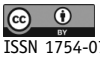
ORIGINAL MANUSCRIPT

emotion review

Emotion Review
Vol. 0, No. 0 (June 2024) 1–12


ISSN 1754-0739
DOI: 10.1177/17540739241259566
https://journals.sagepub.com/home/emr

On How to Develop Emotion Taxonomies

Raamy Majeed
Department of Philosophy, School of Social Sciences, University of Manchester, Manchester, UK

Abstract

How should we go about developing emotion taxonomies suitable for a science of emotion? Scientific categories are supposed to be "projectable": They must support generalizations required for the scientific practices of induction and explanation. Attempts to provide projectable emotion categories typically classify emotions in terms of a limited set of modules, but such taxonomies have had limited uptake because they arguably misrepresent the diversity of our emotional repertoire. However, more inclusive, non-modular, taxonomies also prove problematic, for they struggle to meet the projectability constraint. In this paper, I explain how a developmental approach to emotion, one that utilizes the notion of progressive modularization, can help us approach emotion categorization in a more inclusive and projectable manner.

Keywords
emotion, modularity, categorization, basic emotions, emotion development

Introduction

What sorts of emotion categories should inform a science of emotion remains a source of controversy. For instance, some argue that scientific research into emotion is best served by positing a list of basic emotions, for example, fear, anger, surprise, happiness, sadness, and disgust (Ekman, 1973; Frijda, 1986; Izard, 1977; Panksepp, 1998), whereas others argue that we should also include complex emotions, such as jealousy and guilt (Cowen et al., 2019). At the more liberal end, Plutchik's (2003) emotion wheel posits numerous unique emotion categories, including loathing, admiration, and acceptance. However, distinct from the question of what emotion categories we should employ in scientific research stands a more fundamental question: How should we go about finding out which emotion categories are scientifically suitable in the first place? Put another way, how should we go about developing emotion taxonomies for a science of emotion?

At present, attempts to categorize emotion give way to something resembling a dilemma. On the one hand, we have restrictive emotion taxonomies that are scientifically suitable but too reductive. On the other hand, we have liberal emotion taxonomies which are more inclusive, but scientifically unsuitable. For example, restrictive approaches often classify emotions in terms of a limited set of modules: systems, mechanisms, or programs hardwired into our brains by evolution and purpose-built to generate certain coordinated patterns of expressive, physiological, behavioral, and (perhaps) phenomenological responses (Charland, 1995; Cosmides & Tooby, 2000; Griffiths, 1997). Such approaches yield categories suitable for scientific practice, but they have also had limited uptake because they arguably misrepresent the diversity of our emotional repertoire (de Sousa, 2008; Nussbaum, 2001; Solomon, 2007). By contrast, more liberal approaches categorize emotions via their functional roles (Adolphs & Andler, 2018; Loaiza, 2024), their ecological roles (Cosmides & Tooby, 2000), or their evaluative content (Arnold, 1960; Deonna & Teroni, 2012; Lazarus, 1991). Others go even further, forgoing discrete emotion categories in favor of multidimensional emotion scales (de Sousa, 2008; Scherer, 1993). While such approaches deliver more inclusive emotion categories, they are also thought to be ill-suited for scientific practice (Griffiths, 1997).

In this paper, my aim is to provide an exposition of these issues and explain how taking a developmental approach to emotion categorization can help address them. Central to this approach is Karmiloff-Smith's notion of modularization

Corresponding author: Raamy Majeed, Department of Philosophy, School of Social Sciences, University of Manchester, Oxford Road, Manchester, M13 9PL, UK.
Email: Raamy.majeed@manchester.ac.uk

(Karmiff-Smith 1992; Karmiloff-Smith et al., 2018; Majeed, 2022a, 2022b; Mithen, 1996; Westermann et al., 2007). I argue that appealing to modularization, rather than prespecified modules, can help us approach emotion categorization in a way that better accommodates concerns to do with both inclusivity and scientific suitability.

It is worth being clear that the developmental approach is not intended to guarantee that our final list of emotion categories (whatever they turn out to be) will preserve everything that falls under our folk emotion concepts. However, it is my view that a science of emotion must do more to offer inclusive emotion categories, ideally ones that are not completely removed from our everyday emotion concepts. While the present paper focuses not on what emotion categories there are, but how we should seek them out, the approach taken here is proposed as a modest step in that direction.

The paper is structured as follows. First, I provide some background as to what makes a category scientifically suitable, and then explain how modular approaches provide us with one way to meet this requirement. Next, I explore some problems with such modular approaches, focusing on the charge that they are too reductive. Following this, I provide an exposition of some more liberal non-modular approaches and explain why they are scientifically unsuitable. Finally, I offer a novel, developmental approach to emotion categorization, which I go on to argue is scientifically suitable and more inclusive than the modular approaches.

Scientific Suitability

There are constraints on the sorts of categories that will be useful for scientific practice. On one influential view in the philosophy of science, such categories must be "projectable": They must support generalizations required for the scientific practices of induction and explanation (Goodman, 1954). In particular, we must be able to reliably extrapolate from samples of a category to other members of that category. It is such generalizability that justifies our inductive practices, for example, when we go from observing correlations in some samples of a category to inferring that such correlations can also be found in other instances of the category. Moreover, it is also this sort of generalizability that aids explanation, for example, when we explain correlations in various samples in terms of their membership of the same category.[1]

A common way a category can prove projectable is if its members share some underlying molecular structure; a feature which is sometimes considered to be its essence. For example, *gold* is a projectable category, as all of its instances involve minerals with the atomic number 79. By contrast, *jade* is not a projectable category because there are two different types of minerals, with different molecular structures, associated with jade: nephrite and jadeite. This way of understanding projectability, however, is not appropriate for the life sciences, as biological phenomena (e.g., species) are highly variable in their instances, and are thereby unlikely to have anything akin to micro-structural essences (Boyd, 1989).

Advocates of projectability respond to this issue by appealing to mechanisms. In particular, categories are assumed projectable if they group phenomena based on some similarity-generating mechanism: a mechanism that explains why certain properties cluster together in a projectable manner (Boyd, 1989; Keil, 1989). Following Boyd (1989), such mechanisms are often referred to as "homeostatic" mechanisms. However, it is worth noting that Boyd's (1989) use of the term is potentially misleading as he has a weaker conception in mind than the standard one. As Craver (2009, p. 578) clarifies, "A mechanism is homeostatic in Boyd's [1989] sense if the mechanism explains the regular co-occurrence of phenomenal properties in the cluster." Thus understood, the projectability constraint is met when there is some similarity-generating, or "homeostatic," mechanism that explains why the category is projectable, that is, why we can reliably extrapolate from samples of the category to other instances of that category.

Modularity

What, then, is a suitable homeostatic mechanism for emotion? An obvious suggestion is to go straight to the source and identify emotions with their neural bases. For instance, we might be tempted to categorize emotions with respect to the limbic system, an evolutionarily ancient part of the brain that is said to be responsible for emotion (MacLean, 1952). The suggestion is that we categorize emotion, in general, with respect to the limbic system, and we categorize its sub-categories, such as fear and anger, in terms of distinct circuits located in this system. However, in order to categorize emotions in terms of the limbic system, or any other system for that matter, we first need an account of how to identify the correct mechanisms that would help explain the projectability of our categories.

