Is Fear a Natural Kind?

Raamy Majeed
University of Auckland
Draft 2020

Are emotions natural kinds? While a philosophy of emotion typically proceeds as if they are, those who have turned their attention directly to answering this question are divided. It is hard to find anyone defending the view that everything which falls under our vernacular category 'emotion' forms a natural kind. In his seminal book, *What Emotions Really Are: The Problem of Psychological Categories*, Griffiths (1997) argues for a more cautionary position: though not all emotions form natural kinds, some do, *viz.* evolutionarily hardwired pancultural emotions or what Ekman (1973) calls "basic emotions", e.g. fear and anger.[1] This position has come under attack in recent years by constructionists, e.g. Russell (1991, 2003) and Barrett (2006, 2013, 2017), who argue that emotions, even basic ones, are not natural kinds but rather psychological constructs.[2]

While the constructionist position is getting some traction in current neuroscience and psychology, philosophers of emotion have by and large ignored this position or viewed it with pessimism. (I suspect this is, in part, because philosophers have a hard time believing Barrett's claim that our concepts don't just shape how we categorise our emotions, but that these concept-enabled categorisations are required for the very inception of emotions themselves).[3] I share some of this pessimism, so will leave the constructionist position aside. Instead, I want to offer some reasons for being sceptical of viewing basic emotions as natural kinds, which don't rely on assuming constructionism. Since the target of my critique is Griffiths's argument, for the purposes of this paper I shall take his conception of natural kinds as a given. Instead, I want to challenge the

[1] Note: the question I address in this paper is distinct from the broader question, which Griffiths also addresses, of whether emotion, i.e. the category as whole, forms a natural kind. See Charland (1995, 2002) for a defence of this view, and Griffiths (2004) for a critique.

[2] See Barrett and Russell (2014) for an overview of psychological constructionism, and Faucher (2013) for a taxonomy of constructionism about emotion more broadly.

[3] See Deonna and Teroni (2012: 58-61) for some other reasons for pessimism.

neuro-cognitive data which Griffiths employs to support his argument. In doing so, I aim to offer what I take to be a modest argument for an immodest conclusion, namely if we take recent clarifications concerning this data seriously, basic emotions, in particular fear, don't turn out to form natural kinds. The argument, simply put, is as follows.

According to Griffiths, natural kinds, very roughly, are the sorts of things of which we can make scientific discoveries. More specifically, they are the kinds of categories which allow for reliable extrapolation from samples of the category to all other members of the category. A way to justify such extrapolation is to group things together not on the basis of shared surface properties but rather on the basis of shared underlying causal mechanisms. So far the prevailing assumption has been that basic emotions, e.g. fear and anger, are natural kinds because each basic emotion results from an underlying causal mechanism unique to that very emotion. For example, there is a specific causal mechanism for fear, a specific mechanism for anger, and so on. Here Griffiths (2003) takes LeDoux's (1996) pioneering work on how neural circuits that involve the amygdala help the brain respond to threat as having "confirmed" a neurobiological account of the mechanisms for fear. Recent clarifications in neurobiology, especially the work of LeDoux (2012, 2016, 2017, 2019), however, shows this assumption to be false. In particular, the category 'fear' is employed to categorise entities with two distinct underlying mechanisms: a mechanism for our defensive responses, and another for the subjective feeling of fear, which crucially come apart. Fear, therefore, isn't a natural kind, and most likely neither are the other basic emotions.

The rest of this paper aims to put meat on the bones of this argument. In what follows, I provide some background concerning basic emotions and natural kinds (Section 1), and explain how recent clarifications in neurobiology undermine the orthodox view that basic emotions, especially fear, are natural kinds (Section 2). I then respond to what I take to be both an obvious and *prima facie* plausible objection to my argument (Section 3). I end by making some brief remarks about the implications of this argument for the unity of emotion (Section 4).

