# The New LeDoux: Survival Circuits, and the Surplus Meaning of 'Fear'

Raamy Majeed
University of Auckland
*Philosophical Quarterly* (forthcoming)

ABSTRACT: LeDoux's (1996) pioneering work on the neurobiology of fear has played a crucial role in informing debates in the philosophy of emotion. For example, it plays a key part in Griffiths's (1997) argument for why emotions don't form a natural kind. Likewise, it is employed by Faucher and Tappolet (2002) to defend pro-emotion views, which claim that emotions aid reasoning (de Sousa 1987, Damasio 1994). LeDoux, however, now argues that his work has been misread (2012, 2016, 2017, 2019). He argues that using emotion terms, like 'fear', to describe neuro-cognitive data adds a "surplus meaning": it attributes phenomenal properties to survival circuits which they don't possess. This paper aims to explore LeDoux's new proposal, and examine the potentially devastating consequences that ensue for the aforementioned views. I end by addressing the worry that these lessons are conditional on LeDoux's own higher-order theory of emotional consciousness being true.

KEYWORDS: emotion; consciousness; fear; LeDoux; natural kinds; attention

## 1. Introduction

Neuro-cognitive data concerning the way the brain processes and responds to threat, especially LeDoux's pioneering work outlined in *The Emotional Brain: The Mysterious Underpinnings of Emotional Life* (1996), has played a crucial role in informing debates in the

philosophy of emotion. Much of the philosophical draw here centres on what this data tells us about the neurobiology of emotion, especially how the brain generates fear.

LeDoux's work has proved relevant because he is often credited with discovering that the amygdala is the "source" of fear. Two clarifications. First, as I understand it, this claim is not intended to convey that the amygdala is merely one of the causal factors that contribute to fear. Rather, the claim is something more substantial, i.e. the amygdala is the central underlying causal mechanism responsible for fear. Second understood this way, the claim should not be taken literally. As LeDoux (2016) clarifies, brain functions are to do with brain systems, involving neural circuits, not brain areas. (The talk of brain areas is a hangover from the time when we could only study brain functions based on the effects of brain lesions in specific areas). So claims about the "source" of fear are best understood as claims about the neural circuitry that generate fear. Thus understood, current neuro-cognitive data, including that of LeDoux, is said to provide us with good empirical grounds to suppose that the neural circuitry that generate fear concern subcortical regions of the brain, especially the amygdala.

Another feature of relevance concerns the different ways in which the brain generates fear. Here LeDoux's work is understood to have demonstrated that there are in fact two distinct neural circuits involved in generating fear responses: (i) a thalamus-to-amygdala circuit, which bypasses the cortex, is 'quick and dirty', and occurs without the conscious experience of the stimulus, and (ii) a thalamus-to-cortex-to-amygdala circuit, which is slow, and occurs with the conscious experience of the stimulus. These interpretations of LeDoux's findings, and similar interpretations of other neuro-cognitive data concerning

the amygdala, have seeped into the philosophical literature, and are employed to further various theoretical ends.

Perhaps this is most evident in the evolutionary developmental approach to emotion pioneered by Griffiths (1997). In brief, Griffiths reads the existence of the quick and dirty route to emotion-generation as evidence for what the psychologist Tomkins (1962) described as "affect programs": roughly, an innate neural circuitry, passed down from our ancestors, possibly having certain mechanisms in common with some other animals, and which are responsible for our emotional responses.[1] Drawing on the work of Ekman (1972, 1973, 1980), Griffiths uses this notion of affect programs to argue both for the existence of basic emotions and against the idea that the things picked out by our vernacular category of emotion form a natural kind. Very roughly, basic emotions are those generated by affect programs, and 'emotions' don't refer to a natural kind for it picks out cognitively complex emotions and certain kinds of social pretences, as well as the basic emotions generated by the affect programs.[2]

LeDoux's work has also found a new lease on life in recent arguments for the pro-emotion thesis: the view that emotions aid practical reasoning. There are a series of distinct views which fall under the pro-emotion umbrella, but most share the assumption that emotions aid reasoning by modulating attention. The original proponents of this view, e.g. de Sousa (1987) and Damasio (1994), argue that emotions help with decision-making by

---

[1] In earlier work, Griffiths (1997) relies on various neuro-cognitive data on threat circuitry but doesn't cite LeDoux, whereas he makes explicit reference to LeDoux (1996) in subsequent work, e.g. see Griffiths (2002, 2004a).

[2] For Griffiths, complex emotions, roughly, are those that involve responding in "a more cognitively complex way to more highly analyzed information" (2002: 394).

drawing our attention to some response-options over others. This hypothesis has well-known problems, a major one being that it arguably doesn't stand up to empirical scrutiny.[3] Similar, though crucially distinct, pro-emotion theses, however, have been proposed, and with empirical backing. In their survey paper, Faucher and Tappolet (2002), for instance, point to a series of findings, including LeDoux (1996), to argue that emotions, in particular fear and anxiety, modulate our attention of visual stimuli, and do so in ways that might aid certain developmental ends.

The trouble is, such philosophical implications of the neuro-cognitive data possibly stem from a misunderstanding. In recent work, LeDoux (2012, 2016, 2017, 2019) argues that this data has been misread owing to the way scientists talk about their research. He takes his own work as an example:

> 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)

Crucially, the mistake stems not from new and contradicting empirical findings, but rather from the way scientists talk about, and sometimes conceptualise, the existing data. In particular, it stems from using mental state terms to describe the function of brain circuits, which in turn inflicts the data with a "surplus meaning": roughly, the attribution of

[3] The majority of critiques focus on Damasio, e.g. see Evans (2002), Dunn, Dalgleish, and Lawrence (2006), Gerrans (2007), Linquist and Bartol (2013), and Bartol and Linquist (2015). But see Ransom (2016) for a challenge directed specially at de Sousa. Also see AUTHOR (XXXX) for a discussion of the differences between de Sousa and Damasio's proposals, as well as a brief response to Ransom.

psychological properties to these circuits which they don't posses. With respect to the neural circuitry that deals with threat, describing it as a "fear system" suggests (incorrectly) that it is the same as the circuitry that underly the conscious experience of fear.

Philosophers who typically take conscious experience to be an essential component of emotion, then, seem to be committing an error when they rely on LeDoux, and similar empirical research, to proffer their preferred theories of emotion.[4] This paper aims to investigate this error. In what follows, I disentangle the neuro-cognitive data which is at the heart of LeDoux's diagnosis from the broader revisionist project, i.e. of reframing scientific terminology, which he offers as a remedy (§2). I then reapply the lessons from the data, free of the erroneous readings, to two areas where they have proved crucial: the debate on whether emotions form a natural kind (§3), and pro-emotion views which state that emotions aid reasoning by modulating attention (§4). Finally, I end by addressing the worry that these lessons are conditional on LeDoux's own higher-order theory of emotional consciousness being true (§5).

