

Emotions as Natural and Normative Kinds

Paul E. Griffiths[†]

In earlier work I have claimed that emotion and some emotions are not 'natural kinds'. Here I clarify what I mean by 'natural kind', suggest a new and more accurate term, and discuss the objection that emotion and emotions are not descriptive categories at all, but fundamentally normative categories.

1. Introduction. 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, or so I have argued (Griffiths 1997). The psychological, neural, ecological, evolutionary, and other theories that best explain any particular subset of human emotions will not adequately explain the whole range of human emotions. In a slogan, emotions are not a 'natural kind'. The same may be true of many more specific emotion categories, such as anger and love. On some occasions when it might be properly said in common speech that a person is angry, certain theories will adequately explain that person's state. On other occasions of anger, however, other theories will be needed. In earlier work I have described my position as a form of eliminativism about emotion, because it implies that the term 'emotion' and some specific emotion terms like 'anger' are examples of 'partial reference'. The term 'jade' is the standard example of partial reference. 'Jade' may be either of two different stones, jadeite or nephrite, and the term 'jade' partially refers to each of these two kinds of stone. It follows from this fact that for the purposes of geology or chemistry, jade cannot be treated as a single kind of thing. The properties of the two minerals have to be investigated separately, their geological origins explained separately, and their abundance in unexplored geological deposits predicted separately. In a similar fashion, the sciences of the mind

[†]To contact the author, please write to: Department of History and Philosophy of Science, University of Pittsburgh, Pittsburgh, PA 15232; e-mail: pauleg@pitt.edu.

Philosophy of Science, 71 (December 2004) pp. 901–911. 0031-8248/2004/7105-0021\$10.00
Copyright 2004 by the Philosophy of Science Association. All rights reserved.

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, domain-specific biases in motivation, socially sustained pretences, and other more specific categories of psychological state and process that have been identified or hypothesized in the varied literature that sets out to address human emotion.

While the critical response to my book has been generally positive,¹ most commentators have remained unconvinced by the argument that emotion is not a natural kind. In many cases this seems to reflect different understandings of what it is to be a 'natural kind'. Some critics have been concerned to establish that there is a single domain of related phenomena to be studied by a discipline of 'affective science' or 'affective neuroscience' (Charland 2002). I have argued elsewhere that even if correct, this is consistent with the claim that, like the domain of 'memory science', the domain of 'affective science' encompasses several different phenomena that it is important to distinguish (Griffiths 2003).

Other critics take me to suggest that if emotion is not a natural kind, and thus not subject to a unified scientific explanation, then the concept is unviable in any cognitive role whatsoever. In a brief but thoughtful review, John Doris points out that "Griffiths's argument from heterogeneity to eliminativism applies too widely—the same style of argument might be applied to any number of seemingly viable concepts" (Doris 2000, 618). He gives the examples of 'game' and 'cancer'. 'Cancer' is probably not a good example for his purpose. While there are many specific cancers, they are all cases in which cell lineages escape the mechanisms that normally control the proliferation and differentiation of somatic cell lines. Lenny Moss has recently outlined the various overarching conceptions of the control of development at the cellular level that have guided oncology research in the past century (Moss 2002). If it were to turn out that certain cancers conform to one of these fundamental conceptions of cellular pathology whilst other cancers conform to another, it would not only be important for medical professionals to grasp that 'cancer' had been discovered to denote two very different cellular-level phenomena, it would probably be important for patients to understand it too, since it would almost certainly have implications for treatment and prognosis. If such a discovery were made, the continuing use of the undifferentiated term 'cancer' might, indeed, be something to discourage (I

1. Adolphs 2000; Brown et al. 1999; de Sousa 1999; Dolan 1998; Doris 2000; Gerrans 2000; Hacking 1998; Rottschaefer 1998; Slater 1998; Solomon 1999; Williamson 1998.

should stress that I am *not* suggesting that different cancers *actually* conform to different fundamental conceptions of the nature of cancer).

