the ecological significance of the environment to the organism (Griffiths, In Press-c). That significance is tracked by multiple cognitive subsystems using different environmental cues and *different psychological and neurological mechanisms*. This, of course, returns us to the main theme of this paper, which is that the states and processes we call ‘emotion’ are not all of the same kind.

4. Conclusion: Why emotion is still not a natural kind and why it matters

I have defended the thesis that the psychological states and processes that fall under the vernacular concept of emotion are unlikely to be a single ‘natural kind’. I believe that the same is probably true of some specific vernacular categories of emotion, although I have not defended that view here. I have considered two main alternatives. The first is that the emotions are simply those of states and processes studied in ‘affective neuroscience’. I have argued that, if empirically successful, current approaches to affective neuroscience will define a much larger class of affective and motivational processes, united by the role in all these processes of felt affect. I also think it is important not to foreclose the possibility that some emotions do not involve the basic emotions, or involve them only peripherally, although I have not argued for this here.

The second suggestion was that emotions are appraisals of the significance of the environment to the organism. I agree that multi-level appraisal theory brings many different kinds of emotion under a single taxonomic scheme, but I have argued that does so precisely by abstracting away from the kind of psychological processes that constitute those emotions. A theory of emotions based on their content is not a psychological theory, but an ecological theory, as I have argued at more length elsewhere (Griffiths In Press-c).

It is reasonable to ask why the claim that emotion is not a natural kind matters to philosophy? The simple answer is that many philosophers still take it to be their role to provide an account of the genesis, development and consequences of a 'typical' human emotion (e.g. Wollheim, 1999). If I am correct, then there is no such thing as a *typical* emotion. Instead, there are different kinds of emotion, or of emotional process, each of which should be treated in its own terms and whose various possible interactions should be studied. Similarly, the idea that all emotions are intentionally directed at aspects of the environment *in the same sense*, is a core methodological assumption of much current philosophical work on the emotions. If I am correct, then we should be more concerned with the distinctive properties of the different kinds of emotional intentionality and with how these different kinds of emotion process interact in real emotional episodes. Cases in which people have an emotion in one sense and do not have it in another should be as illuminating for the philosophy of psychology as they have been for psychology itself.

References

- Blackburn, S. (1998). *Ruling Passions: A theory of practical reasoning*. Oxford and New York: Oxford University Press.
- Buss, D. M. (2000). *The Dangerous Passion: Why Jealousy is as Essential as Love and Sex*. New York: Simon and Schuster.
- Charland, L. C. (2001). In defence of 'emotion'. *Canadian Journal of Philosophy*, 31(1), 133-154.
- Cosmides, L., & Tooby, J. (2000). Evolutionary Psychology and the Emotions. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of the Emotions* (2 ed., pp. 91-115). New York and London: Guildford Press.
- Damasio, A. R. (1994). *Descartes Error: Emotion, Reason and the Human Brain*. New York: Grosset/Putnam.
- Damasio, A. R. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace.
- Delancey, C. (2001). *Passionate Engines: What emotions reveal about mind and artificial intelligence*. New York, Oxford: Oxford University Press.
- Doris, J. M. (2001). Review of Griffiths, Paul E. What Emotions Really Are: the problem of psychological categories. *Ethics*, xxx, xxx-xxx.
- Ekman, P., Friesen, W. V., & Simons, R. C. (1985). Is the startle reaction an emotion? *Journal of Personality and Social Psychology*, 49 (5), 1416-1426.

- Gaulin, S. J. C., & McBurney, D. H. (2001). *Psychology: An Evolutionary Approach*. Upper Saddle River, NJ: Prentice Hall.
- Goodman, N. (1954). *Fact, Fiction and Forecast* (1 ed.). London: Athlone Press, University of London.
- Gordon, R. M. (1987). *The Structure of Emotions*. Cambridge: Cambridge University Press.
- Greenspan, P. (1988). *Emotions and Reasons: An Inquiry into Emotional Justification*. New York: Routledge.
- Griffiths, P. E. (1994). Cladistic classification & functional explanation. *Philosophy of Science*, 61 (2), 206-227.
- Griffiths, P. E. (1996a). Darwinism, Process Structuralism and Natural Kinds. *Philosophy of Science*, 63(3 Supplement: PSA 1996 Contributed Papers), S1-S9.
- Griffiths, P. E. (1996b). The Historical Turn in the Study of Adaptation. *British Journal for the Philosophy of Science*, 47(4), 511-532.
- 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* (pp. 208-228). Cambridge, MA: MIT Press.
- Griffiths, P. E. (2001a). Emotion and the problem of psychological categories. In A. W. Kazniak (Ed.), *Emotions, Qualia and Consciousness* (pp. 28-41). Singapore, New Jersey, Hong Kong, London: World Scientific.
- Griffiths, P. E. (2001b). From adaptive heuristic to phylogenetic perspective: Some lessons from the evolutionary psychology of emotion. In H. H. I. Holcolmb (Ed.), *The Evolution of Minds: Psychological and Philosophical Perspectives* (pp. 309-325). Dordrecht: Kluwer.
- Griffiths, P. E. (In Press-a). Basic Emotions, Complex Emotions, Machiavellian Emotions. In A. Hatzimoysis (Ed.), *Proceedings of the Royal Institute of Philosophy, 2001* (pp. xxx-xxx).
- Griffiths, P. E. (In Press-b). Emotions as Natural Kinds and Normative Kinds. *Philosophy of Science, Supplement: Proceedings of the 2002 Biennial Meeting of the Philosophy of Science Association*, xxx-xxx.
- Griffiths, P. E. (In Press-c). Towards a Machiavellian Theory of Emotional Appraisal. In D. Papineau & D. Evans & P. Cruise (Eds.), *Emotion, Evolution and Rationality* (pp. xxx-xxx). Cambridge: Cambridge University Press.
- Kenny, A. (1963). *Action, Emotion & Will*. London: Routledge & Kegan Paul.
- Lazarus, R. S. (1991). *Emotion and Adaptation*. New York: Oxford University Press.
- LeDoux, J. (1996). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Simon and Schuster.
- LeDoux, J. E. (1993). Emotional networks in the brain. In M. Lewis & J. M. Haviland (Eds.), *Handbook of Emotions* (pp. 109-118). New York: Guildford Press.
- Lyons, W. (1980). *Emotion*. Cambridge: Cambridge University Press.