**2. The Surplus Meaning of 'Fear'**

[4] The view that emotional experience is essential to emotion is the assumed view in current philosophy, e.g. see Deonna and Teroni (2012), but is also prevalent amongst scientists, including Freud (1915), Clore (1994), Frijda (1999), Russell (2003) and Barret (2009). For a rival view, see Prinz (2004).

LeDoux's recent work on rethinking the emotional brain can be summed up as follows. The use of emotion terms to describe various brain systems attributes to these systems psychological properties which they don't possess; in particular, it attributes to them the conscious subjective feelings that typically accompany our emotional responses. Fear is a central example. Describing the defensive survival circuitry which triggers our threat responses as "fear circuits" attributes to them, amongst other things, the conscious, subjective feelings of fear. This is misleading because, though the survival circuitry influences our conscious feelings of fear, the circuitry that underly such feelings is something altogether distinct. One way to correct this is to reserve emotion terms for subjective feelings.

Note that there are four distinct claims being made here, which are all part of LeDoux's revisionary project: (i) a claim about a misunderstanding in the literature, (ii) the related empirical claim about what justifies this reading, (iii) a diagnosis of what causes the misunderstanding, and (iv) a recommendation for how to preempt future misunderstanding. Consider the example of fear:

> The vernacular meaning of emotion words is simply too strong. When we hear the word 'fear', the default interpretation is the conscious experience of being in danger, and this meaning dominates. For example, although I consistently emphasized that the amygdala circuits operate nonconsciously, I was often described in both lay and scientific contexts as having shown how feelings of fear emerge from the amygdala. Even researchers working in the objective tradition sometimes appear confused about what they mean by fear. (LeDoux 2017: 303)

The neurobiological account of how the brain processes threat, as outlined in LeDoux (1996) and elsewhere, concerns how certain defensive survival circuits in the brain respond to threatening stimuli. (i*) This account is read by philosophers and many in the scientific community as an account concerning the emotion fear, where 'fear' picks out, amongst other things, the conscious feeling of fear. (ii*) This reading, however, is problematic because although the threat circuitry modulates the kinds of fearful feelings we experience, actual conscious experiences of fear are caused by a distinct neural system. For example, the cortical circuits, which give rise to the subjective feelings of emotion, need not always be activated when the threat circuitry triggers its defensive responses. (iii*) The misreading stems from the way LeDoux, as well as others, describe the threat circuity: employing terminology like the "fear system" inflicts the neuro-cognitive data with a surplus meaning, i.e. we treat the system not just as that which triggers our defensive responses but also the subjective feeling of fear. (iv*) A way to avoid this confusion is to reserve the term 'fear' for the conscious feeling of fear.

LeDoux (2012, 2016) is at pains to note that this revisionist project aims not to redefine or explain emotions but to provide a way to move forward in a scientific study of our survival circuitry without incurring the aforementioned kinds of confusion. Nevertheless, in future work, he also appears to endorse a stronger position: "subjective emotional experience, the feeling, is the essence of an emotion" (LeDoux and Hoffman 2018: 67). The former is a pragmatic position about scientific practice, whereas the latter, intended or not, amounts to a substantive metaphysical hypothesis. Both claims are important and worthy of attention but will be kept aside, as the the philosophical implications which we will be concerned with in this paper can be brought out by simply

focussing on claims (i), (ii) and (iii). The philosophical confusions that result from misreading the empirical data are best seen by example, which will be the topic of the next two sections, but for the remainder of this section, let me say a bit more about the data itself and how it has been misread.

In very simple terms, the data is misread because it doesn't speak to emotional experience while it is interpreted as doing just this. Now, there are well-known reasons why we think that third-person scientific methodology cannot give us direct access to first-person introspectable conscious experience.[5] The problem that concerns us is not anything near as grand. The problem for us isn't that we cannot directly read off claims about emotional experience from neuro-cognitive data, but that the specific data in question doesn't give us anything resembling even a third-person science of emotional experience. To elaborate, LeDoux accepts that verbal reports give us adequate means to study conscious experience: "the hallmark of conscious states in humans is that they can be reported" (2016: 149). The problem is that the neural circuitry that underly our emotional responses are distinct from those responsible for our emotions experiences, which we can measure by verbal reports.

In the case of fear, this problem isn't really new. As LeDoux (2016) notes, the scientific press, e.g. *Nature*, *Science*, *Wired*, *Scientific American* and *Discover*, have all recently run stories which counter what they suppose to be orthodoxy, *viz.* that the amygdala is the source of fear — where 'fear' is intended to mean the feeling of fear. Based on findings by Feinstein *et al* (2013), fear, they report, can be experienced by a subject with bilateral amygdala damage. These reports, however, only have the shock-value they do on the basis

[5] See Nagel (1974), Jackson (1982) and Chalmers (1995).

of the surplus reading; when we suppose that the survival circuits that underly our threat responses, and for which the amygdala plays a key role, are also what constitutes the feeling of fear. What LeDoux is at pains to convey is that what these reports assume to be orthodoxy itself rests on a misunderstanding.

The surplus reading of the old empirical literature is mistaken because the survival circuits that trigger our defensive responses come apart from the neural circuitry that underly the conscious feeling of fear. The new findings, e.g. by Feinstein *et al* (2013), strengthens, rather than undermines, this. For LeDoux (2017), the main reason to suppose that these circuitries come apart has to do with the fact that it accounts for a whole host of puzzles in the field:

1) The feelings of fear don't correlate well with measures of behavioural and physiological defence responses.[6]
2) Patients with damage to the amygdala don't manifest the physiological defence responses but still feel fear.[7]
3) Threats processed unconsciously still trigger the amygdala and activate the defence responses even when the subject lacks the feeling of fear.[8]
4) The feeling of fear is not tied to a single subcortical circuit.[9]

---

[6] LeDoux (2016), LeDoux and Brown (2017), and LeDoux and Hoffman (2018).

[7] See Feinstein *et al* (2013) but also Anderson and Phelps (2002) for a discussion of the relationship between the amygdala and subjective experiences more broadly.

[8] Ohman (2002), Tamietto and Gelder (2010), and Bornemann *et al* (2012).

[9] LeDoux (2016, 2019), and LeDoux and Brown (2017).

5) Drug discovery based on the defence responses of animals comes up with medications that are more likely to alter behavioural tendencies than the feelings of fear.[10]

These findings undercut the presupposition that the amygdala, and survival circuits more broadly, are the main source of the conscious feeling of fear. One further point to note is that they do so without the need to assume a positive theory of what in fact underlies the conscious experience of fear. As we shall see in the final section, LeDoux does offer a positive account of emotional consciousness, which strengthens the existing reasons for a separation between survival circuits and those which underly the feelings of fear. Nevertheless, the proceeding philosophical implications of LeDoux's recent work can be drawn out without the need to buy into his more controversial views concerning emotional consciousness.