The example of 'game' is more apposite, because there is clearly no reason to discourage its use, even though, as Wittgenstein famously showed, it is very far from denoting a group of things that are all 'of the same kind'. Once one considers the full range of things that can be counted as 'games' it is quite a challenge to say why it is epistemically or pragmatically valuable to have a single term for such a broad category, rather than different terms for sports, pastimes, etc. 'Emotion' represents a similarly complex and productive conceptual challenge. As I mentioned at various points in my 1997, anthropologists and social-constructionist psychologists have made a number of valuable contributions to understanding the social functions of the emotion category. Many of these center on its role in removing behavior from the realm of blame and responsibility. Perhaps as a matter of conceptual analysis it is illuminating to look at the category of emotion not referentially, but in terms of the role played by classifying behavior as emotional in a set of social practices.

Doris suggests just such an account of the emotion category: it derives its unity from a role in normative practices such as self-criticism, rather than from any role it may have in describing and explaining behavior. This line of thought, however, seems to me more a valuable complement to my claim that emotion is not a natural kind of psychological state or process, rather than an objection to it. While I did not explore this aspect of the subject in my earlier book, I did stress that

scientific understanding is not the sole or main goal of everyday life. Vernacular emotion concepts serve other purposes besides those of explanation and induction. The future development of these concepts is unlikely to be as simple as their refinement or replacement by the concepts best suited to scientific understanding. (Griffiths 1997, 228)

I will return to this theme in Section 4. First, however, I want to clarify what I mean by 'natural kind' before suggesting a new term that may better convey this meaning.

2. What Do I Mean by 'Natural Kind'? I use the term 'natural kind' to denote categories which admit of reliable extrapolation from samples of the category to the category as a whole. In other words, natural kinds are categories about which we can make inductive scientific discoveries. This usage is in direct line of descent from the initiators of the modern discussion of natural kinds in the 1840s, John Stuart Mill and William Whewell. The existence and importance of what Mill called 'real kinds' was one of the few topics in the philosophy of science on which these two authors agreed. The proper objects of investigation in the inductive

sciences are those “classes in which we have not a finite but an *inexhaustible* body of resemblances amongst individuals, and groups made by nature, not by mere definition” (Whewell [1860] 1971, 290).

Whewell’s brief definition perfectly expresses the two, linked ideas that I intended to convey when I used the term ‘natural kind’ in my earlier work. First, good scientific categories have the property that the criteria we use to judge whether an individual is a member of the category do not exhaust what unites all the members of that category. Second, the ability to clearly define a category is neither necessary nor sufficient to make that category a productive object of scientific inquiry. Hence it came as a considerable surprise to me when some philosophers took my argument that emotion is not a natural kind to be premised on the claim that ‘emotion’ lacks a clear definition, as Martha Nussbaum does in the following attempt to show that my position is internally inconsistent:

It is particularly odd that Griffiths, who is a stern critic of the reliance on ordinary use and ordinary conceptions, should rely on them himself in a quite uncritical way when arguing that the category “emotion” contains such great heterogeneity that no interesting single account is possible. He uses the word quite loosely in order to establish that the things falling under it are multiple and not unified; and yet it is he who holds that our loose use is not to be trusted. (Nussbaum 2001, 8)

But issues of vagueness and ‘loose use’ are simply irrelevant when assessing whether something is a natural kind. The Aristotelian category of ‘superlunary objects’ can be defined by a single necessary and sufficient condition—being outside the orbit of the moon—but it ceased to be a natural kind with Galileo’s astronomical discoveries. Conversely, the history of the gene concept might not unreasonably be summed up by saying that the more clearly ‘gene’ has been defined, the more quickly that definition has been refuted (Falk 1986, 2000). But the study of this ill-defined category has been fantastically productive. The question about emotion, therefore, is not whether we can give a single ‘account’ of the category in the sense of a philosophical analysis of the emotion concept, but whether the category thus singled out is a productive object of scientific enquiry.