1. Background

Notions of natural kinds are legion. For the purposes of this paper, the conception of natural kinds at issue isn't the familiar metaphysical notion; roughly, things which 'carves nature at its joints'. Rather, natural kinds, following Griffiths (1997, 2002, 2004), are categories which can inform the scientific practices of induction and explanation. In particular, they are categories which allow for reliable extrapolation from samples of the category to all other members of the category.[4]

Consider the mineral jade. Jade isn't a natural kind. Though the things referred to as 'jade' share certain surface properties — e.g. they have the same perceptual appearance — they are, in fact, two distinct kinds of minerals with distinct structures at the molecular level. By the same token, the minerals that make up jade itself, *viz.* nephrite and jadeite, are natural kinds. All instances of nephrite, for example, share an underlying molecular structure, i.e. a silicate comprising of calcium and magnesium.

In *What Emotions Really Are*, Griffiths makes a similar claim about emotion. The vernacular category 'emotion' doesn't pick out a natural kind because although it refers to things which share certain surface properties — e.g. certain functional and possibly phenomenal profiles — they don't share the relevant profile. In the above example, the relevant profile was a molecular structure. According to Griffiths, the relevant profile for emotions is an underlying causal mechanism. 'Emotion' refers to basic emotions, cognitively complex emotions and social pretences, all of which are brought about by very different kinds of causal mechanisms. The "really" in the title, it turns out, is facetious. There is no such thing as a typical emotion, not *really*. Philosophy of emotion, which takes as its aim the genesis, development and consequences of a 'typical' emotion rests on a foundational mistake.[5]

---

[4] This conception of natural kinds has its roots in Goodman (1954) and Boyd (1989, 1991), and is developed by Griffiths (1997). In later work, Griffiths (2004), following Brigandt (2003), prefers the term "investigative kinds". See Hacking (2007) for a survey of the various notions of natural kinds.

[5] Charland (2002) argues that the category 'emotion' is a natural kind because all emotions have the property of being 'felt affective states', which Griffiths (2002) criticises as being too broad.

Nevertheless, this isn't to say that we can't have a science of emotion. Out of the three sorts of things picked out by the category 'emotion', basic emotions are likely candidates for natural kinds. For Griffiths, basic emotions are what Tomkins (1962) calls "affect-programs": roughly innate neural circuits responsible for the short-lived stereotypical physiological changes which constitute our emotional responses. They include changes in facial expression, the autonomic nervous system, arousal etc.[6] According to Griffiths, there is a distinct affect-program for (roughly) each of the emotions which feature in Ekman's (1973) list of basic emotions. There is an affect-program for fear, anger, surprise, happiness, sadness and disgust.[7] Moreover, since each basic emotion can be characterised by a specific underlying causal mechanism, these emotions are natural kinds.

This is the barebones version of the argument. Griffiths's original argument, in addition, tells us precisely how we should group things together on the basis of sharing an underlying causal mechanism. Since causal mechanisms in nature can show considerable variability at the individual level, such mechanisms should be grouped together not on the basis of possessing certain essential properties, but rather on the basis of manifesting various kinds of causal resemblances. The background assumption here is that natural kinds, in biology and the social sciences, are what Boyd (1989, 1991, 2010) calls 'Homeostatic Property Cluster Kinds': objects whose properties clusters together more or less reliably due to one or more causal homeostatic mechanisms.[8] As Craver (2009: 578) clarifies, Boyd's use of the term 'homeostatic' is potentially misleading as he has a weaker conception in mind than the standard one: "A mechanism is homeostatic in Boyd's sense if the mechanism explains the regular co-occurrence of phenomenal properties in the cluster." Craver critiques this account on grounds that there is no objective non-

---

[6] Other proponents of basic emotions posit basic emotion "systems" (Panksepp and Watt 2011) or "mechanisms" (Levenson 2011) which are similar to affect-programs. See Scarantino (2015) for a discussion.

[7] Ekman's (1999) revised list is more inclusive, and contains emotions such as guilt, contentment and relief.