Griffiths (1997) develops such an account by drawing on evolutionary developmental biology, where homeostatic mechanisms are thought to be patterns of descent. As Griffiths (1997) argues, unlike classifications based on analogy (shared functions), those based on homology (shared ancestry) are supposed to be "deep": Even when the function has been transformed, there is more convergence in the underlying causal mechanisms. Thus understood, the point is not that the limbic system, or whatever we find to be the neural underpinnings of emotion in the future, is irrelevant to emotion categorization. The point, rather, is that it is by looking at descent that we can identify the relevant mechanisms, neural or otherwise, that explain the projectability of our emotion categories.

So where does taking this approach lead us? According to Griffiths (1997), it tells us that emotions are best thought of in terms of what is sometimes called "Darwinian modules": systems, mechanisms, or programs hardwired into our brains by evolution and purpose-built to generate certain coordinated patterns of expressive, physiological, behavioral, and (perhaps) phenomenological responses (de Sousa, 2008; Griffiths, 1997). Note that Griffiths (1997) himself uses Tomkins's (1962) notion of affect programs, which are innate neural circuits responsible for the short-lived stereotypical physiological changes that constitute our emotional responses. For our purposes, we can think of affect programs as a variant of Darwinian modules. The important point is that this way of categorizing emotions assumes that there is a Darwinian module (e.g., affect program) for roughly each of the basic emotions. For example, there is a module for fear, a module for anger, one for surprise, and so on. Moreover, if it turns out that any of our basic emotion categories lack a corresponding module, this is a reason to eliminate that emotion from our list of emotion categories.

Overall, there are two lessons to draw so far. First, scientifically suitable categories must be projectable. Second, one way to ensure such projectability is to adopt modular emotion taxonomies. In what follows, I take the projectability constraint as a given, but question the wisdom of satisfying it by appealing to emotional modularity.

Reductive Concerns

Despite the fact that modular emotion taxonomies have proved highly influential within certain strands of emotion research, they have also become highly controversial. The major source of controversy is over the existence of emotion modules themselves. LeDoux's (1996) findings on how an amygdala-based circuit is implicated in how rodents respond to threat was thought to confirm an affect program for fear. Moreover, it was assumed that future research would uncover specific circuits for other basic emotions as well (Griffiths, 2004). However, despite initial optimism, whether the empirical data supports the existence of any neural circuits that we could identify as emotion-specific modules are presently a point of contention.

For example, Barrett (2013, 2017) argues that recent meta-analyses have failed to uncover any biological markers, either in the body or the brain, which correspond to our emotional concepts. For example, while the amygdala is implicated in fear, it is also implicated in various other emotional and non-emotional responses. By contrast, others argue we simply have not looked hard enough. Scarantino (2015), in particular, argues that while there is a lack of evidence to show that there are hardwired mechanisms in the brain that correspond one-to-one with our folk emotion concepts, this does not preclude the possibility that there are hardwired emotion systems that outstrip such concepts. For instance, there might not be one specific emotion-generating mechanism for our vernacular category fear. But this does not rule out the existence of several distinct fear-generating systems, ones which correspond to distinct fear categories. There might, say, be a distinct system for exteroceptive fear and another for interoceptive panic (Adolphs & Anderson, 2018).

We need not take a stand on this debate here. The point is that modular emotion taxonomies typically prove controversial owing to skepticism over the existence of emotion modules. However, there is also a further, less explored, worry; one that remains even if it could be established that there are such things as emotion-specific modules. That is, modular emotion taxonomies are arguably too reductive: They misrepresent the diversity of our emotional repertoire (de Sousa, 2008; Nussbaum, 2001; Solomon, 2007).

There are a series of related concerns here, most of them raised in the philosophy of emotion. One concern is that to categorize emotions in terms of a limited set of modular-cum-basic emotions is to fail to represent the sheer variety of emotions we can experience. We experience not just basic emotions, such as anger, but a lot of cognitively complex emotions, such as jealousy and schadenfreude, as well. Relatedly, some argue that we experience a far greater number of emotions than those for which we have names (de Sousa, 2008). To categorize emotions simply according to a limited set of basic emotions is to neglect such emotions.

To be fair, proponents of basic emotion taxonomies, for example, Ekman (1973), do not take it as their aim to provide an account of everything that falls under our vernacular emotion category.[2] Likewise, Griffiths (1997) envisions a sort of burden-sharing. That is, cognitive psychologists and cognitive neuroscientists should attempt to investigate basic emotions, while the task of investigating complex emotions is best undertaken by social psychologists, or perhaps even sociologists and anthropologists. The suggestion is that we can have different emotion taxonomies to suit different research ends. There is nothing, in and of itself, wrong with such a "divide and conquer" approach to emotion research. However, we can also recognize the point raised by philosophers, that is, if we take such an approach, we will find modular approaches to emotion (e.g., Ekman's) limiting. It will exclude a lot of the emotions philosophers find interesting, such as jealousy, guilt, and love. So instead of fostering interdisciplinary research, say between cognitive science and philosophy of mind, modular emotion taxonomies drive philosophers to look elsewhere (de Sousa, 2008; Nussbaum, 2001; Solomon, 2007).[3] The point, then, is that insofar as we are moved by fostering such interdisciplinary work, more needs to be done to bridge the divide between the sciences and the humanities. One way to do so is to ensure that we bring our different research tools to bear on the same explanatory targets. A divide and conquer approach, for all its advantages, looks to be in tension with such a bridge-building endeavor.

A second concern provides us with a reason to think the situation is a lot worse than that. There is a growing consensus within cognitive neuroscience that activity in the subcortical regions of the brain, which are taken to be the neural underpinnings of affect programs, while perhaps necessary are unlikely to be sufficient for "subjective feelings," that is, the conscious experience of emotion (Adolphs & Anderson, 2018). For example, LeDoux (2016) himself, whose earlier work was thought to confirm a Darwinian module for fear, has clarified that what he has been investigating all along is how we respond to threat, not fear as such—at least not if we use emotion terms to pick out subjective feelings. As a point of reference, LeDoux (2016) now argues that we should call the circuits responsible for threat "defensive survival circuits" instead of "fear circuits," lest we imply that these circuits underpin the conscious experience of fear.

Some proponents of Darwinian modules, for example, Scarantino (2018), are well aware of this, and yet are happy to use basic emotion labels to refer to survival circuits, that is, circuits responsible for various behavioral and physiological responses associated with emotion. Likewise, Adolphs and Anderson (2018) argue that we should identify emotions with internal brain states responsible for various behavioral markers for emotion (what they call "emotion primitives"), even if these states are not responsible for the conscious experience of emotion. How to use emotion terms within a science of emotion, then, proves controversial. But that is not the main point here. Rather, the point is that if the growing consensus is right, to categorize emotions with respect to things such as Darwinian modules, basic emotions, or affect programs, is to offer emotion categories that might exclude the conscious components of emotion (Majeed, 2020).