In my 1997 book I built on the work of several other philosophers and scientists to construct a detailed account of natural kinds in psychology and biology, an account further elaborated in Griffiths 1999 and 2001, and briefly sketched here. The fundamental scientific practices of induction and explanation presume that some of the observable correlations between properties are ‘projectible’ (Goodman 1954). That is, correlations observed in a set of samples can be reliably ‘projected’ to other instances

of the category. Scientific classifications embody our current understanding of where such projectible clusters of properties are to be found. The species category, for instance, classifies organisms into sets that reliably cluster as yet undiscovered morphological, physiological, and behavioral properties. These properties of the species can be discovered by studying a few members of the species.

The traditional requirement that natural kinds be the subjects of spatiotemporally universal and exceptionless laws of nature 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 concept of a law of nature to the notion that statements are to varying degrees 'lawlike' (have counterfactual force). This allows a broader definition of a natural kind. A category is (minimally) natural if it would be reasonable to place some degree of reliance on some inductive predictions about unobserved 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 allow *reliable* predictions in a *large domain* of properties. The classic examples of natural kinds, chemical elements and biological species, meet these desiderata.

It is important to note that categories are natural relative to specific domain(s) of properties to which they are connected by background theories. The category of domestic pets is not a good 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. Similarly, emotion is not a natural kind relative to the domains of properties that are the focus of investigation in psychology and the neurosciences, or so I have argued. For emotion to be a natural kind in this sense, it would need to be 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.

3. From Natural Kinds to Investigative Kinds. The term 'natural kind' carries a lot of unwelcome baggage. To many philosophers it suggests a concern with the most fundamental kinds of the physical sciences, kinds far removed from everyday human activity and even from the mere terrestrial process of evolution (Ellis 2001). To other philosophers the term 'natural kind' implies a single best taxonomy of nature independent of any particular human purposes (Dupré 1993). Neither of these ideas is tightly connected to my own two concerns: the need for reliable property

correlations for induction and explanation, and the fact that whether a category has a clear definition and sharp boundaries is orthogonal to whether it is scientifically useful. For something to be a natural kind in this sense it is sufficient, as Richard Boyd has stressed, that there exist a cluster of correlated properties and that we can offer some defeasible justification for extrapolating those correlations (Boyd 1991, 1999; see also Wilson 1999). Biological species are natural kinds, I have argued, simply because current theory provides ample reason to suppose that the successes to date in investigating the properties of living systems by extrapolating from samples of one species to the rest of that species are not accidental! This is true despite the fact that the very same theories strongly suggest that species have no definitive 'nature'. Any property of a species may exhibit diversity within the species at a time and may evolve over time. Species merely represent the intersection of several processes that cause biodiversity to be 'clumpy', the process of heredity being perhaps the most important of these. In the same way, the parts of animals, such as the liver, the human expression of surprise, and, I would argue, the primate disgust response, are best regarded as biological homologues: characteristics that arose at some point in evolution and that now mark out more or less reliably the set of organisms 'downstream' of that evolutionary event.

This conception of a natural kind is intended not as a contribution to metaphysics but as a contribution to the philosophy of science. Inductive enquiry presumes that there are projectible properties, which is as much as to say that there are some categories that are natural kinds in the sense just outlined. This realization yields insights into the nature of scientific concepts and of conceptual change in science. The actual name 'natural kind' has been rendered increasingly inappropriate by what Boyd has termed the 'enthusiasm' for the concept to which it refers (Boyd 1991). The insights into scientific practice that can be obtained from the 'natural kind' model are applicable not only to the natural sciences, but also to the human sciences. They are even applicable to those parts of the human sciences in which, as Ian Hacking has pointed out, the practices of classifying and theorizing feed back into substantive changes in the subject matter of those sciences (Hacking 1992, 1995). Given the historical baggage that attaches to the term 'natural kind' and its literal inappropriateness to many of the scientific categories to which it is now applied, it seems preferable to introduce an alternative.