**3. The Natural Kind Debate**

In his seminal book, *What Emotions Really Are: The Problem of Psychological Categories*, Griffiths (1997) argues that emotions don't form a natural kind. As Griffiths (2002) explains, this matters because philosophers of emotion often take as their aims the genesis, development and consequences of a 'typical' emotion. If emotions don't form a natural kind, these projects are doomed from the onset for there won't be any such thing as a typical emotion.

[10] See LeDoux and Pine (2016), Pine and LeDoux (2017), and Fanselow and Pennington (2018) for a critique.

Griffiths's argument for this controversial position is broken into two parts. In the first part, he develops a "psychoevolutionary" approach to emotion, and in the second he provides an exposition of the notion of natural kinds at play in the natural sciences. Let us take each in turn.

In part one, Griffiths puts more meat on the bones of the the psychoevolutionary approach to emotion he develops earlier (1990), which he takes to be Darwinian in conception. Here, and in future work, he draws on Darwin (1859, 1872), but more so on the Darwin-inspired work on the facial expression of emotion developed by Ekman (1972, 1973, 1980), to defend the existence of Tomkin's affect programs: "short-term, stereotypical responses involving facial expression, autonomic nervous system arousal, and other elements. The same patterns of response occur in all cultures and homologues are found in related species" (Griffiths 1997: 8). Griffiths, following Ekman, sometimes refers to these responses as 'basic emotions', and following Ekman's list, takes them to include anger, fear, happiness, sadness, surprise and disgust.[11]

While Tomkins regarded affect programs as being hypothetical, Griffiths, along with Ekman, endorses a realist position which rests on an argument from abduction. In brief, affect program responses possess certain characteristics: they are both complex and coordinated, occur very quickly after the stimulus is registered, and are typically involuntary. These features can be best explained by the existence of "a single neural program that is triggered by the stimulus and controls the various elements of the

[11] Ekman's (1999) updated list includes 16 basic emotions.

unfolding response" (pg. 84). In subsequent work (2004a), Griffiths cites LeDoux (1996) in support for there being a neural basis for the affect programs concerning fear.[12]

The other significant development in part one concerns the contrast between basic emotions and other kinds of responses that fall under the vernacular category of emotion. These include more cognitively-involving emotions or what he laters calls "complex emotions" (2002), e.g. jealousy, and various forms of social pretences, like love. Crucially, Griffiths argues that these kinds of responses don't share enough similarities to be unified under one emotion banner. For example, whilst the affect program responses may affect complex emotions, these two kinds of responses come apart: affect programs can be triggered *sans* the complex emotions, and *vice versa*. The affect programs, *ergo*, can't be said to underly all of the kinds of things picked out by our ordinary emotion terms.

In part two, Griffiths develops what he takes to be the notion of natural kinds at play in the natural sciences: kinds which are useful for explanation and induction. This notion of natural kinds is distinct from that employed in metaphysics, i.e. the kinds of things which 'carve nature at its joints'. Instead, drawing on Goodman (1954) and Boyd (1991), Griffiths endorses a "theory view" of kind terms, the basic idea being that our categorisation of things into kinds aims to locate "projectable properties", roughly a

[12] Or more accurately, Griffiths takes LeDoux (1996) to have confirmed a twin-pathway neurobiological account of fear, where the first pathway lends support for the existence of affect programs concerning fear.

"cluster of correlated properties and that we can offer some defeasible justification for extrapolating those correlations" (Griffiths 2004b: 906).[13]

Natural kinds in this sense are supposed to be domain specific. A natural kind in nosology, for example, might not offer the kinds of property clusters that would allow for induction and explanation in physics. That said, much of Griffiths's discussion of the natural kinds that should be at play in psychology, and thereby of relevance for an investigation into emotion, is informed by the notion in evolutionary biology.

To elaborate, according to Griffiths, "the purpose of categorization in the sciences is to group together things which resemble one another in many different ways because of some underlying, similarity-generating mechanism" (1997: 16). It is the existence of such shared underlying mechanisms that offer (defeasible) justification for extrapolation. Moreover, a possible way to facilitate research into these underlying mechanisms is to categorise things using the notion of evolutionary homology (shared ancestry) in contrast to analogy (shared function). In more detail:

> 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. (Griffiths 2002: 396)

[13] In later work, Griffiths (2004b) opts to use Brigandt's (2003) term 'investigative kinds' in order to distance his account from other notions of natural kinds. See Hacking (2007) for an overview of the various conceptions of natural kinds.

For example, the wings of a fruit bat are analogous to those of a pigeon, as the resemblances between them are a product of being adapted for the same ecological role, *viz*. work amongst branches. However, these wings aren't homologous with each other because they don't result from the same ancestral form. What is important to note here is that *contra* classifications based analogy, those based on homology are 'deep'. Despite superficial differences, even when the function has been transformed, there is more convergence in the underlying mechanisms. For this reason, Griffiths argues that natural kinds should be classified on the basis of homology not analogy, for only the former provide us with resemblances of an underlying (causal) mechanism which is crucial for scientific explanation and induction. This holds true for natural kinds in evolutionary biology, but also for such kinds in psychology, and thereby informs the psychoevolutionary approach to emotion spelt out earlier.

On the basis of parts one and two, Griffiths goes on to argue that only affect programs should count as natural kinds, as only they can be grouped together on the basis of having homologous traits. Moreover, since the other things referred to as 'emotion' can't be grouped together on the basis of homology, nor have a shared basis in basic emotions which do have such a basis, emotions, i.e. as a whole, don't form a natural kind. That is, while some emotion types, e.g. 'fear', might form natural kinds, the category 'emotion' doesn't. There is, it turns out, no such thing as a typical emotion.[14]

---

[14] Griffiths's final conclusion is stronger: he runs together the idea that emotions don't form a natural kind with the claim that the category of emotion should be eliminated. As several commentators have observed, eliminativism is a far more controversial position, which is not entailed by his overall argument, e.g. see Solomon (1999). I turn to this issue at the end of the section.

How much this should prove a threat to a philosophy of emotion depends on whether it shares the same subject-matter as the psychoevolutionary approach to emotion. It is at this very juncture that the surplus meaning of emotion terms threatens the discourse. Philosophers, as we noted, tend to view the subjective experience of an emotion as being either the essence or an essential component of an emotion. Lessons from Griffiths's extended Darwinian approach to emotion, then, need only take hold if his account speaks to this concern. All philosophical responses to Griffiths, as far as I can tell, assume that it does.[15] Nevertheless, the textual evidence for this is shaky at best.