It is important to recognize that this too need not, in and of itself, be a problem. Many psychologists and neuroscientists are not focused on the conscious aspects of emotions but rather on their behavioral aspects. However, once again, we can also note that this poses a challenge for fostering interdisciplinary collaboration between such sciences and philosophy, for example. Philosophers of emotion are typically interested in emotions as first-person accessible phenomena: states with subjective feelings, in particular, states with distinctive sorts of phenomenology to other mental states. For this reason, they are bound to find research into (just) the neural underpinnings of certain sorts of expressive, physiological, and behavioral responses at a disconnect from their central concerns (Majeed, 2020). This is not to say that all scientific research into emotion is irrelevant. There is growing research into the cognitive and neural processes that might give rise to conscious emotions (Barrett, 2017; LeDoux, 2016). The worry is that modular emotion taxonomies, which often underpin scientific research into emotion, will not be of much interest to those who are mainly interested in emotions understood as conscious phenomena.

Finally, quite removed from anything to do with emotional experience, there is also a further concern, namely while Darwinian modules might have been adaptive in the Pleistocene where they evolved, it remains unclear how they can be useful for us now (Jones, 2008). To elaborate, explaining how emotions often make us behave in irrational ways is typically considered a desideratum of any emotion theory, however, there is also a growing "pro-emotion" consensus that attempts to explain how emotions can (simultaneously) play a positive role in human reasoning (Damasio, 1994; de Sousa, 1987; Jones, 2008). The worry is that while positing Darwinian modules may help explain the ways emotions can hinder practical rationality, such a framework is incapable of explaining the ways emotions can also aid such rationality.

Proponents of Darwinian modules can mitigate some of these concerns. For example, evolutionary psychologists such as Cosmides and Tooby (2000) argue that emotions evolved to meet certain adaptive challenges can be recalibrated to address challenges that are evolutionarily novel. In a similar vein, following Ekman and Cordaro (2011), Scarantino (2015) argues that affect programs are best conceived of as "open programs" (Mayr, 1974). That is, instead of being programmed to trigger a set of fixed responses, what is more likely is the evolutionary selection of programs capable of producing flexible outputs that can respond efficiently to context-dependent environmental challenges (Prinz, 2004). Such responses take us some way toward addressing the present worry, but some aspects of it still remain.

The remaining problem is that modular-cum-basic emotions are usually thought of as short-term impulses and thereby might not account for *all* the ways emotions help with our long-term goals. For example, as Solomon (2007, p. 19) argues, "anger is much more than a basic emotion or a set of feelings. It is a way of interacting with another person (or with a situation or a task) and a way of situating oneself in the world." This is an old existentialist point, one that is perhaps easy to write off. For instance, it is tempting to think that while a short-lived burst of anger might not explain all the ways your anger helps situate yourself in the world, the disposition to have such bursts might. However, this is too quick. Of course, basic emotions can come in the form of "emotional episodes," that is, particular occurrent emotions, or "emotional dispositions," that is, tendencies to have such occurrent emotions (Goldie, 2000). Solomon's (2007) criticism rests not on whether anger could be understood in a dispositional sense, but on how exactly we should do so. His main charge, I gather, is that something is lost when we treat its dispositional profile simply as a tendency to manifest a certain affect program. Griffiths (2002, p. 239) himself makes a similar point about complex emotions: "When a woman's feeling of guilt explains her behavior through a long session of negotiation with her husband and their lawyers, it does more than dispose her to

intermittently display affect-program sadness and affect-program fear."

Griffiths (2003) aims to tackle the strategic role of emotions via his "Machiavellian" view of emotion. On this view, emotions are Machiavellian "to the extent that they find their dominant evolutionary functions in social competition" (Griffiths, 2003, p. 53). As I read him, Griffiths (2003) thinks both basic and complex emotions are Machiavellian and often work together to shape our long-term plans. However, this does not directly help address Solomon's (2007) critique. Griffiths (2003) still understands basic emotions in terms of affect programs. What is gained by treating them as Machiavellian is the recognition that they "show an evolved sensitivity to strategically significant aspects of the organism's social context" (Griffiths, 2003, p. 62). We need not deny this to also recognize, as Solomon (2007) does, that emotions might be strategic in ways that go beyond such evolved sensitivities. Griffiths (2003), as we have seen, acknowledges such strategic potential when it comes to complex emotions but stops short of acknowledging that the same might be true of some (non-modular) instances of basic emotions. We miss both when we take a modular perspective.

There is plenty more a proponent of emotional modularity can say in response to all three concerns (Majeed, 2022c). However, for now, it suffices to note the sorts of worries that underlie the charge that modular emotion taxonomies are too reductive. To this, I should add that categorization is always a reductive act. To categorize is to offer abstractions that simplify complex phenomena. Categorization of emotion proves no exception. To sort our emotions into categories such as *anger* or *fear* is to abstract away from the rich detail manifest in instances that fall under these categories. This is not the problem. Rather, the problem is that the sorts of emotion categories that result from modular emotion taxonomies are reductive in ways that threaten the legitimacy of these very categories. They are not just reductive, they are *too* reductive. In other words, they leave too much out. How much is too much, of course, will be vague, but as we have seen, they leave out the conscious aspects of emotion, certain types of mental states that typically count as emotion (e.g., complex emotions), and some of the ways emotions can be strategic.

Projectability Concerns

An obvious way to avoid such reductive worries is to endorse a more liberal approach to emotion categorization. The problem with more inclusive emotion categories, however, is that they arguably fail to meet the projectability constraint. Consider the vernacular concept: "emotion is a putative psychological category of motivational states that exhibit passivity" (Griffiths, 1997, p. 246). According to Griffiths (1997), this concept picks out three distinct kinds of phenomena: (modular) basic emotions, (non-modular) complex emotions, and social pretenses. The problem with this sort of heterogeneity is that it seems to conflict with projectability. As Griffiths (1997, p. 242) explains, "Retaining the vernacular concepts is not an option, at least for the purposes of induction and explanation, because there is now no epistemic warrant for supposing that discoveries about some emotions will extend to all other emotions."

To elaborate, to say that emotions are "passive" is to say that they are independent of our long-term planned actions (Griffiths, 1997). The issue is that emotions arguably turn out to be passive in two very different ways. First, emotions can be passive because they stem from affect programs: innate neural circuits responsible for the short-lived stereotypical physiological changes that constitute our emotional responses (Tomkins, 1962). Crucially, affect programs are typically thought of as modular systems, for they bear certain Fodorian hallmarks of modularity (Fodor, 1983). For example, affect programs are supposed to be quick, automatic, and encapsulated from thought.

However, emotions can also be passive in a different sense. This can be brought out using Frank's (1988) theory of emotion, where emotions are seen as irruptive patterns of motivation. Unlike affect programs, which are useful to explain basic emotions, Griffiths (1997) sees Frank's (1988) theory as a way to explain complex emotions, such as jealousy. We need not get bogged down in the details of the theory. The main point is that such emotions irrupt our long-term patterns not because they stem from modular systems. As Griffiths (1997, p. 246) points out, "Whatever psychological mechanism underlies the irruption of these clusters of desires into belief-desire causation, it is not the same mechanism that allows the affect programs to rapidly engage various effector stores without reference to consciously accessible beliefs and desires."