One possibility would be to recruit Hans-Jörg Rheinberger's term 'epistemic object' (Rheinberger 1997). Rheinberger uses this phrase to refer to entities, like the gene, that enter science as targets of research—putative entities that are discussed and investigated through certain experimental practices. Stating at the outset a precise theoretical definition would vitiate

the whole purpose of an epistemic object. Nor can such objects be defined operationally, since that would tie them to a particular set of experimental practices that will likely be superseded (this is not to deny the very real role of 'operational definitions' in science, merely to deny that these cognitive tools are aptly named.) The idea of an epistemic object as something that cannot be defined but only investigated is certainly resonant with the idea of a natural kind that I have outlined above. However, Rheinberger has associated this term very strongly with a nonreferential view of language. Epistemic objects are not merely investigated through a set of experimental practices, but constituted by those practices. There is no reason to link the account of natural kinds sketched above to a rejection of scientific realism. A commitment to scientific realism is no impediment to recognizing that natural-kind concepts are open ended, in the sense that any account of what it is to be a member of the kind may be refuted by new empirical findings.

A more neutral substitute for 'natural kind' that carries many of the right connotations is 'investigative kind' (Brigandt 2003). This term highlights the fact that the emphasis in the model of natural kinds outlined above is on open-ended investigation. A natural-kind *concept* is a concept that it makes sense to seek to clarify through empirical inquiry. Such concepts are ongoing projects of inquiry in which extension and intension are altered to preserve inductive and explanatory power. The extension of a natural-kind concept should change to ensure that the instances of the kind reliably possess a rich cluster of the properties the kind is used to investigate. Untranscribed regulatory regions in the genome, for example, fit the criteria for classical Mendelian genes. But now that it is possible to describe the genetic material and its activities at the molecular level, these sequences stand in stark contrast to sequences that template for polypeptides and other functional gene products. Hence, it seems, these sequences are no longer classified as genes (Waters 1994). The intension of a natural-kind concept may need to be altered to reflect the fact that the background theories that license treating a category as a natural kind have altered. It is clear that in pre-Darwinian biology the particular historical origins of taxa were not essential to those taxa. Today, however, nearly all the competing accounts of systematics accept as a necessary condition of taxon membership a shared origination at a particular point in a genealogical network. The fact that members of a taxon share a common ancestor is a central part of the network of theory that licenses extrapolating from a sample to the whole. Hence that shared origin is treated as a defining feature of the taxon (Griffiths 1999).

In the remainder of this article I will use Brigandt's term 'investigative kind' rather than 'natural kind'. Amongst its other virtues, the new term nicely captures what seems to be the predominant attitude to 'definitions'

of emotion in the scientific literature: “we must sharpen our conceptual tools as best we can and have faith that in using them to untangle the complexity [of emotional behavior] we shall see how to fashion better ones” (Hinde 1985, 990).

4. Emotion as a Normative Kind. Dominic Murphy and Stephen Stich have suggested that arguments similar to those I have offered concerning emotion show that ‘mental illness’ or ‘psychopathology’ does not represent a ‘natural kind’ of psychological or neurological processes. They conclude, however, that in this case the category remains valuable as a normative kind, since it brings together all the processes that are “ways of going *wrong*” (Brown et al. 1999, 25; their emphasis). Building on this suggestion, John Doris has argued that all instances of the vernacular category of anger may form a single kind for the purposes of asking, for example, whether anger is *warranted* by the situation at hand. Similarly, he suggests, all ‘negative emotions’ may share the normative feature that we wish not to experience them. This is an important point, and it is clear that emotion categories do function as normative kinds in this sense. However, I will try to show here that this does not provide any basis for rejecting my view that emotion is not a single investigative kind, but rather complements and enriches that view. I will suggest that, firstly, normative kinds are open ended in the same way as investigative kinds, although for different reasons; secondly, emotions are simultaneously investigative and normative kinds; thirdly, the two kinds of conceptual dynamics interact, so that the realization that emotion (or some particular emotion) is not a single investigative kind will impact on its role as a normative kind.

Normative kinds are open ended because normative projects as well as epistemic projects provide motives for altering extension and intension. Just as a category can be altered for the sake of inductive power, a category can be altered, for example, as part of a project of social reform. Ian Hacking has presented a valuable case study of such a change in the evolving definition of child abuse from the 1960s to the present day (Hacking 1992). The change from viewing a pattern of childcare as normal to viewing it as abusive need not reflect an epistemic project, such as maximizing the predictive power of child abuse as a diagnostic category in psychiatry. The change can equally well represent the spread of a new normative model of the relationship between parent and child, or a change in the relative value placed on various traits of the older child or adult, such as placing a higher value on personal fulfillment and a lower value on social conformity.