Part of the problem is that the psychoevolutionary approach to emotion neglects the phenomenology of emotion. The other is that when Griffiths does mention emotional experience within the context of this approach, his position isn't entirely consistent:

i) Emotion feelings and cognitive phenomena such as the directing of attention are obvious candidates to be added to this list [of affect program responses]. (Griffiths 1997: 77)

ii) The output of perceptual systems is a mental event, a perception, whereas that of the emotion system is behavioral. (Griffiths 1997: 94)

iii) The affect program theory can (and should) incorporate emotional feelings either as an additional element of the program or as the perception of the physiological elements of the program. (Griffiths 1997: 121)

Griffiths (1997), then, seems to be wavering between three distinction positions on how emotional feelings stand with respect to the coordinated set of changes that constitute the

[15] See Deonna and Teroni (2012: §2) for a sample.

affect program responses. Roughly, these feelings are (i) relevant, (ii) irrelevant or (iii) essential features of our affect program responses. At face value, the lesson is that Griffiths doesn't hold a consistent view on emotional feelings *vis-a-vis* affect programs. A more charitable reading has it that emotional feelings aren't the primary explanatory targets of the psychoevolutionary approach to emotion, which seeks to draw on neuroscience and evolutionary biology to offer a psychological science of the emotions. Insofar as the sciences are in the business of studying third-person accessible physiological and behavioural responses, as opposed to first-person accessible phenomenally conscious experiences, the neglect is understandable. Nevertheless, it does little address the present concern, i.e. whether Griffiths's position should inform a philosophy of emotion.

The reason to suppose that it does stems from the assumption made by Griffiths that these physiological and behavioural responses coincide with the phenomenology of emotion. This is evident from the third position, but also from the way he draws on Ekman to develop his account of affect programs:

> The psychoevolutionary theory describes a set of well-defined physiological responses that can be used to create a new classification of emotional responses. Ekman and his co-workers call the responses they have isolated fear, anger, joy, surprise, disgust and sadness. *They use these traditional labels because the new categories coincide more or less well with the occurrent, phenomenologically salient instances of these traditional categories.* I shall refer to these six categories as the affect-program responses. (Griffiths 1990: 189, my italics)

Both Ekman and Griffiths assume that the physiological responses that are a part of the affect programs coincide with emotional feelings. On this assumption, empirical

discoveries about affect programs do have implications for a philosophy of emotion. However, whether the assumption itself is justified is a different matter.

One important lesson from LeDoux's recent work is that this assumption is now found to be incorrect. The threat circuitry, which concerns the amygdala and is responsible for the affect program responses, comes apart from the subjective experiences of emotion. Though the neural circuitry that trigger these responses can affect our subjective experiences and alter them in various ways, the conscious experience of emotion is grounded in a different neural circuitry. The upshot is that discoveries about affect programs don't generalise to emotional feelings.

The point can also be made in another way. It is plausible that both Ekman and Griffiths, in their discussion of affect programs, take 'fear' to refer to a property cluster, which includes the conscious experience of fear, as well as our defensive responses generated by the survival circuits. Nevertheless, as mentioned earlier, Griffiths also assumes 'fear', like other kind categories, can only be categorised using such property clusters if the properties in question co-occur due to an underlying causal mechanism. (Recall, the purpose of categorisation is to allow extrapolation from observed to unobserved instances, which is justified only if the properties co-occur due to such shared mechanisms). It is precisely this assumption which is now shown to be false. While phenomenological and physiological properties associated with fear do typically co-occur, they result from distinct causal mechanisms, which in turn means that we can no longer group things together as 'fear' on the basis of such property clusters. As a result, we can no longer justifiably suppose that discoveries about the affect programs that control our defensive responses hold for the conscious experience of fear as well.

On either way of making the point, Griffiths's discussion of affect programs is subject to a surplus meaning. In using emotion terms to describe these responses, e.g. 'basic emotions', the affect program responses are read as having psychological properties they don't have, *viz.* the conscious experience of emotion.[16] Philosophers have engaged with Griffiths on the basis of this surplus meaning. However, if the neuro-cognitive data on offer by LeDoux is on the right track, this surplus meaning is also seen to be unfounded.

There are at least two lessons to draw from this. First, and perhaps most pressingly, what Griffiths takes be the contender for a natural kind, and worthy of a science of emotion, i.e. affect programs, is something that is only indirectly related to emotional experience. So philosophers who take such experience to be essential to emotion shouldn't look to the psychoevolutionary approach to emotion to underpin their metaphysics. This lesson should be both surprising and alarming, as it tells us that more than twenty years of philosophical discussion generated by Griffiths's influential work, *What Emotions Really Are*, is off the mark.

This is not to say that philosophers have nothing to learn from the psychoevolutionary approach. Here is the second lesson. Griffiths's argument for the category 'emotion' not forming a natural kind centres on there being distinct kinds of emotion types; some that involve affect programs and others which don't. He argues that only some emotion types, i.e. those constituted by affect program responses, form natural kinds. However, we can now make the case against any given type of emotion being a natural kind — including basic emotion types. Take fear. It is plausible that the vernacular

[16] Scarantino's (2018) defence of basic emotions *qua* affect programs takes the lessons from LeDoux's recent work into account by forgoing the need for such emotions to have phenomenal properties.

category of fear picks out various affect program responses, as well the conscious feelings of being afraid. Since the mechanisms that underly these two kinds of phenomena come apart, there is no guarantee that all things picked out by the term 'fear' will share a mechanism which offers justification for extrapolation. The upshot is that though affect programs concerning fear still plausibly form a natural kind, we can no longer unequivocally say that fear itself is a natural kind.

Perhaps this is more grist to Griffiths's mill. Despite arguing that basic emotions are natural kinds, Griffiths (1997) claims that a future psychological science should ultimately get rid of the category 'emotion'. (The category creates confusion because it refers to things which are natural kinds, as well as things which aren't). If no emotions form a natural kind, not even basic emotions, then such a position is indeed warranted. We may have a science of affect programs, but we can no longer have a science of emotion. Nevertheless, it is also hard not to see this as a reductio of Griffiths's whole approach. Though Griffiths thinks that a future science can make do without the vernacular category of emotion, he does think that there is something picked out by this category, which forms a natural kind and is worthy of a science of emotion, *viz.* basic emotions. It is for this reason that the psychoevolutionary approach is an approach to emotion. However, if even basic emotion types don't form a natural kind, the link between the approach and what it claims to investigate is severed. We can have a science of affect programs, which we do already, but this can no longer provide an empirical basis for a psychology of emotion. If we want a scientifically robust approach to emotion itself, this must be found elsewhere.