[8] Also see Kornblith (1993), Wilson *et al* (2007) and Samuels and Ferreira (2010).

interest relative way of individuating mechanisms, which in turn means there will be multiple (incompatible) answers to the question of whether a given phenomenon counts as a natural kind.[9]

Griffiths's account mitigates against at least some of Cravers's concerns. First, Griffiths concedes that what counts as a natural kind will be domain-specific, e.g. something that counts as a natural kind in nosology is unlikely to offer property clusters adequate for the inductive and explanatory pursuits of physics. Second, Griffiths goes beyond Boyd's conception of Homeostatic Property Cluster Kinds by offering us a way of individuating the homeostatic mechanisms relevant for emotion. For Griffiths, the causal resemblances which replace essences are best captured by employing the notion of homology (shared ancestry) in evolutionary developmental biology. In comparison to analogies (shared function), classifications based on homologies are supposed to be "deep": even when the function has been transformed, there is more convergence in the underlying causal mechanisms, as homologues descend from a common ancestral form. For this reason, natural kinds should be classified on the basis of homology, and not analogy, as it is classifications of the former variety which can provide us with resembles in causal mechanisms required for the explanatory and inductive practices of the sciences. Doing so tells us that only affect-programs (i.e. basic emotions) count as natural kinds, as only they can be grouped together on the basis of having homologous traits. The empirical evidence for this tends to focus on fear. As Griffiths notes, "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" (2002: 238).

Whether basic emotions, e.g. fear, actually are natural kinds is a source of controversy. However, this is an issue which tends to be lumped under the even more controversial topic of

[9] I am grateful to xxxx for emphasising this point. Similar worries about individuating mechanisms are discussed by Bechtel (2008), Machamer *et al* (2000), and Illari and Williamson (2012), and the problem this raises for the natural kind debate is also discussed by Taylor (2018).

whether any basic emotions exist at all.[10] In what follows, I offer a critique of the view that fear is a natural kind whilst steering clear of the thorny issue of whether it also manifests the salient features of a basic emotion.

2. Why Fear isn't a Natural Kind

Basic emotions are (supposedly) natural kinds because they share not just certain surface properties — e.g. certain functional and possibly phenomenal profiles — but an underlying causal mechanism as well. This is taken to be evident in the emotion fear. There is a great body of work in neurobiology to suggest that our fear responses are triggered by subcortical regions of the brain, especially the amygdala. The neuroscientist LeDoux, in particular, is often credited with discovering that the amygdala is the "source" of fear.

A clarification. In summing up his research, LeDoux notes that the "amygdala has been tagged as the hub of the emotion of fear" (1996: 168). This quote is often misunderstood. As LeDoux himself notes, talk of identifying mental capacities with specific brain regions is a hangover from the time when we could only study brain functions by the effects of brain lesions in specific areas of the brain. Claims about the "hub" or "source" of fear, then, are really to be understood as claims about the neural circuitry which generate our fear responses. LeDoux's body of work provides good empirical grounds to show that the neural circuitry responsible for generating our fear responses involve subcortical regions of the brain, especially the amygdala.

It is precisely this, and related work, which is taken up by proponents of the basic emotion theory, e.g. Griffiths (1997), to argue that basic emotions form natural kinds. Fear is a natural kind because it turns out that there is, as a matter of fact, a specific affect-program for fear. More specifically, our fear responses are triggered by specific innate subcortical neural circuits which

[10] E.g. see Barrett (2006, 2017), Barrett *et al* (2007), and Warnick, LaPorte and Kalueff (2011) for scepticism about basic emotions *vis-a-vis* natural kinds. See Ortony and Turner (1990) for a critique of basic emotions more generally, and Panksepp and Watt (2011) along with Scarantino and Griffiths (2011) for a reply.

concern the amygdala. The hope is that a future neurobiology will deliver specific affect-programs for other basic emotions as well.[11]

The problem is that LeDoux has thrown a wrench in the works. This problem isn't due to the discovery of any new, contradicting, data from neuroscience. Rather, it stems from the way the original empirical data has been interpreted. The fault, according to LeDoux, lies in the way scientists, including himself, talk about their research:

> In retrospect, I now believe that it was a mistake to use the expression "fear system" to describe the role of the amygdala in detecting and responding to threats, and also erroneous to talk about fear stimuli and fear responses in this context. (LeDoux 2016: 36)