This creates a problem for the projectability of the vernacular category. Since emotions so understood are brought about by distinct processes, or mechanisms, there is no "epistemic warrant" for supposing that the discoveries we make about some instances of the category will generalize to other members of that category. To clarify, the vernacular category was introduced to capture mental states with certain clusters of properties, including passivity. Part of what we want from an investigation into emotion is to help us move beyond our knowledge of just these clusters. For instance, we want to know further properties possessed by such mental states, as well as the sorts of mechanisms responsible for them. If it turns out, however, that the category captures phenomena with two very different mechanisms, we cannot reliably extrapolate from some samples that manifest the relevant clusters (e.g., passivity) to other instances of the category.

This is not just a problem for the vernacular category. Projectability is also a concern for more inclusive conceptions of emotion found in both psychology and philosophy. Consider ecological conceptions of emotion found in

evolutionary psychology (Cosmides & Tooby, 2000). Such views categorize emotions in terms of the sorts of adaptive problems they solve, for example, fear solves the problem of dealing with danger, anger solves the problem of specific kinds of obstacles, and so on. In general, ecological conceptions categorize emotions at the level of task description. Such conceptions of emotions are typically criticized because they rest on adaptationist assumptions which are independently deemed problematic (Buller, 2005; Gould & Lewontin, 1979; Sterelny & Griffiths, 1999). However, when it comes to categorization, the problem once again is that such conceptions track two very different kinds of phenomena: basic emotions triggered by modules and complex emotions brought about by non-modular systems.

The same issue arises for other liberal conceptions of emotion, such as those provided by functionalism in the philosophy of mind. Functionalist accounts analyze mental states in terms of their causes, their effects, and the relationships they bear to one another, that is, instead of their underlying physical makeup (Armstrong, 1968; Braddon-Mitchell & Jackson, 2006; Putnam, 1960).[4] Functional analyses are commonplace in the philosophy of mind, so it is natural for philosophers investigating emotions to analyze them functionally (Loaiza, 2024). Functional analyses of emotions, however, are not just confined to philosophy. While scientific accounts of emotion typically characterize emotions in terms of their neurological profiles (Panksepp, 1998), there has been a recent push toward functional conceptions here too. For instance, according to Adolphs and Andler (2018, p. 195), emotions are to be "individuated by the web of causal relations they have to all other internal states, as well as to stimuli and behaviors … In a nutshell: emotions are individuated by what they do, not by how they are physically constituted."

Functionalist accounts are also, arguably, too general to satisfy the projectability constraint. As before, they capture modular and non-modular emotions alike and therefore do not provide us with the sorts of epistemic guarantees that license scientific generalizations. To illustrate, consider a functional analysis of fear: Fear is a response to threat, which modulates our attention, and brings about certain threat-avoidance behavior. Thus understood, the category *fear* captures modular and non-modular instances of fear. It captures the sorts of "passive" responses you get when you think you have seen a dangerous animal; rapid responses that seem to be triggered without any conscious deliberation. However, it also captures more deliberative patterns of responses, for example, the fear of interest rates rising, which are still "irruptive" to your long-term plans.

As with ecological conceptions of emotion, functional conceptions of emotion prove too inclusive to satisfy the projectability constraint. We lack the required sort of "epistemic warrant" for supposing that discoveries we make about some instances of the category will extend to other members of that category. We see this with the broader category *emotion*, but we also see this with discrete emotion categories, as demonstrated in our functionalist account of fear.

The Developmental Approach

Thus far we have seen that attempts to categorize emotion give way to something resembling a dilemma: Current emotion taxonomies are said to be either (a) scientifically suitable but too reductive or (b) more inclusive but scientifically unsuitable. In what follows, we see how a developmental approach to categorization can take us closer to resolving this dilemma.

The developmental approach I have in mind draws on developmental psychology (Karmiloff-Smith, 1992, 1998, 2009, 2015), neuroconstructivism (Westermann et al., 2007; Karmiloff-Smith et al., 2018), and the developmental systems theory (DST; Griffiths & Gray, 1994, 2004). Central to this approach is Karmiloff-Smith's (1992) notion of "modularization." The basic idea is that instead of thinking of certain cognitive capacities as being underpinned by Darwinian or innately specified modules, we should view such capacities as developing through a process of progressive modularization. That is, the formation of module-like structures through multidirectional interactions between an agent's genes, brain, body, cognition, behavior, and environment.

To give this idea some context, we can note that all three strands of developmental research aim to bypass the nature versus nurture dichotomy. Instead of thinking of a cognitive trait as (mainly) a product of nature (e.g., Darwinian modules) or nurture (e.g., a social construct), we view a trait as being "constrained" by a series of factors, including an organism's genes, brain, body, and physical, and social environment. In that regard, all three strands problematize the assumption that we can isolate a key overarching factor that shapes development. For example, Griffiths and Gray (1994, p. 277) note that the "Developmental systems theory rejects the dichotomous approach to development. The genes are just one resource that is available to the developmental process."

To elaborate, the DST can be understood to comprise two main theses. First, *the Parity Thesis*: Genes are just one of the inputs into a developing system (Griffiths & Gray, 1994, 2004; Oyama, 1985). Second, the thesis of *Extended Inheritance:* We inherit not just genes but a developmental niche, which includes certain social and cultural features, such as customs, norms, and concepts (Heyes 2018; Oyama, 2002; Sterelny, 2003).[5] The upshot is that it no longer makes sense to talk about whether a cognitive trait is a product of biological or cultural evolution (Griffiths & Gray, 1994). Instead, cognitive traits are seen to be shaped by physiological factors (e.g., locomotion), social factors (e.g., parent reactions), and cultural factors (e.g., emotional concepts), as well as genetic factors.

We find a very similar approach to development in neuroconstructivism. The thought here is that cognitive

development is "constructed," as opposed to being innate. However, unlike social constructionist models found in sociology, neuroconstructivists emphasize how such development is "constrained" in certain ways:

> The neuroconstructivist approach characterizes development as a trajectory that is shaped by multiple interacting biological and environmental constraints. The central aspect of understanding cognitive development in this framework is the explanation of how these constraints affect the development of the neural networks of the brain that give rise to progressively more complex mental representations (Westermann et al., 2007, p. 724).

As in the DST, the development of a cognitive trait for neuroconstructivism is seen to be shaped by numerous interacting factors.

It is worth noting that similarities aside, there are differences between the DST and neuroconstructivism. While both frameworks utilize the notion of progressive modularization, they do so in different ways. Within the context of the DST, modularization is an explanation of how certain module-like structures are constructed in phylogenetic development (i.e., the development of traits within a species, e.g., see Mithen, 1996), whereas in the context of neuroconstructivism, it is employed in an explanation of how such structures are constructed in ontogenetic development (i.e., the development of traits during an individual's lifespan, e.g., see Westermann et al., 2007).

Both forms of modularization, I take it, are important for understanding emotion development. However, *emotional* modularization has so far only been discussed with respect to ontogenetic development (Majeed, 2022a, 2022b). The basic idea here is that we need not commit to innately specified modules (or affect programs) to explain why some emotions bear certain hallmarks of modularity, for example, being quick, automatic, and encapsulated from thought. Rather, we can think of some emotions as developing such features through progressive modularization. Since modularization plays a key role in how we are to resolve the categorization dilemma, it is worth spending some time seeing how it works.