**4. The Pro-Emotion Debate**

According to Jones (2006), a new consensus is emerging from both philosophy and psychology, which replaces the old dogma that emotions (only) hinder reasoning for the thesis that they, as a matter of fact, can aid practical reasoning. On this rival view, emotions are said to be "clever design solutions to the problem of making fast decision in response to practical problems posed by the natural and social worlds: we perceive a danger and fear immediately primes us to take protective action" (Jones 2006: 3). Precisely how emotions aid decision-making is hypothesised by both de Sousa (1987) and Damasio (1994), which has to do with how emotions modulate attention. On both accounts, emotions help us make decisions promptly by highlighting some options as relevant, whilst eliminating others from consideration. Emotions, on these accounts, aren't the sole drivers of what to choose, but rather narrow the scope of what is considered by our rational faculties. In Damasio's words, emotions act as "biasing devices" which bias us in favour of some options at the expense of others (pg. 174).

The claim that emotions aid decision-making, especially in the specific ways spelt out by de Sousa and Damasio, arguably, doesn't hold up to empirical scrutiny.[17] Nevertheless, there appears to be something right in the idea that emotions meet certain evolutionary ends by modulating attention. In their survey article, Faucher and Tappolet (2002) review the empirical data, from both psychology and neuroscience, in support of this idea.

[17] One major criticism is that the original hypotheses are vague, and none of their disambiguations are plausible in light of either the cited or new empirical data, e.g. see Evans (2002), Gerrans (2007), Colombetti (2008), Bartol and Linquist (2015), and Ransom (2016).

Though they take their survey to lend support for de Sousa's hypothesis, they acknowledge that the data considered underdetermines his thesis in certain respects:

> de Sousa claims that the emotions' function is to direct our attention. More precisely, he suggests that emotions determine a) what information is processed, b) what inferences are drawn, and c) what options are considered in deliberation. The data we have considered concern only the first part of this tripartite thesis. (Faucher and Tappolet 2002: 127)

It is my view that Faucher and Tappolet discuss an even more restricted thesis than is let on by this qualification. While de Sousa, as well as Damasio, hypothesise that emotions aid decision-making by helping us choose which course of action to take, the hypothesis considered in the survey holds that emotions, in particular fear and anxiety, direct our attention to certain visual stimuli. For several reasons, this is a much more modest hypothesis than that originally ascribed to pro-emotion views. First, nowhere is there a claim about emotions actually aiding practical rationality. Rather, the hypothesis explored, i.e. that emotions modulate attention, if true, points in favour of a possible route by which emotions might aid reasoning. Second, it is not that emotions modulate attention wholesale, but rather particular kinds of negative emotions do, namely fear and anxiety. Third, the effects of emotion on attention are restricted to the domain of visual stimuli. These considerations suggest a much more restricted hypothesis than those typically on offer by pro-emotion camps. Nevertheless, this more restricted thesis, though it doesn't quite establish a pro-emotion thesis, let alone a consensus, can be cited in support of it,

and crucially, isn't undermined by sceptical challenges to the original pro-emotion positions.

My cause for contention here isn't with the empirical data, nor in the way it is used to support the more restricted thesis. Rather, it concerns how we are to interpret the restricted thesis itself, especially in relation to de Sousa's pro-emotion thesis. The worry is that this is precisely an instance where the surplus meaning of 'fear' threatens to muddy the waters of what can be concluded on the basis of empirical data. In brief, the restricted thesis holds that fear and anxiety modulate our attention of visual stimuli. The neuro-cognitive data cited in favour of this thesis concern the effects of threat circuitry on attentional phenomena, which suggests that we should interpret 'fear' and 'anxiety' in the thesis as the physiological responses associated with these emotion labels. However, the pro-emotion thesis owing to de Sousa seems to be one about how emotional experiences aid practical reasoning. The only way we can read the neuro-cognitive data as supporting de Sousa's thesis, then, is if we interpret the restricted reading as having a surplus meaning: 'fear' and 'anxiety', which modulate attention, refer to the conscious experience of fear and anxiety. The upshot is that there is a faulty inference from the neuro-cognitive data to de Sousa's hypothesis, which appears to be a product of the surplus meaning of emotion-terms used to spell out the restricted thesis.

Let us look at this charge more closely. To start with, de Sousa, I take it, assumes that emotions are conscious phenomena. This is not made explicit, but several points carry in favour of this interpretation. First, this is the way he is usually interpreted in the philosophical literature; plausibly owing to the deep-seated nature of the assumption that emotions are, at least typically, conscious phenomena. (Given the prevalence of this

assumption, at the very least, we would expect de Sousa to provide us with a caveat if he means something else). Second, this is something that appears to be born out of de Sousa's metaphysics. de Sousa is typically read as endorsing a perceptual theory of emotions, which holds that emotions are, or somehow are akin to, perceptions, e.g. to fear Fido is to perceive danger. Such perceptual theories of emotion tend to be derived on the basis of an analogy between emotion and perpetual experience. For instance, in emphasising the importance of this analogy, Tappolet notes that "emotions and perceptual experiences share many important features, such as phenomenal qualities." (2016: xii). Since perceptual theories are derived in this way, it stands to reason that plausible versions have emotional experiences as their explanatory target. Third, de Sousa argues for his thesis solely on the basis of armchair reasoning, which renders it plausible that he is, at least partially, relying on phenomenological evidence when he describes emotions as determining the "salience" of response-options. Such evidence can only have emotional experiences as the subjects of their explanans.

Now consider the neuro-cognitive data Faucher and Tappolet think lends support for de Sousa's hypothesis. This includes several findings on emotional phenomena, attentional phenomena, and the connection between the two. The problem manifests itself regarding the first of these findings. Faucher and Tappolet begin their survey on emotions by citing LeDoux's findings which he now describes as being mistaken: "the amygdala has been tagged as the hub of the emotion of fear" (LeDouX 1996: 168). This assumption that the amygdala underlies fear is central to their exploration of the connection between emotion and attention. The implicit reasoning, which results from this assumption, and informs this part of their survey, goes something like this: 'LeDoux etc. show that

amygdala is responsible for fear. Other findings show that amygdala modulates attention. Inference: fear modulates attention'. For example, they cite LeDoux in support of the claim that "Efferents projections of the amygdala to the cortical structures explain how fear can influence cognitive processes" (pg. 125). Such examples scatter their discussion. In fact, Faucher and Tappolet make six intermediary conclusions on the basis of their survey of the neuro-cognitive data. These conclusions establish various links between the amygdala and attention, but the ensuing discussion is carried out as one where 'fear' directs our attention.

The inferences drawn here are not problematic in and of themselves. The conclusions about fear and its effects on attention are unproblematic so long as we remain clear that by 'fear' here, we mean either the threat circuitry or the physiological responses to which they give rise. What's problematic is when we read 'fear' in this context as also involving the conscious feeling of fear. This is what makes the inference fallacious.