This mistake is owing to the misleading nature of the terminology. According to LeDoux, when we use mental state terms to describe the function of brain circuits, we inflict the data with what he calls a "surplus meaning": roughly, we attribute psychological — specially phenomenal — properties to circuits which they lack. For LeDoux, what he's been working on all along is how the brain responds to threat. The problem with describing the circuits responsible for defensive behaviours as "fear systems" is it suggests (incorrectly) that it is the very same system which generates the conscious experience of fear; or in his terminology, the "subjective feeling" of fear. However, "feelings of fear or anxiety are not products of circuits that control defensive behavior" (LeDoux and Pine 2016).[12]

Much of LeDoux's recent work consists in pointing out this error. His solution is to use emotion terms for the conscious experience of emotion. With regards to 'fear', he thinks this term should denote the subjective feeling of fear, and we should instead use "threat circuits" or

---

[11] Scarantino and Griffiths (2011) note that our final list of basic emotions won't, in all likelihood, correspond to our folk categories, e.g. anger, joy, sadness etc.

[12] Several empirical findings support this claim, including Ohman (2002), Tamietto and Gendler (2010), Bornemann *et al* (2012), and Feinstein *et al* (2013). See LeDoux (2016, 2017, 2019) and LeDoux and Hoffman (2018) for a discussion.

"survival circuits" to refer to those neural mechanisms which trigger our defensive responses. This revisionist project has generated controversy, with several neuroscientists arguing that it is good scientific practice to use the term "emotion" to refer to internal brain states which are measurable via changes in facial expression, physiology and behaviour, and "to distinguish emotion states from conscious experiences of those emotions" (Adolphs and Anderson 2018: 129). We needn't weigh in on this debate here. The point is that LeDoux's clarification problematises the argument employed to show that basic emotions form natural kinds.

As now evident, the crux of the problem concerns the conscious experience of emotion. The physiological changes which constitute the affect-program responses don't involve conscious experience. Nevertheless, for Griffiths, the affect-program theory "can (and should) incorporate emotional feelings" (1997: 121). Griffiths himself doesn't provide us with any details about how this is to take place, but assumes that it is unproblematic to do so. The reason being that we can use emotion terms, like "fear","sadness", "joy", "surprise", "anger", and "disgust", to refer to affect-program responses "because the new categories coincide more or less well with the occurrent, phenomenologically salient instances of these traditional categories" (1990: 189).

Nevertheless, it is precisely this assumption which is now called into question. If LeDoux is right, the neural circuits which generate our affect-program responses come apart from those responsible for the conscious experience of emotion. To be clear, he grants that the threat circuitry can modulate our emotional experiences in various ways. However, the experiences themselves are the products of a distinct neural circuitry which can be activated without the activation of the threat circuitry, and vice versa. Subsequently, there is no longer a guarantee that the affect-program responses "coincide" with our emotional feelings.

We are now in a position to revisit the argument for why basic emotions are natural kinds. Natural kinds are to be grouped together on the basis of a shared underlying causal mechanism. Basic emotions are natural kinds because our basic emotion categories refer to things which share an underlying causal mechanism. For instance, fear is a natural kind because the kinds of things

picked out by the category 'fear' all share a certain underlying causal profile. This argument, however, has a false premise. The category 'fear' categorises things based on both physiological and phenomenological surface properties. This, in itself, isn't a problem. However, it turns out that the causal mechanisms that give rise to the physiological properties come apart from those responsible for the phenomenal ones. *Ergo*, the category 'fear' isn't a natural kind, as it doesn't refer to things with share the relevant underling causal mechanism.

3. A Worry

It is tempting to resist this argument on the basis of the belief that all instances of fear *must* share a common core. For instance, it could be argued that fear is a natural kind on account of the two aforementioned mechanisms having certain features in common. In this section, let me make some clarifications in order to preempt any objections along these lines.