My (2022a, 2022b) account of emotional modularization is indebted to Karmiloff-Smith's (1992, 1998, 2009, 2015) more general account of modularization. Perhaps the simplest way to unpack her account is to see what she says about two of the main hallmarks of modularity: domain-specificity (functional specialization) and information encapsulation (insensitivity to background information). First, she replaces the notion of domain-specificity with "domain-relevance." As Karmiloff-Smith (1998, p. 390) notes, "it is more plausible to think in terms of a variety of what one might call domain-relevant mechanisms that might gradually *become* domain-specific as a result of processing different kinds of input." Likewise, instead of supposing that we are born with encapsulated systems, such encapsulation, as well as some of the other hallmarks of modularity, are features that develop through the process of gradual modularization.

Here is a brief outline: Certain domain-relevant mechanisms (in other words, certain biological constraints), along with other developmental factors (e.g., social constraints) dispose or "bias" us towards processing certain kinds of inputs over others. Such biases can lead to the gradual development of brain circuits *progressively* selected to process inputs of that kind. These inputs are processed more often, but also more quickly and automatically. The upshot of this is that such biases can lead to the gradual development of domain-specific systems. However, such specialization also results in the relevant brain circuits becoming less sensitive to what is going on elsewhere in the brain. In other words, such specialization results in these circuits being more and more informationally encapsulated. Overall, this explains how certain developmental constraints can lead to the gradual development of modular characteristics (Karmiloff-Smith, 1992; Majeed 2022a, 2022b).

Precisely how this works at the algorithmic and implementation levels remains to be worked out. A possible way to model the process of modularization might be found in the Dynamical Systems Theory (DST*), which originated in mathematics as a way of explaining the behaviors of complex systems over time (Karmiloff-Smith, 1992). The DST* has proved effective in modeling a range of cognitive skills, and does so in accordance with dynamical systems principles, such as self-organization and emergence (Clark, 1997; van Gelder & Port, 1995; Wheeler, 2005). The suggestion here is that we utilize such principles to generate mathematical models of modularization itself. It is worth noting that there are existing dynamical approaches to emotion that draw on the DST* (Camras & Witherington, 2005; Colombetti, 2014; Lewis, 2005). These accounts do not mention modularization, but future collaborative research between them and the developmental approaches already mentioned could turn out to be mutually beneficial. For example, Colombetti (2014) employs the DST* to offer a way to conceptualize emotional episodes which is markedly different from the way they are conceived by modular (e.g., affect program) accounts. In brief, instead of being the products of rigid, innately specified modules, Colombetti (2014, p. 82) argues that "each emotional episode should be seen as a complex dynamical pattern that has been shaped over evolutionary and developmental time." This promises to be an effective framework from which to examine whether, and if so how, emotions develop module-like traits in accordance with dynamical systems principles.

The picture is slightly clearer for how modularization might be neutrally implemented. One suggestion is that it occurs through a process called "parcellation." Johnson and Karmiloff-Smith (1992, p. 36, footnote*) define this as "the process whereby neural circuits or cognitive structures become isolated from others by means of loss." Moreover, they consider this process to be the equivalent of

modularization at the cognitive level. To clarify, parcellation, as originally conceived, is a process that occurs at the cellular level (Ebbesson, 1980). The claim here is that parcellation, or modularization at the cognitive level, is underpinned by a similar process of loss at the neural level. This in turn could be explained by the more familiar notion of synaptic pruning. That is, the weakening and ultimate elimination, or "pruning," of unused synapses during brain development (Karmiloff-Smith, 1992).

A similar account is also provided by "neurodynamical" approaches to emotion, which draw on the DST* to explain structural changes in the brain during emotion development (Colombetti, 2014; Lewis & Liu, 2011). According to such approaches, the exponential synaptic growth that occurs during early development is followed by a process of synaptic pruning. There are, however, slight differences in detail. For example, neural development is explained here in terms of neural Darwinism, not parcellation. To elaborate, according to the parcellation picture, isolation between neural circuits is a product of neural selection at the level of individual synapses (Ebbesson, 1980), whereas according to that of neural Darwinism, this occurs at the level of neuronal groups (Edelman, 1987). Details aside, what is striking is that both approaches take neural isolation to be a significant feature of emotion development. It is my view that such isolation will prove crucial to an explanation of how emotional modularization is realized in the brain.

Of course, much of this is speculative and can presently be put aside. The key point for us is that modularization, regardless of how it works algorithmically or neurally, also helps explain how emotions can acquire certain hallmarks of modularity, including those often associated with the notion of "passivity" in the vernacular category emotion, such as automaticity and encapsulation (Majeed, 2022a, 2022b). This holds important lessons for emotion categorization.

Revisiting the Projectability Concerns

To recap, categories are said to be projectable if they group phenomena based on some similarity-generating "homeostatic" mechanism: a mechanism that explains why certain properties cluster together in a projectable manner. The suggestion here is that modularization itself can be thought of as a homeostatic mechanism in this sense. To begin, we can follow Fodor (1983) in observing that if a cognitive system has some of the hallmarks of modularity (e.g., being quick and automatic), it is also likely to have others (e.g., being relatively encapsulated). One way of explaining the existence of such clusters is by positing an innate program or module. However, this is not the only way. Modularization can also explain why such features cluster together. For instance, relative encapsulation can be seen to be a side effect of our emotion-generating processes becoming more quick and automatic (Majeed, 2022a, 2022b).

We can exploit this idea to explain how categorizing emotions based on modularization might deliver projectable emotion categories. One way of categorizing emotions is in terms of "passivity": their independence of our long-term planned actions (Griffiths, 1997). We can observe that emotions are often passive because they bear certain hallmarks of modularity. For instance, they are quick, automatic, and encapsulated from thought—all ways that mark them as independent of our long-term plans. Griffiths (1997) explains these features of emotion in terms of modular affect programs. However, we can also explain why such features cluster together by appealing to modularization. The upshot is that categorizing emotions based on a certain modular type of passivity, insofar as it mirrors a classificatory system that tracks the developmental process of modularization, will deliver categories that are projectable.

For example, growing up with a parent who fears flying might gradually tune your fear responses to track conditions related to air travel, in such a way that these responses become more quick and automatic, and at the same time, less sensitive to your thoughts; in other words, more informationally encapsulated. This explains how you might develop a fear of flying that exhibits features associated with a modular kind of passivity. Significantly, a similar process of modularization might also explain other passive fear responses that bear hallmarks of modularity, for example, someone's fear of lifts which proves quick, automatic, and insensitive to their thoughts. It is worth noting that modularization is not an inevitable developmental process. The various constraints that act on development ensure that no cognitive trait develops in a predetermined trajectory. So for instance, you might have a fear of interest rates rising, but it might remain an emotion that is brought on by conscious deliberation and is sensitive to your background assumptions. The point relevant for us with respect to categorization is that "instances of fear that exhibit a modular kind of passivity," to the extent that they exhibit such passivity on account of modularization, will prove to be a projectable sub-category in its own right. There is, in other words, a similarity-generating mechanism (viz. modularization), which ensures that we can reliably extrapolate from samples of the category to the other instances of that category.