The surplus meaning of fear threatens Faucher and Tappolet's survey at two levels. First, it threatens the discussion of the restricted thesis. In discussing the actual empirical findings themselves, Faucher and Tappolet are often careful to make qualifications that pre-empt the problematic readings. For instance, they note that "we will review some recent evidence showing that cerebral structures, *traditionally* thought to be part of the emotional network, might contribute to attentional function" (p. 123, my italics). Likewise, elsewhere they make connections between emotional *responses* and attention whilst staying clear of talking about emotions *simpliciter*.

However, confusion still results from how this discussion of the neuro-cognitive data is to be viewed in the context of their overall survey. Before considering the neuro-

cognitive data, Faucher and Tappolet evaluate psychological data in favour of the restricted thesis. Here they consider emotional stroop tasks, probe detection tasks etc, where it is explicit that their target is not emotional responses, but the conscious experience of emotion itself:

> [T]he data we considered suggest strongly the emotion of fear, *as experienced* by normal subjects, involves attentional bias toward three stimuli of the same kind observed in pathological cases. (Faucher and Tappolet 2002: 120, my italics)

The psychological data and the neuro-cognitive data, then, seem to be about two distinct phenomena, and nowhere is this flagged in their discussion. Consequently, the ensuing discussion, which proceeds on the basis of both kinds of data, is confusing for we are left without knowing what the "emotional phenomena" are that is supposed to be related to attention.

At the second level, the surplus meaning of fear threatens the discussion of how the restricted thesis supports de Sousa's pro-emotion hypothesis. As noted, de Sousa takes emotions to be conscious phenomena, which means we can only extrapolate from the restricted thesis to a discussion of this more general hypothesis by deploying the problematic interpretation of the restricted thesis. This results in not just the threat of a surplus meaning, but of the fallacious inference itself: 'the amygdala underlies our threat circuitry, and also modulates attention, therefore, fear modules attention.' It is hard to see how something like this isn't going on when Faucher and Tappolet draw their survey to a

close: "The conclusion is that de Sousa's hypothesis is on the right track with respect to fear and anxiety" (pg. 135).

The aim of this section isn't to undermine pro-emotion views[18], or even to offer a critique of de Sousa's view in particular.[19] The point was to see how the surplus meaning of 'fear' might influence discussions of certain pro-emotion views. That said, one obvious lesson to draw is that, *contra* Faucher and Tappolet, the neuro-cognitive data concerning our defensive survival circuitry doesn't support de Sousa's hypothesis. Another is that to the extent to which such data can be utilised to support pro-emotion views, they can only be employed to support views that hold that threat circuitry modulates attention, and thereby possibly plays a role in resolving various practical problems. We find the underpinnings of such a view in LeDoux himself. Very roughly, the threat circuitry raises our vigilance, i.e. we become more attuned to our sensory environment, focussing on present dangers and being alert for future ones, whilst lowering our threshold for additional defensive responses (2016: 44-5).[20] What we can't do, however, is establish a close link between such neuro-cognitive data and the pro-emotion theses espoused in the

---

[18] Nothing I have said here undermines behavioural data from psychology, which explains some other ways "felt" emotions can both positively and negatively affect decision-making. For example, the Appraisal Tendency Framework successfully links appraisal processes specific to certain emotions with different judgement and choice outcomes. See Lerner *et. al.* (2015) for an overview.

[19] E.g. it may be the case that the survival circuits modulate emotional experiences, which in turn modulate attention. But here the key part of the pro-emotion hypothesis — that emotional *experiences* modulate attention — is supported by the behavioural, as opposed to neuro-cognitive, data.

[20] Also see Prinz (204: 202) for a pro-emotion view which explicitly rejects the idea that 'emotions' need to be felt to modulate attention in ways that foster survival.

philosophical literature; views which assume that emotions *qua* conscious phenomena are what modulate attention and thereby aid practical reasoning. This is not to say that such views cannot have a basis in neuroscience. One positive lesson is that if we are to insist on defending such views, we should shift our focus from the survival circuitry, which concern the amygdala, to those that actually underly the subjective feelings of emotion.

**5. Neurobiological Theories of Emotional Consciousness**

The case made by LeDoux for a separation between the survival circuitry which processes and responds to threat and that which constitute the conscious experience of fear is tied to his assumptions about the latter. For this reason, there is a worry that the lessons which we can draw on the basis of this separation are conditional on LeDoux's own account of emotional consciousness being true. This worry is made especially acute owing to the *prima facie* viability of rival neurobiological accounts of emotional consciousness. If any of these rival accounts prove correct, this threatens to undermine the philosophical implications of LeDoux's recent work. In this final section, I address this worry.

LeDoux's positive account of emotional consciousness is fleshed out in LeDoux (2016), and developed further in LeDoux and Brown (2017) and LeDoux (2019). As is made clear by LeDoux and Brown, while it is traditional to view cognitive states of consciousness as arising from a different brain system from that which generates conscious feelings, on their account both phenomena are grounded in the very same system. For example, on some key traditional views, e.g. Panksepp's (1998, 2016), the conscious

experience of fear is constituted by subcortical regions of the brain. By contrast, on their view, while these regions can modulate emotional feelings by providing nonconscious inputs that alter the kinds of feelings we have, the actual conscious feelings themselves are constituted by a different system: the cortical regions of the brain that also underly the conscious experience of external stimuli. This they dub a neuro-cognitive account of emotional consciousness, for "a general network of cognitions underlies both cognitive and emotional states of consciousness" (pg. 69). Moreover, they draw on this to inform their preferred account, i.e their higher-order theory of emotional consciousness.[21]

According to higher-order theories of consciousness more broadly, first-order representations, say a perceptual representation of an external stimulus, don't suffice for consciousness. In addition to having a first-order representation, one must also be aware of it to have a conscious experience of the external stimulus.[22] (Note: this not to say that one needs to be aware *that* one is aware of the first-order representation). This state of being aware of the first-order representational state is the higher-order state, which is crucial for consciousness. LeDoux and Brown offer a high-order theory not of consciousness in general, but of emotional consciousness.[23] Moreover, they draw on the aforementioned assumption about brain systems to explain how such a higher-order theory might be implemented at the neurobiological level.

On their account, first-order states e.g. perceptual representations of external stimuli, are associated with the prefrontal cortex, whereas higher-order states concern the

---

[21] See the Global Workspace Theory, e.g. Baars (2005), Dehaene (2014), and Naccache (2018), for a different neuro-cognitive account.

[22] E.g. see Rosenthal (2005), Carruthers (2005), and Lau and Rosenthal (2011).