To begin, it is important to acknowledge that there is controversy over the extent to which the survival circuitry comes apart from those which give rise to the conscious experience of emotion. Some neuroscientists, including LeDoux, take current neuro-cognitive data to show that the survival circuitry is neither necessary nor sufficient for emotional experience, whereas others take the line that the survival circuitry is necessary but not sufficient for such experience.[13] The exception is Panksepp (1998) who argued that activation in the subcortical regions of the brain which give rise to our defensive responses also underly the conscious experience of emotion.[14]

We can put Panksepp aside because recent discoveries in neurobiology put pressure on the claim that the activation in the subcortical regions of the brain, which underly our physiological responses, suffice for the conscious experience of emotion, at least in humans. The point of contention in contemporary neuroscience is over whether activation in these regions is necessary

[13] E.g. see Scarantino (2018), and Adolphs and Anderson (2018).

[14] This argument is premised on the (contentious) assumption that fear behaviours triggered by subcortical stimulation in decorticate animals must result from underlying affective experiences, *viz.* feelings of displeasure (e.g. see Panksepp 2011: 2).

for the conscious experience of emotion. This is an empirical question, an open one at that, to which I won't take a stand in this paper.

Now suppose the activation of certain neural circuits in the subcortical regions turn out to be necessary for the conscious experience of emotion. The world turning out this way won't, despite appearances, undermine the present argument. This is because simply sharing certain properties — even certain causal mechanisms — won't by itself suffice to categorise things as being of the same natural kind.

To clarify, Griffiths's argument for why emotions don't form a natural kind rests on two claims. First, as we noted earlier, the category 'emotion' refers to three distinct kinds of things — basic emotions, complex emotions and social pretences — with very different underlying causal natures. Opponents argue that complex emotions are of the same kind of emotion as basic emotions since both types of emotion share a basic emotion core. This brings us to the second claim. Griffiths (2002) concedes that complex emotions may have basic emotions as parts, but denies that the former are blends or elaborations of the latter. Simply having a necessary feature in common isn't enough to group things together as a natural kind. By analogy, silicon and oxygen are essential components of both jadeite and nephrite but this doesn't suffice for these two minerals to fall into the same natural kind category in the earth sciences.

What is missing is an account of how we are precisely to group things together on the basis of sharing a causal mechanism. An obvious suggestion is that natural kinds are categories whose members share *enough* properties, e.g. they have the right amount of underlying causal features in common. But to offer this suggestion is to fail to appreciate the non-essentialist nature of the conception of natural kinds at play in the present dialectic. On an essentialist notion of natural kinds, such kinds have essences: roughly necessary (and perhaps sufficient) properties which all members of the kind share. According to Boyd (1989, 1991, 2010), the kinds of entities we study in the special sciences are unlikely to have a set of essential properties in common. Griffiths, following Boyd, uses the category 'species' to illustrate this point: "it has been accepted that

natural populations are pools of variation, that they are continually evolving, and consequently that it is futile to try to define species in terms of a common intrinsic structure" (2007: 14). This is remedied, he argues, by categorising members of a species based on patterns of ancestry and descent. *Mutatis mutandis* for the category 'emotion'.

To elaborate, Griffiths's "psychoevolutionary" approach to emotion is, at heart, a project which aims to draw on evolutionary developmental biology to inform a psychological science of emotion. To that end, emotions are natural kinds provided they can be categorised on the basis of a shared ancestry; in other words homology. Affect-programs, arguably, are natural kinds in this respect. Our affect-program responses are triggered by subcortical regions of the brain, which we share with our hominin ancestors, and possibly other mammals as well. What proves controversial, however, is whether emotions themselves can be given the same treatment.

On the conception of homology Griffiths favours, a homologue is "The same organ in different animals under every variety of form and function" (Owen 1843: 374). The subcortical regions of the brain aren't the same organ as the neocortex under different varieties of form and function. Rather, they are distinct organs altogether, with distinct evolutionary histories. (The neocortex is newer, hence the name). For this reason, the neocortex-involving circuitry responsible for generating the conscious experience of emotion — whether or not it also involves the subcortical regions — aren't technically homologues of our survival circuitry. Categorisations based on homology must, therefore, treat these as distinct categories.