Crucially, the account applies to complex emotions, such as guilt and jealousy, as well as basic emotions, such as fear and anger. The claim is not that both *types* of emotion are modular, or even take on the appearance of modularity. Rather, we can begin by recognizing that both types of emotion have *instances* that look modular. For instance, as Griffiths (1997) points out, there are instances of fear that look modular (e.g., a fear of snakes) and non-modular (e.g., a fear of interest rates rising). However, going beyond Griffiths (1997), we can also note that the same can be said of complex emotions. For example, though jealousy is regarded as a complex emotion, *some* instances of jealousy also bear certain hallmarks of modularity,

for example, they are quick, automatic, and encapsulated. (Majeed, 2022a). The point is that when we encounter the appearance of modularity in emotions—be it in instances of basic or complex emotions—we can explain such appearances, not by positing innate emotion modules, but via the developmental process of gradual modularization.

This turns out to be important for emotion categorization. Earlier we saw that while the vernacular notion of emotion might be defined as "motivational states that exhibit passivity," there is (allegedly) no epistemic warrant for supposing that discoveries about some emotions will extend to all other emotions. Why not? The main reason is that they result from different underlying homeostatic mechanisms: Basic emotions are passive (in a modular sense) because they are triggered by modules, whereas complex emotions are passive (in a non-modular sense) because of some other mechanism (Griffiths, 1997). By contrast, insights from the DST and neuroconstructivism suggest that at least some of the property clusters that make up the emotional phenotype (including a certain modular type of passivity), for both basic and complex emotions, might be products of the same underlying homeostatic mechanism, that is, modularization. If that is right, it no longer makes sense to divide emotions into basic (modular) and complex (non-modular) emotion categories, only one of which is suitable for scientific practice. Instead, we will have modularized and non-modularized instances of the same emotion sub-categories, for example, modularized and non-modularized fear, modularized and non-modularized jealousy, and so on. The present proposal is that modularized emotion sub-categories (i.e., the modular passive ones) can form projectable categories suitable for a science of emotion.

To elaborate, earlier we saw that attempts to categorize emotion face a dilemma: Our emotion taxonomies are said to be either (a) scientifically suitable (i.e., projectable) but too reductive or (b) more inclusive but scientifically unsuitable (i.e., not projectable). Our developmental account can shed light on how we might resolve this dilemma. That is, it can help explain how certain emotion sub-categories are projectable. They are projectable because they stem from the same similarity-generating mechanism, that is, the developmental process of gradual modularization. Moreover, since the process explains the projectability of a sub-category of both basic and complex emotions alike, it also captures a wider set of emotions than modular emotion taxonomies.

Why not, then, say that the account solves the dilemma? I doubt modularization, or any other developmental notion, can carry the burden of emotion categorization, all by itself. In particular, it does not address the broader, and more contentious, question of what emotions really are. (To illustrate, even modular features, such as automaticity and encapsulation, don't help us distinguish emotions from non-emotional phenomena that also happen to be automatic and encapsulated). For what it is worth, I do not think there is such a thing as an adequate definition of emotion. The notion has been used in so many different contexts (folk, philosophical, scientific, historical, cultural, and so on) that we are unlikely to find a definition that will satisfy everyone (see also Duffy, 1941; Moors, 2022; Russell, 2003). However, such skepticism aside, we can note that there are ways of defining or analyzing emotion that are informative and further certain research ends. For instance, we can analyze emotions in terms of their ecological roles, functional roles, or evaluative content. The trouble with such analyses is that they seem incapable of delivering projectable categories of the sort required for a science of emotion (Griffiths, 1997). Nevertheless, *pace* Griffiths (1997), this need not be a problem.

The trick is to recognize that we can prise apart the task of defining or analyzing emotion from the task of explaining its projectability, both of which are relevant to emotion categorization. Thus understood, you could adopt your preferred conception of emotion and explain the projectability of some of their sub-categories (e.g., the modular passive ones) by appealing to modularization. If you are drawn to functionalism, say, you could define emotions in terms of whatever occupies the relevant functional roles. For example, you could claim that fear is a response to threat, which modulates our attention and brings about certain threat-avoidance behavior. Such an analysis is broad enough to capture modular and non-modular fear responses alike. The suggestion here is that we can explain the projectability of a sub-category of such fear responses, that is, the modular "passive" ones, by appealing to modularization. The take-home message, then, is that modularization will not make all our vernacular emotion categories suitable for a science of emotion, but it might render a lot of their sub-categories scientifically suitable.

Revisiting the Reductive Concerns

To end, let us consider how our account handles the specific worries to do with modular emotion taxonomies being too reductive. The first concern is relatively easy to dispense with. This is the concern that to categorize emotions in terms of a limited set of modular-cum-basic emotions is to fail to represent the sheer variety of emotions we can experience. As we saw, modularization can help us explain the projectability of complex emotions as well as basic ones. Here we should not forget the point that we might even experience emotions beyond those for which we have words (de Sousa, 2008). It is hard to say too much about such emotions without a clear account of what they are supposed to be like. Nevertheless, we can observe that there is no reason why modularization cannot operate on the mechanisms that underlie such emotions.

For example, suppose there is a certain kind of affective phenomenon that is halfway between guilt, shame, and anger, which passes the benchmark for counting as an emotion in its own right. Say it has its own unique appraisal

criteria; ones that are too complex to be obtained by simply adding those of other emotions (Griffiths, 2002). It might manifest itself when you feel something akin to shame and guilt for having wronged someone, but you also feel anger at that person for making you feel that way. Such an emotion most likely will not be evolutionarily ancient, but some of its instances might still undergo modularization, and in such a way that they take on the appearance of a basic emotion. That is, they could become quick, automatic, specific to certain situations, encapsulated from thought, and so on. To the extent that instances of this emotion can actually undergo modularization, they can form a projectable emotion category suitable for scientific inquiry. Put another way, modularization can help explain why some instances of the category have certain modular features, and in a way that we can reliably extrapolate from these samples to other (modularized) instances of the category.

The second concern has to do with emotional consciousness. Darwinian modules are increasingly identified with activity in the subcortical regions of the brain responsible for the physiological and behavioral components of emotion, as opposed to activity in the neocortex which is more likely to account for their conscious components. In a way, the proponent of emotional modularization does not have much to say about this issue, as an appeal to modularization does not involve a commitment to any specific claims about regions of the brain responsible for emotion. Nevertheless, this also proves to be an advantage, for they can bypass the charge that our emotion taxonomies leave out the conscious component of emotion. To this we can add that modularization can, in theory, operate on neural circuitry in both the subcortical and cortical regions of the brain. For example, a threat response brought on by a subcortical neural circuit might become modularized over time such that it takes on the appearance of a modular response. Likewise, a conscious fear response brought on by activity in the neocortex might also gradually become modularized in such a way that it is triggered in a quick, automatic, encapsulated, and domain-specific manner. Therefore, to categorize emotions on the basis of modularization is inclusive of both conscious and non-conscious emotional phenomena.