- Millikan, R. G. (1996). 'Pushmi-pullyu Representations'. In L. May & M. Friedman & A. Clark (Eds.), *Mind and Morals: Essays on cognitive science and ethics* (pp. xxx-xxx). Cambridge, MA: MIT Press.
- Nussbaum, M. C. (2001). *Upheavals of thought: The intelligence of emotions*. Cambridge, New York: Cambridge University Press.
- Öhman, A. (1999). Distinguishing Unconscious from Conscious Emotional Processes: Methodological Considerations and Theoretical Implications. In T. Dalgleish & M. J. Power (Eds.), *Handbook of Emotion and Cognition* (pp. 321-352). Chichester: John Wiley and sons.
- Öhman, A. (2002). Automaticity and the amygdala: Nonconscious responses to emotional faces. *Current Directions in Psychological Science*, 11(2), 62-66.
- Owen, R. (1843). *Lectures on the Comparative Anatomy and Physiology of the Vertebrate Animals, Delivered at the Royal College of Surgeons, in 1843*. London: Longman, Brown, Green and Longmans.
- Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford and New York: Oxford University Press.
- Prinz, J. (2000). Getting to the Heart of Emotions. New York: Paper delivered at American Philosophical Association.
- Robinson, J. (1995). Startle. *The Journal of Philosophy*, 92, 53-74.
- Scarantino, A. (In Press). Affordances Explained. *Philosophy of Science, Supplement: Proceedings of the 2002 Biennial Meeting of the Philosophy of Science Association: Contributed Papers*.
- Scherer, K. R. (1999). Appraisal Theory. In T. Dalgleish & M. J. Power (Eds.), *Handbook of Emotion and Cognition* (pp. 637-663). Chichester: New York.
- Solomon, R. (1976). *The Passions*. New York: Doubleday.
- Solomon, R. C. (1993). The philosophy of emotions. In M. Lewis & J. M. Haviland (Eds.), *Handbook of Emotions* (pp. 3-15). New York: Guildford.
- Teasdale, J. D. (1999). Multi-level Theories of Cognition-Emotion Relations. In T. Dalgleish & M. J. Power (Eds.), *Handbook of Cognition and Emotion* (pp. 665-681). Chichester: John Wiley and sons.
- Tooby, J., & Cosmides, L. (1992). The Psychological Foundations of Culture. In J. H. Barkow & L. Cosmides & J. Tooby (Eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (pp. 19-136). Oxford and New York: Oxford University Press.
- Wollheim, R. (1999). *On the Emotions*.
- Zajonc, R. B. (1980). Feeling & thinking: preferences need no inference. *American Psychologist*, 35, 151-175.

* To appear in *Philosophers on Emotion*, Robert Solomon (ed). Oxford University Press.

ⁱ Page numbers without citations refer to Griffiths (1997).

ⁱⁱ E.g. *Nature*, 391, 1998; *Times Literary Supplement* July 17 1998; *Philosophy in Review* 18, 1998; *Australian's Review of Books*, April 1998; *Metascience* 8 (1) 1999; *Australasian Journal of Philosophy* 77 (4) 1999; *Philosophical Review*, 108 (1) 1999; *American Journal of Psychology*, Fall 2000: 472-478; *Dialogue* 38 (4) 2000.

ⁱⁱⁱ I take this to be an open question. Many psychologists of emotion reject the idea that emotions come in discrete types and support instead a dimensional account of the emotion system. Even if basic emotions form discrete types, the best theory of complex emotions might still be dimensional rather than typological.

^{iv} One reason I take this possibility seriously is that contemporary research in evolutionary psychology does not proceed under the assumption that all emotions are composed of or have at their core one of the basic emotions. In fact, so-called ‘Santa-Barbara school’ evolutionary psychologists make considerable play of the claim that ‘all emotions are equally basic’ (Buss, 2000; Cosmides & Tooby, 2000; Gaulin & McBurney, 2001). These researchers define an emotional adaptation as any motivational mechanism designed to influence behavior in some specific problem domain and whose operation cannot be understood as the application of domain-general processes to that problem. Their commitment to the ‘massive modularity hypothesis’ makes them suspicious of attempts to explain all specific emotions via the interaction of a smaller number of general-purpose mechanisms, such as the basic emotions and our capacity for cognitive evaluation of stimulus situations.

^v I do not think it is yet clear how many appraisal levels are needed for an adequate representation of emotional appraisal. When I talk of ‘low’ and ‘high’ levels I do not mean to imply that there are just two levels of appraisal, but rather to mark the existence of a dimension of difference.

^{vi} Some would say that the appraisal mechanism has ‘innate knowledge’ that this cue reliably predicted conspecific aggression in ancestral environment (Tooby & Cosmides, 1992), but the scientific substance of this claim is simply that the appraisal mechanism consistently makes certain inferences.