In summary, affect-programs responsible for generating our fear (threat) responses, arguably, are natural kinds. These programs may also be necessary for generating the conscious experience of fear, in which case these experiences and our threat responses would share a common causal core. Nevertheless, this is beside the point. The vernacular category 'fear', like the broader category 'emotion', picks out things with very different causal mechanisms when

individuated at the level of homology. Subsequently, the very reasons Griffiths provides for demonstrating that emotion isn't a natural kind tell against fear being a natural kind as well.[15]

4. The Unity of Emotions

If what I have argued thus far is correct, does this mean that the category 'fear' doesn't refer to a unified class of mental phenomena? Griffiths (2002) is at pains to point out that the question of whether emotions are a natural kind is a different question from whether the concept 'emotion' can be given a univocal analysis. He thinks the latter is possible, and only takes his project to show that things which fall under the vernacular category 'emotion' don't form a natural kind for the purposes of scientific inquiry. I end by making the very same point here with regards to the category 'fear'.

Things which fall under the vernacular category 'fear' don't form a natural kind for the purpose of scientific inquiry. LeDoux's broader revisionist project aims to employ the term 'fear' exclusively for the conscious experience of fear, whereas others argue that 'fear' should refer to whatever internal brain states trigger our defensive responses.[16] Far be it from me to tell scientists how to use their terminology. The point is, we can only have a science of fear once we make some adjustments to how we employ the term 'fear'.

LeDoux's clarifications raise a similar challenge for a philosophy of emotion. Should a philosophical analysis of 'fear' refer to the conscious feeling of fear or whatever underlies our physiological defensive responses? What is worth stressing, however, is that the constraints for categorisation in philosophy needn't be as stringent as those in the natural sciences. In particular, philosophical analyses needn't run deep: we needn't analyse things in terms of underlying causal mechanisms. Subsequently, we needn't decide whether fear is fearful feelings or brain states that

---

[15] This is compatible with fear being a natural kind for proponents of the New Basic Emotion Theory, but precisely because they revise the term 'basic emotion' to refer to affect-programs, e.g. see Scarantino (2018: 78).

[16] See Scarantino (2014, 2018), Adolphs and Anderson (2018), and Adolphs and Andler (2018).

cause our defensive responses on the basis of these features having different underling causal mechanisms. 'Fear' can be analysed in a way that is inclusive of both of these features.

By way of illustration, the folk concept 'jade' picks out two distinct kinds of minerals but it can still be given a univocal analysis, e.g. 'a green colour mineral(s) used as an ornament'. Likewise, according to Griffiths, our folk concept has it that "emotion is a putative psychological category of motivational states that exhibit passivity" (1997: 246). Nothing I have said about fear failing to be a natural kind precludes fear, or other basic emotions for that matter, from being given an analysis along these lines. For instance, 'fear' might be 'motivational states — with either physiological or phenomenological surface profiles — that exhibit passivity, and which are specifically responses to danger'. Present-day philosophy of emotion aims at these very sorts of univocal analyses of both the broader category 'emotion' and specific instances of the category, like 'fear'. This remains a viable project, i.e. so long as we don't expect philosophical analyses to mirror the categorisation practices in the natural sciences.

To my mind, what's really significant about the discovery that fear isn't a natural kind isn't any consequences that ensue for the unity of fear *per se*. Rather, it's the cautionary note it provides for an empirically-driven philosophy of emotion. Namely, it is no longer enough to draw on neuro-cognitive data concerning fear. We need to know exactly what this data is about: whether it concerns the conscious feeling of fear, our threat responses or both. Given the history of surplus meaning of 'fear' in neurobiology, this won't be easy.

REFERENCES

- Adolphs, R. and Anderson, D. J. (2018). *The Neuroscience of Emotion: A New Synthesis.* Princeton and Oxford: Princeton University Press.
- Adolphs, R. and Andler, D. (2018). 'Investigating Emotions as Functional States Distinct from Feelings'. *Emotion Review* 10(3): 191-201.