Addressing the third concern proves more difficult. This is the existentialist worry that short-term impulses, such as those triggered by Darwinian modules, will not explain all the ways our emotions help situate ourselves in the world. This is a problem for a science of emotion. The various disciplines in science that study emotion, such as cognitive psychology, cognitive neuroscience, and evolutionary biology, have nothing to say about long-term emotions (i.e., except as dispositions to have short-lived emotional episodes). This has led some to dismiss such emotions, for example, love, as "essentially social pretenses" (Griffiths, 1997). By contrast, the "soft sciences," such as social science, social psychology, and anthropology, have plenty to say about such emotions. To my mind, an adequate response to the existentialist cannot be found until we embark on wider interdisciplinary endeavors, where we seek to bring insights from disciplines such as social psychology, anthropology, and philosophy of mind, to the more "serious" sciences, such as cognitive psychology and cognitive neuroscience. That said, a psychological science of the mind must start somewhere. I have argued that appealing to modularization instead of Darwinian modules provides us with a more inclusive starting position.

Acknowledgment

I am grateful to two referees for this journal, as well as the previous editor Agnes Moors, for their helpful comments.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Raamy Majeed https://orcid.org/0000-0001-8864-1643

Notes

1. Some (e.g., Griffiths,1997) take projectability to be the basis of what makes a category a natural kind, though we should take care to note that the term "natural kind" has been employed in a variety of different, and often incompatible, ways (Hacking, 2007).
2. Ekman (1973) also argues that basic emotions are families with shades, for example., the anger family comprises phenomena such as irritation, annoyance, and rage. This makes his account more inclusive than it initially appears.
3. Of course, not all philosophers of emotion look elsewhere. For example, Prinz (2004) offers an empirically driven, "partially modular", approach to emotion categorization that establishes a "unity" between basic and complex emotions. (In brief, he does so by treating both types of emotions as embodied appraisals). However, significantly, Prinz (2004) does not address how such a unified category can meet the projectability constraint.
4. Note that Griffiths (1997) includes functional taxonomies in his critique of ecological emotion taxonomies, but his focus is on accounts that conceive of emotions in terms of their *adaptive* function. Functionalist accounts in the philosophy of mind do not typically conceive of function in this way.
5. This is a point often emphasized in situated approaches to cognition, which take cognitive processes to be "embedded" in our social and physical environment (Robbins & Aydede, 2008). For situated approaches to emotion, see Griffiths and Scarantino (2005) and Greenwood (2015).

References

Adolphs, R., & Anderson, D. J. (2018). *The neuroscience of emotion: A new synthesis.* Princeton University Press.

Adolphs, R., & Andler, D. (2018). Investigating emotions as functional states distinct from feelings. *Emotion Review*, *10*(3), 191–201. https://doi.org/10.1177/1754073918765662

Armstrong, D. M. (1968). *A materialist theory of the mind*. Routledge.

Arnold, M. B. (1960). *Emotion and personality, vol. 1*. Columbia University Press.

Barrett, L. F. (2013). Psychological construction: The Darwinian approach to the science of emotion. *Emotion Review*, *5*(4), 379–389. https://doi.org/10.1177/1754073913489753

Barrett, L. F. (2017). *How emotions are made: The secret life the brain*. Houghton-Mifflin-Harcourt.

Boyd, R. (1989). What realism implies and what it does not. *Dialectica*, *43*(1–2), 5–29. https://doi.org/10.1111/j.1746-8361.1989.tb00928.x

Braddon-Mitchell, D., & Jackson, F. (2006). *Philosophy of mind and cognition: An introduction*. Wiley-Blackwell.

Buller, D. J. (2005). Evolutionary psychology: The emperor's new paradigm. *Trends in Cognitive Sciences*, *9*(6), 277–283. https://doi.org/10.1016/j.tics.2005.04.003

Camras, L. A., & Witherington, D. C. (2005)). Dynamical systems approaches to emotional development. *Developmental Review*, *25*(3–4), 328–350. https://doi.org/10.1016/j.dr.2005.10.002

Charland, L. (1995). Emotion as a natural kind: Towards a computational foundation for emotion theory. *Philosophical Psychology*, *8*(1), 59–84. https://doi.org/10.1080/09515089508573145

Clark, A. (1997). *Being there: Putting brain, body, and world together again*. MIT Press.

Colombetti, G. (2014). *The feeling body: Affective science meets the enactive mind*. MIT Press.

Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and the emotions. In M. Lewis, & J. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed, pp. 91–115). The Guilford Press.

Cowen, A., Sauter, D., Tracy, J. L., & Keltner, D. (2019). Mapping the passions: Toward a high-dimensional taxonomy of emotional experience and expression. *Psychological Science in the Public Interest*, *20*(1), 69–90. https://doi.org/10.1177/1529100619850176

Craver, C. F. (2009). Mechanisms and natural kinds. *Philosophical Psychology*, *22*(5), 575–594. https://doi.org/10.1080/09515080903238930

Damasio, A. (1994). *Descartes' error: Emotion, rationality and the human brain*. Avon Books.

Deonna, J. A., & Teroni, F. (2012). *The emotions: A philosophical introduction*. Routledge.

de Sousa, R. B. (1987). *The rationality of emotion*. MIT Press.

de Sousa, R. B. (2008). Against emotional modularity'. In L. Faucher, & C. Tappolet (Eds.), *The modularity of emotions* (pp. 29–50). University of Calgary Press.

Duffy, E. (1941). An explanation of "emotional" phenomena without the use of the concept "emotion". *Journal of General Psychology*, *25*(2), 283–293. https://doi.org/10.1080/00221309.1941.10544400

Ebbesson, S. O. (1980). The parcellation theory and its relation to interspecific variability in brain organization, evolutionary and ontogenetic development, and neuronal plasticity. *Cell and Tissue Research*, *213*(2), 179–212. https://doi.org/10.1007/BF00234781

Ekman, P. (1973). *Darwin and facial expression*. Academic Press.

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic? *Emotion Review*, *3*(4), 364–370. https://doi.org/10.1177/1754073911410740

Eldeman, G. M. (1987). *Neural Darwinism: The theory of neuronal group selection*. Basic Books.

Fodor, J. A. (1983). *The modularity of mind*. MIT Press.

Frank, R. H. (1988). *Passions within reason: The strategic role of the emotions*. Norton.

Frijda, N. H. (1986). *The emotions*. Cambridge University Press.

Goldie, P. (2000). *The emotions: A philosophical exploration*. Oxford University Press.

Goodman, N. (1954). *Fact, fiction and forecast* (1st ed.). Athlone Press, University of London.

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of The Royal Society Of London. Series B. Biological Sciences*, *205*(1161), 581–598. https://doi.org/10.1098/rspb.1979.0086

Greenwood, J. (2015). *Becoming human: The ontogenesis, metaphysics, and expression of human emotionality*. MIT Press.

Griffiths, P. E. (1997). *What emotions really are: The problem of psychological categories.* Chicago University Press.

Griffiths, P. E. (2002). Is emotion a natural kind? In R. C. Solomon (Ed.), *Thinking about feeling: Contemporary philosophers on emotions* (pp. 233–249). Oxford University Press.