[23] A general theory is provided in LeDoux (2019).

representation of these first-order states by 'cortically based general networks of cognition' (GNC). More specially, "subcortical circuits are not responsible for feelings, but instead provide lower-order, nonconscious inputs that coalesce with other kinds of neural signals in the cognitive assembly of conscious emotional experiences by cortical circuits" (pg. 1). The same is assumed to be true on the higher-order theory of emotional consciousness. What differentiates emotional experiences from other kinds of experiences here are the inputs that are processed by the cortical areas; not the processing systems themselves.[24]

This neuro-cognitive approach to emotional consciousness has at least two plausible rivals: what LeDoux and Hofmann (2018) call the "neuro-Darwinian" and the "neuro-Jamesian" approaches.[25] For those who subscribe to the former, e.g. Panksepp (1998, 2016), emotional feelings arise from amygdala-centred subcortical circuits, which we have inherited from our animal ancestors and are highly preserved amongst mammals. By contrast, advocates of the latter, e.g. Damasio (1994) and Craig (2009), assume that such feelings result from activity in the body-sensing circuits in the neocortex (somato-sensory and/or insula areas), e.g. those triggered when threat activates the amygdala circuits. In more recent work, Damasio has shifted his emphasis from cortical to subcortical body-sensing areas in the brainstem.[26] The key difference between these approaches and LeDoux's is that they take the traditional line, *viz.* that emotional consciousness is a product of neural circuits distinct from those that underly other kinds of conscious experience. Which of these accounts is correct is an empirical matter. For LeDoux, the fact

---

[24] Also see Brown, Lau and LeDoux (2019) and LeDoux (2019)

[25] They also mention a neuro-behaviourist position, which I shall leave aside for it forgoes realism about subjecting feelings for an eliminativist position.

[26] E.g. see Damasio *et al.* (2013) and Damasio and Carvalho (2013).

the the neuro-cognitive account does a better job of explaining various puzzles in the field, provides abductive grounds to prefer it over these rival accounts.[27] This is often supplemented by other considerations, e.g. against the neuro-Jamesian approach, LeDoux and Hofmann note that while the body-sensing subcortical and cortical "circuits clearly represent body states, convincing evidence that these representations are the main causes of emotional experiences is lacking" (2018: 68).[28]

Let us now turn to addressing the present worry, i.e. that the lessons we explored in the previous sections are conditional on LeDoux's own theory of emotional consciousness being true. We now see that there are two ways of disambiguating this worry: the lessons are conditional on the neuro-cognitive approach, or they are conditional on the higher-order theory of emotional consciousness, where the neuro-cognitive approach gives us an account of how the theory is implemented at the neurobiological level. However, we don't need to assume either to draw out our lessons. What we require is not a positive thesis about the sub-regions of the brain that give rise to conscious feelings, but rather a negative thesis about which regions don't. In particular, what we require is a rejection of the traditional line: that conscious feelings are constituted by subcortical circuits which include the amygdala.

Fear is a case in point. The survival circuits that involve the amygdala, and trigger our defensive affective program responses, aren't what constitute the conscious experience of fear. Subsequently, while the affect programs may indirectly shape the kinds of fearful feelings we have, empirical findings concerning this circuitry can't directly inform a

---

[27] See page 9.

[28] See LeDoux (2016: 133-138, 2019: §53, §64-65) for an extended discussion.

philosophy of emotion — where we understand emotion to be a conscious phenomenon. Likewise, defensive survival circuits, including the activation of the amygdala, plausibly play a role in modulating attention of threatening stimuli. Nevertheless, for the same reason as before, we can't take this as a data point to bolster pro-emotion views which hold that emotion *qua* conscious feelings play a role aiding practical reason by drawing our attention to salient features of the environment. This is to say nothing of what in fact constitutes the conscious experience of emotion. Rather, the lessons from this paper stand or fall depending on how much we should look to the traditional neurobiological conception of emotional consciousness to inform a philosophy of emotion. LeDoux's body of work, both old and new, strongly suggests that we should look elsewhere.[29]

[29] Many thanks to Alex Grzankowski, as well as two referees for this journal, for their comments on the draft. I am also grateful to audience members at the Philosophy, Psychology and Neuroscience Seminar at the University of Glasgow and the New Directions in Emotion Workshop at the University of Manchester for their helpful comments.

## References


- Anderson, A. K and Phelps, E. A (2002). 'Is the human amygdala critical for the subjective experience of emotion? Evidence of intact dispositional affect in patients with amygdala lesions'. *Journal of Cognitive Neuroscience* 14(5):709–720.
- Baars, B. J (2005). 'Global workspace theory of consciousness: Toward a cognitive neuroscience of human experience'. *Progress in Brain Research* 150:45–53.
- Barret, L. F (2009). 'Variety is the spice of life: A psychological construction approach to understanding variability in emotion'. *Cognition and Emotion* 23:1284-1306.
- Bartol, J. and Linquist, S. (2015). 'How do Somatic Markers Feature in Decision Making?' *Emotion Review* 7(1):81-89.
- 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* 85:116-124.
- Boyd, R. (1991). 'Realism, Anti-foundationalism and the Enthusiasm for Natural Kinds'. *Philosophical Studies* 61:127–148.
- Brigandt, I. (2003). 'Species Pluralism Does Not Imply Species Eliminitivism'. *Philosophy of Science* 70:1305–1316.
- Brown, L. , Lau, H., and LeDoux, J. E. (2019). 'The Misunderstood Higher-Order Approach to Consciousness'. *Trends in Cognitive Science* 23(9): 754-768.
- Carruthers. P. (2005). 'Consciousness: Essays from a High-Order Perspective'. Oxford: Oxford University Press.

- Chalmers, D. J. (1995). 'Facing Up to the Problem of Consciousness'. *Journal of Consciousness Studies* 2(3):200–219.
- Clore, G. (1994). 'Why Emotions Are Never Unconscious'. In *The Nature of Emotion: Fundamental Questions*, eds. P. Ekman and R.J Davidson. New York: Oxford University Press: 285-90.
- Colombetti, G. (2008). 'The Somatic Marker Hypotheses, and What the Iowa Gambling Task Does and Does not Show'. *British Journal for the Philosophy of Science* 59: 51-71.
- Craig, A. D. (2009). 'How do you feel — now? The anterior insula and human awareness'. *Nature Reviews Neuroscience* 10:59-70.
- Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam, and London: Vintage.
- Damasio, A. R., Damasio, H., and Tranel, D. (2013). 'Persistence of Feelings and Sentience After Bilateral Damage of the Insula'. *Cerebral Cortex* 4:833-846.
- Damasio, A. R. and Carvalho, G. B. (2013). 'The Nature of Feelings: Evolutionary and Neurobiological Origins'. *Nature Reviews Neuroscience* 14:143-152.
- Darwin, C. (1859/1964). *On The Origin of Species: A Facsimile of the First Edition*. Cambridge: Harvard University Press.
- Darwin, C. (1872/1965). *The Expression of the Emotions in Man and Animals*. Reprint, Chicago: University of Chicago Press.
- Dehaene, S. (2014). *Consciousness and the Brain: Deciphering how the brain codes our thoughts.* New York: Viking Press.
- Deonna, J. A. and Teroni, F. (2012). *The Emotions: A Philosophical Introduction*. London and New York: Routledge.