- Barrett, L. F. (2006). 'Are Emotions Natural Kinds?'. *Perspectives on Psychological Science* 1(1): 28-58.
- Barrett, L. F (2013). 'Psychological Construction: The Darwinian Approach to the Science of Emotion'. *Emotion Review* 5(4): 379–389
- Barrett, L. F. (2017). *How Emotions Are Made: The Secret Life the Brain*. New York: Houghton-Mifflin-Harcourt.
- Barrett, L. F., Lindquist, K. A., Bliss-Moreau, E., Duncan, S., Gendron, M., Mize, J., and Brennan, L. (2007). 'Of mice and men: Natural kinds of emotion in the mammalian brain?' *Perspectives on Psychological Science* 2: 297–312.
- Barrett, L. F., and Russell, J. A. (eds.) (2014). *The Psychological Construction of Emotion*. Guilford Press.
- Bechtel, W. (2008). *Mental mechanisms*. Sussex: Psychology Press.
- Bornemann, B., Winkielman, P., and van der Meer, E. (2012). 'Can You Feel What You Do Not See? Using Internal Feedback to Detect Briefly Presented Emotional Stimuli'. *International Journal of Pschyophysiology: Official Journal of the International Organization of Psychophysiology* 85: 116-24.
- Boyd, R. (1989). 'What Realism Implies and What it Does Not'. *Dialectica* 43 (1–2):5-29.
- Boyd, R. (1991). 'Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds'. *Philosophical Studies* 61 (1):127-148.
- Boyd, R. (2010). 'Realism, natural kinds, and philosophical methods'. In He. Beebee & N. Sabbarton-Leary (eds.), *The Semantics and Metaphysics of Natural Kinds*. Routledge. pp. 212--234.
- Brigandt, I. (2003). 'Species Pluralism Does Not Imply Species Eliminitivism'. *Philosophy of Science* 70: 1305–1316.
- Charland, L. (1995). 'Emotion as a natural kind: Towards a computational foundation for emotion theory'. *Philosophical Psychology* 8: 59–84.

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

- Craver, C. F. (2009). 'Mechanisms and Natural Kinds'. *Philosophical Psychology* 22(5): 575-594.
- Deonna, J. A. and Teroni, F. (2012). *The Emotions: A Philosophical Introduction*. London and New York: Routledge.
- Ekman, P. (1973). *Darwin and Facial Expression*. New York: Academic Press.
- Ekman, P. (1999). 'Basic Emotions'. In T. Dalgleish and M. Power (eds.), *Handbook of Cognition and Emotion*. Chichester: John Wiley and Sons, Co.: 45–60.
- Faucher, L. (2013). Comment: Constructionisms? *Emotion Review* 5 (4):374-378.
- Feinstein J. S, *et al.* (2013) 'Fear and panic in humans with bilateral amygdala damage'. *Nature Neuroscience* 16(3): 270–272.
- Goodman, N. (1954). *Fact, Fiction and Forecast* (1 ed.). London: Athlone Press, University of London.
- Griffiths, P. E. (1990). 'Modularity, and the psychoevolutionary theory of emotion'. *Biology and Philosophy* 5(2): 175-196.
- Griffiths, P. E. (1997). *What Emotions Really are: The Problem of Psychological Categories*. Chicago and London: Chicago University Press.
- Griffiths, P. E. (2002). 'Is emotion a natural kind?' In R. C. Solomon (ed.), *Thinking About Feeling: Contemporary Philosophers on Emotions*. New York: Oxford University Press.
- Griffiths, P. E. (2003). 'Basic Emotions, Complex Emotions, Machiavellian Emotions'. *Royal Institute of Philosophy Supplement* 52:39-67.
- Griffiths, P. E. (2004). 'Emotions as natural and normative kinds'. *Philosophy of Science* 71(5): 901-911.
- Griffiths, P. E. (2007). 'Evo-Devo Meets the Mind: Towards a Developmental Evolutionary Psychology'. In R. Sansom and R. Brandon (eds.), *Integrating Evolution and Development: From Theory to Practice*. Cambridge, MA: MIT Press: 195-225.
- Hacking, I. (2007). 'Natural Kinds: Rosy Dawn, Scholastic Twilight'. *Royal Institute of Philosophy Supplement* 61: 203-239.