It is unlikely that description and prescription can be cleanly separated. The very same ‘human kinds’ (Hacking 1999) are used to predict and explain human behavior and to prescribe and condemn that behavior.

This is self-evident in the case of many emotion categories, such as love, jealousy, or Doris's example of anger. There are normative criteria for whether anger is warranted and perhaps also normative standards about how to behave in anger, but there is also a body of common sense about what *actually* makes people angry and about how angry people *actually* behave. Diagnostic psychiatric categories provide another rich source of examples. There are both skeptics and enthusiasts about the diagnosis of attention deficit hyperactivity disorder (ADHD), but surely neither group would deny that the diagnosis came into being as the result of a perception that certain behavior was unacceptable. The issue that separates them is how much these children have in common and how much they differ from other children, over and above the fact that their behavior is unacceptable.

Given that the very same categories serve as both investigative kinds and normative kinds, the remaining issue is how such categories behave when the investigative projects in which they serve pull them in one direction and the normative projects pull them in another. Doris seems to envisage that anger would be retained as a single normative category even if it were discovered that different instances of anger are very different psychologically. This seems to me unlikely both from a consideration of this particular case and on general grounds. In this specific case, the nature of the psychological processes underlying anger is directly relevant to whether the behavior is voluntary and thus to its normative status. Suppose it comes to be believed that certain cases of anger represent a highly automated, species-typical response of some natural subset of human beings. It is not hard to imagine such a picture being accepted for the following kind of anger: A young male is waiting in line outside a nightclub. Someone unexpectedly pokes him sharply in the small of the back. He spins around, making a threat expression, probably the 'square-mouthed' variety that represents subordinate threat in chimpanzees. His body adjusts physiologically for violent action. Now suppose that a young woman has fallen against him and is smiling and apologizing profusely. The episode of anger is over almost immediately. Accepting the model of what happened in this episode associated with the 'affect program' theory of basic emotions seems to me to move it into an entirely different normative category from many other episodes of anger. I can imagine someone arguing that it simultaneously removes this instance from the category of true anger, but since this implies that no chimpanzee is ever truly angry, that seems merely a backhanded way to admit that there is more than one kind of anger. More generally, I suggest that normative judgments about emotion are no more likely to be independent of our beliefs about the nature of emotion than normative beliefs about gender or race are independent of our scientific beliefs about the validity and nature of those categories.