- de Sousa, R. B. (1987). *The Rationality of Emotion*. Cambridge and London: MIT Press.
- Dunn, B. D., Dalgleish, T. and Lawrence, A. D. (2006). 'The Somatic Marker Hypothesis: A Critical Evaluation'. *Neuroscience and Biobehavioral Reviews* 30: 239–71.
- Ekman, P. (1972). 'Universals and Cultural Differences in Facial Expressions of Emotions'. In *Nebraska Symposium on Motivation*, ed. J. Cole. Lincoln, Nebraska: University of Nebraska Press: 207-283.
- Ekman, P. (1973). *Darwin and Facial Expression*. New York: Academic Press.
- Ekman, P. (1980). 'Biological and cultural contributions to body and facial movement in the expression of emotions'. In *Explaining Emotions*, ed. A. O. Rorty. Berkeley: University of California Press.
- Ekman, P. (1999). 'Basic Emotions'. In *Handbook of Cognition and Emotion*, eds. T. Dalgleish and M. Power. Chichester: John Wiley and Sons, Co.:45–60.
- Evans, D. (2002). 'The Search Hypothesis of Emotions'. *British Journal for the Philosophy of Science* 53(4):497-509.
- Fanselow, M. S. and Pennington, Z. T. (2018). 'A return to the psychiatric dark ages with a two-system framework for fear'. *Behaviour Research and Therapy* 100:24-29.
- Faucher, L. & Tappolet, C. (2002). 'Fear and the focus of attention'. *Consciousness and Emotion* 3(2):105-144.
- Feinstein J. S, *et al.* (2013) Fear and panic in humans with bilateral amygdala damage. Nature Neuroscience 16(3):270–272.
- Freud, S. (1915). 'The Unconscious'. In *The Standard Edition of the Complete Pyschological Works of Sigmund Freud*, vol 14, ed. J. Strachey. London: The Hogarth Press:161-215.

- Frijda, N. H. (1999). 'Emotions and Hedonic Experience'. In *Well-being: The foundations of hedonic psychology*, eds. D. Kahneman, E. Diener, & N. Schwarz. New YorkRussell Sage Foundation:190-210.
- Gerrans, P. (2007). 'Mental Time Travel, Somatic Markers and "Myopia for the Future" '. *Synthese* 159:459–74.
- 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 *Thinking About Feeling: Contemporary Philosophers on Emotions*, ed. R. C. Solomon. Oxford and New York: Oxford University Press: 389-417
- Griffiths, P. E. (2004a). 'Toward a "machiavellian" theory of emotional appraisal'. In *Emotion, Evolution, and Rationality*, eds. D. Evans & Pierre Cruse. Oxford and New York: Oxford University Press: 89-106
- Griffiths, P. E. (2004b). 'Emotions as natural and normative kinds'. *Philosophy of Science* 71(5):901-911.
- Hacking, I. (2007). 'Natural kinds: Rosy Dawn, Scholastic Twilight'. *Royal Institute of Philosophy Supplement* 61:203-239.
- Jackson, F. (1982). 'Epiphenomenal qualia'. *Philosophical Quarterly* 32 (April):127-136.

- Jones, K. (2006). 'Quick and Smart? Modularity and the pro-emotion consensus'. *Canadian Journal of Philosophy* 32(5S):3-27.
- Lau, H. and Rosenthal, D. M. (2011). 'The higher-order view does not require consciously self-directed introspection: Response to Malach'. *Trends in Cognitive Science* 15(11):508–509.
- LeDoux, J. E. (1996). *The Emotional Brain.* New York: 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 Under- stand and Treat Fear and Anxiety.* New York: 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 Brown, R. (2017). 'A higher-order theory of emotional consciousness'. *Proceedings of the National Academy of Sciences* USA, 114: E2016-E2025.
- LeDoux, J. E., and Hoffman, S. G (2018). 'The subjective experience of emotion: a fearful view'. *Current Opinion in Behavioral Sciences* 19:67–72.
- Lerner, J. S., Li, Y., Valdesolo, P., and Kassam, K. S (2015). 'Emotion and Decision Making'. *Annual Review of Psychology* 66:799-823.
- Linquist, S. & Bartol, J. (2013). 'Two Myths about Somatic Markers'. *British Journal for the Philosophy of Science* 64(3):455-484.

- Naccache, L. (2018). 'Why and how access consciousness can account for phenomenal consciousness'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 373 (1755).
- Nagel, T. (1974). 'What is it like to be a bat?' *Philosophical Review* 83 (October):435-50.
- Ohman, A. (2002). 'Automaticity and the amgydala: nonconscious responses to emotional faces'. *Current Directions in Psychological Science* 11:62-66.
- 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*. New York: Oxford University Press.
- Panksepp, J. (2016). 'The cross-mammalian neurophenomenology of primal emotional affects: from animal feelings to human therapeutics'. *Journal of Comparative Neurology* 524:1624-1635.
- Pine, D.S. and LeDoux, J.E. (2017). 'Elevating the Role of Subjective Experience in the Clinic: Response to Fanselow and Pennington'. *American Journal of Psychiatry* 174(11): 1121- 1122.
- Prinz, J. (2004). *Gut Reactions: A Perceptual Theory of Emotion*. Oxford: Oxford University Press.
- Ransom, M. (2016). 'Why Emotions Do Not Solve the Frame Problem'. In *Fundamental Issues of Artificial Intelligence*, ed. V. C. Müller. Springer: 353-365.
- Rosenthal D. M (2005). *Consciousness and Mind*. Oxford: Oxford University Press.
- Russell J. A. (2003): 'Core affect and the psychological construction of emotion'. *Psychollogical Review* 110:145-172.

- Scarantino, A. (2018). 'Are LeDoux's Survival Circuits Basic Emotions Under a Different Name?' *Current Opinion in Behavioral Sciences* 24: 75-82.
- Solomon, R. C. (1999). 'Review of What Emotions Really are: The Problem of Psychological Categories'. *Philosophical Review* 108 (1):131.
- Tamietto, M. and de Gelder, B. (2010). 'Neural bases of the non-conscious perception of emotional signals'. *Nat Rev Neurosci* 11(10):697-709.
- Tappolet, C. (2016). *Emotions, Value, and Agency*. Oxford: Oxford University Press.
- Tomkins, S. S. (1962). *Affect, Imagery, Consciousness*. New York: Springer.