- Illari, P. M., & Williamson, J. (2012). 'What is a mechanism? Thinking about mechanisms *across* the sciences'. *European Journal for the Philosophy of Science* 2: 119–135.
- Kornblith, H. (1993). *Inductive inference and its natural ground: An essay in naturalistic epistemology*. Cambridge, MA: MIT Press.
- LeDoux, J. E. (1996). *The Emotional Brain*. Simon and Schuster.
- LeDoux, J. E. (2012). 'Rethinking the emotional brain'. *Neuron* 73:653–676
- LeDoux, J. E. (2016). A*nxious: Using the Brain to Understand and Treat Fear and Anxiety.* Viking.
- LeDoux J. E, and Pine, D. S. (2016). 'Using neuroscience to help understand fear and anxiety: A two-system framework'. *American Journal of Psychiatry* 173(11): 1083–1093.
- LeDoux, J. E (2017). 'Semantics, surplus meaning, and the science of fear'. *Trends in Cognitive Science* 21(5): 303-306.
- LeDoux, J. E. (2019). *The Deep History of Ourselves: The Four-Billion Year Story of How We Got Conscious Brains*. New York: Viking.
- LeDoux, J. E., and Hoffman, S. G. (2018). 'The subjective experience of emotion: a fearful view'. *Current Opinion in Behavioral Sciences* 19: 67–72.
- Levenson, R. W. (2011). 'Basic Emotion Questions'. *Emotion Review* 3(4): 379–386.
- Machamer, P., Darden, L., & Craver, C. (2000). 'Thinking about mechanisms'. *Philosophy of Science,* 67: 1–25.
- Ohman, A. (2002). 'Automaticity and the amgydala: nonconscious responses to emotional faces'. *Current Directions in Psychological Science* 11:62-66.
- Ortony, A., and Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review* 97: 315–331.
- 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*. New York: Oxford University Press.
- Panksepp, J. (2011). 'Cross-Species Affective Neuroscience Decoding of the Primal Affective Experiences of Humans and Related Animals'. *PLOS One* 6(9): 1-15.
- Panskepp, J., and Watt, D. (2011). 'What is basic about basic emotions?: Lasting lessons from affective neuroscience'. *Emotion Review* 3(4): 387–396.
- Russell, J. A. (1991). Culture and the categorization of emotion. *Psychological Bulletin* 100: 426–50.
- Russell, J. A. (2003): 'Core affect and the psychological construction of emotion'. *Psychological Review* 110:145-172.
- Samuels, R., and Ferreira, M. (2010). Why *don't* concepts constitute a natural kind? *Behavioral and Brain Sciences* 33: 222–223.
- Scarantino, A. (2015). 'Basic Emotions, Psychological Construction and the Problem of Variability'. In J. Russell and L. F. Barrett (eds.), *The Psychological Construction of Emotion.* New York and London: Guilford Press: 334-376.
- Scarantino, A. (2018). 'Are LeDoux's Survival Circuits Basic Emotions Under a Different Name?' *Current Opinion in Behavioral Sciences* 24: 75-82.
- Scarantino, A., and Griffiths, P. (2011). 'Don't give up on basic emotions'. *Emotion Review* 3(4): 444–454.
- Tamietto, M. and de Gelder, B. (2010). 'Neural bases of the non-conscious perception of emotional signals'. *Natural Reviews Neuroscience* 11(10): 697-709.
- Taylor, H. (2018). 'Emotions, Concepts and the Indeterminacy of Natural Kinds'. *Synthese* 10.1007/s11229-018-1783-y.
- Tomkins, S. S. (1962). *Affect, Imagery, Consciousness*. New York: Springer.
- Warnick, J., LaPorte, J., and Kalueff, A. V. (2011). 'Domain interplay in mice and men: New possibilities for the "natural kinds" theory of emotion'. *New Ideas in Psychology* 29*:* 49–56.

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