REFERENCES

- Adolphs, Ralph (2000), "Emotions Really Are Social", *American Journal of Psychology* 3 (Fall): 472–478.
- Boyd, Richard (1991), "Realism, Anti-foundationalism and the Enthusiasm for Natural Kinds", *Philosophical Studies* 61: 127–148.
- (1999), "Homeostasis, Species and Higher Taxa", in Robert A. Wilson (ed.), *Species: New Interdisciplinary Essays*. Cambridge, MA: MIT Press, 141–185.
- Brigandt, Ingo (2003), "Species Pluralism Does Not Imply Species Eliminativism", *Philosophy of Science* 70 (Proceedings): 1305–1316.
- Brown, Richard, Dominic Murphy, Stephen Stich, Donald Dryden, Paul Redding, Neil MacNaughton, and Paul E. Griffiths (1999), "Eliminating Emotions?" (review symposium on *What Emotions Really Are: The Problem of Psychological Categories* by Paul E. Griffiths), *Metascience* 8(1): 5–62.
- Charland, Louis C. (2002), "The Natural Kind Status of Emotion", *British Journal for the Philosophy of Science* 53(4): 511–537.
- de Sousa, Ronald (1999), review of *What Emotions Really Are* by Paul E. Griffiths, *Dialogue: Canadian Philosophical Review* 38(4): 908–911.
- Dolan, Ray J. (1998), "How We Feel", *Nature* 391: 35–36.
- Doris, John M. (2000), review of *What Emotions Really Are: The Problem of Psychological Categories* by Paul E. Griffiths, *Ethics* 10(3): 617–619.
- Dupré, John (1993), *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, MA: Harvard University Press.
- Ellis, Brian (2001), *Scientific Essentialism*. Cambridge: Cambridge University Press.
- Falk, Raphael (1986), "What Is a Gene?", *Studies in the History and Philosophy of Science* 17: 133–173.
- (2000), "The Gene: A Concept in Tension", in Peter J. Beurton, Raphael Falk, and Hans-Jörg Rheinberger (eds.), *The Concept of the Gene in Development and Evolution*. Cambridge: Cambridge University Press, 317–348.
- Gerrans, Phil (2000), review of *What Emotions Really Are* by Paul E. Griffiths, *Australasian Journal of Philosophy* 77(4): 512–515.
- Goodman, Nelson (1954), *Fact, Fiction and Forecast*, 1st ed. London: Athlone Press, University of London.
- Griffiths, Paul E. (1997), *What Emotions Really Are: The Problem of Psychological Categories*. Chicago: University of Chicago Press.
- (1999), "Squaring the Circle: Natural Kinds with Historical Essences", in Robert A. Wilson (ed.), *Species: New Interdisciplinary Essays*. Cambridge, MA: MIT Press, 208–228.
- (2001), "Emotion and the Problem of Psychological Categories", in Alfred W Kaszniak (ed.), *Emotion, Qualia and Consciousness*. Singapore, New Jersey, Hong Kong, and London: World Scientific, 28–41.
- (2003), "Is Emotion a Natural Kind?", in Robert C. Solomon (ed.), *Thinking about Emotion: Contemporary Philosophers on Emotion*. Oxford and New York: Oxford University Press, 233–249.
- Hacking, Ian (1992), "World-Making by Kind-Making: Child Abuse for Example", in Mary Douglas and David L. Hull (eds.), *How Classification Works: Nelson Goodman among the Social Sciences*. Edinburgh: Edinburgh University Press, 180–238.
- (1995), *Rewriting the Soul: Multiple Personality and the Sciences of Memory*. Princeton, NJ: Princeton University Press.
- (1998), "By What Links Are the Organs Excited?" *Times Literary Supplement* 4972 (July 17): 11–12.
- (1999), "Making Up People", in Mario Biagioli (ed.), *The Science Studies Reader*. New York: Routledge, 161–171.
- Hinde, Robert A. (1985), "Was 'The Expression of Emotions' a Misleading Phrase?", *Animal Behaviour* 33: 985–992.
- Moss, Lenny. (2002), *What Genes Can't Do*. Cambridge, MA: MIT Press.

- Nussbaum, Martha C. (2001), *Upheavals of Thought: The Intelligence of Emotions*. Cambridge: Cambridge University Press.
- Rheinberger, Hans-Jörg (1997), *Towards a History of Epistemic Things: Synthesising Proteins in the Test Tube*. Stanford, CA: Stanford University Press.
- Rottschaefer, William A. (1998), review of *What Emotions Really Are* by Paul E. Griffiths, *Bridges* (Fall), 274–283.
- Slater, Carol (1998), review of *What Emotions Really Are* by Paul E. Griffiths, *Philosophy in Review* 18(5): 335–337.
- Solomon, Robert C. (1999), review of *What Emotions Really Are* by Paul E. Griffiths, *The Philosophical Review* 108(1): 131–134.
- Waters, Kenneth (1994), “Genes Made Molecular”, *Philosophy of Science* 61: 163–185.
- Whewell, William ([1860] 1971), *On the Philosophy of Discovery*. New York: Burt Franklin.
- Williamson, David (1998), “Mixed Feelings”, *Australian’s Review of Books* (April), 14–15.
- Wilson, Robert A. (1999), “Realism, Essence, and Kind: Resuscitating Species Essentialism?”, in Robert A. Wilson (ed.), *Species: New Interdisciplinary Essays*. Cambridge, MA: MIT Press, 187–207.