Griffiths, P. E. (2003). Basic emotions, complex emotions, machiavellian emotions. *Royal Institute of Philosophy Supplements*, *52*, 39–67. https://doi.org/10.1017/S1358246100007888

Griffiths, P. E. (2004). Emotions as natural and normative kinds. *Philosophy of Science*, *71*(5), 901–911. https://doi.org/10.1086/425944

Griffiths, P. E., & Gray, R. D. (1994). Developmental systems and evolutionary explanation. *Journal of Philosophy*, *91*(6), 277–304. https://doi.org/10.2307/2940982

Griffiths, P. E., & Gray, R. D. (2004). The developmental systems perspective: Organism-environment systems as units of development and evolution. In K. Preston, & M. Pigliucci (Eds.), *The evolutionary biology of complex phenotypes* (pp. 335–339). Oxford University Press.

Griffiths, P. E., & Scarantino, A. (2005). Emotions in the wild: The situated perspective on emotion. In P. Robbins, & M. Aydede (Eds.), *Cambridge handbook of situated cognition* (pp. 437–453). Cambridge University Press.

Hacking, I. (2007). Natural kinds: Rosy Dawn, scholastic twilight. *Royal Institute of Philosophy Supplements*, *61*, 203–239. https://doi.org/10.1017/S1358246100009802

Heyes, C. (2018). *Cognitive gadgets: The cultural evolution of thinking*. Harvard University Press.

Izard, C. E. (1977). *Human emotions*. Springer Science & Business Media.

Johnson, M. H., & Karmiloff-Smith, A. (1992). Can neural selectionism be applied to cognitive development and its disorders? *New Ideas in Psychology*, *10*(1), 35–46. https://doi.org/10.1016/0732-118X(92)90046-3

Jones, K. (2008). Quick and smart? Modularity and the pro-emotion consensus. In L. Faucher, & C. Tappolet (Eds.), *The modularity of emotions* (pp. 3–27). University of Calgary Press.

Karmiloff-Smitch, A. (1992). *Beyond modality: A developmental perspective on cognitive science*. MIT Press.

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences*, *2*(10), 389–398. https://doi.org/10.1016/S1364-6613(98)01230-3

Karmiloff-Smith, A. (2009). Nativism versus neuroconstructivism: Rethinking the study of developmental disorders. *Developmental Psychology*, *45*(1), 56–63. https://doi.org/10.1037/a0014506

Karmiloff-Smith, A. (2015). An alternative to domain-general or domain-specific frameworks for theorizing about human evolution and ontogenesis'. *AIMS Neuroscience*, *2*(2), 91–104. https://doi.org/10.3934/Neuroscience.2015.2.91

Karmiloff-Smith, A., Thomas, M. S. C., & Johnson, M. H. (2018). *Thinking developmentally: From constructivism to neurconstructivism*. Routledge.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. MIT Press.

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press.

LeDoux, J. E. (1996). *The emotional brain.* Simon and Schuster.

LeDoux, J. E. (2016). *Anxious: Using the brain to understand and treat fear and anxiety*. Viking.

Lewis, M. D. (2005). Bridging emotion theory and neurobiology through dynamical systems modelling. *Behavioral and Brain Sciences*, *28*(2), 169–245. https://doi.org/10.1017/S0140525X0500004X

Lewis, M. D., & Liu, A. (2011). Three time scales of neural self-organization underlying basic and nobasic emotions. *Emotion Review*, *3*(4) 416–423. https://doi.org/10.1177/1754073911410748

Loaiza, J. R. (2024). Functionalism and the emotions. *British Journal for the Philosophy of Science*, *75*(1), 1–34. https://doi.org/10.1086/715207

Maclean, P. D. (1952). Some psychiatric implications of physiological studies on frontotemporal portion of limbic system (visceral brain). *Electroencephalography & Clinical Neurophysiology*, *4*(4), 407–418. https://doi.org/10.1016/0013-4694(52)90073-4

Majeed, R. (2020). The new LeDoux: Survival circuits and the surplus meaning of 'fear'. *Philosophical Quarterly*, *70*(281), 809–829. https://doi.org/10.1093/pq/pqaa009

Majeed, R. (2022a). The "puzzle" of emotional plasticity. *Philosophical Psychology*, *35*(4), 546–568. https://doi.org/10.1080/09515089.2021.2002293

Majeed, R. (2022b). Does the problem of variability justify Barrett's emotion revolution? *Review of Philosophy and Psychology*, *14*, 1421–1441. https://doi.org/10.1007/s13164-022-00650-0

Majeed, R. (2022c). Modularity and the politics of emotion categorization. A tribute to Ronald de Sousa. https://www.unige.ch/cisa/related-sites/ronald-de-sousa/assets/pdf/Majeed_Paper.pdf

Mayr, E. (1974). Behavior programs and evolutionary strategies. *American Scientist*, *62*, 650–659. http://www.jstor.org/stable/27845170

Mithen, S. (1996). *The prehistory of the mind*. Thames and Hudson Ltd.

Moors, A. (2022). *Demystifying emotions: A typology of theories in psychology and philosophy*. Cambridge University Press.

Nussbaum, M. C. (2001). *Upheavals of thought: The intelligence of emotions*. Cambridge University Press.

Oyama, S. (1985). *The ontogeny of information: Developmental systems and evolution*. Cambridge University Press.

Oyama, S. (2002). The nurturing of natures. In A. Grunwald, M. Gutmann, & E. M. Neumann- Held (Eds.), *On human nature: Anthropological, biological and philosophical foundations* (pp. 163–170). Springer.

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions.* Oxford University Press.

Plutchik, R. (2003). *Emotions and life: Perspectives from psychology, biology, and evolution*. American Psychological Association.

Prinz, J. (2004). *Gut reactions: A perceptual theory of emotion*. Oxford University Press.

Putnam, H. (1960). Minds and machines. In H. Putnam (Ed.), *Mind, language, and reality* (pp. 362–385). Cambridge University Press.

Robbins, P., & Aydede, M. (Eds.). (2008). *The Cambridge handbook of situated cognition*. Cambridge University Press.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. https://doi.org/10.1037/0033-295X.110.1.145

Scarantino, A. (2015). Basic emotions, psychological construction and the problem of variability. In J. Russell, & L. F. Barrett (Eds.), *The psychological construction of emotion* (pp. 334–376). Guilford Press.

Scarantino, A. (2018). Are LeDoux's survival circuits basic emotions under a different name? *Current Opinion in Behavioral Sciences*, *24*, 75–82. https://doi.org/10.1016/j.cobeha.2018.06.001

Scherer, K. R. (1993). Studying emotion-antecedent appraisal process: An expert system approach. *Cognition And Emotion*, *7*(3–4), 325–355. https://doi.org/10.1080/02699939308409192

Solomon, R. C. (2007). *True to our feelings: What our emotions are really telling us*. Oxford University Press.

Sterelny, K. (2003). *Thought in a hostile world*. Blackwell.

Sterelny, K., & Griffiths, P. E. (1999). *Sex and death: An introduction to philosophy of biology*. University of Chicago Press.

Tomkins, S. S. (1962). *Affect, imagery, consciousness*. Springer.

van Gelder, T. J., & Port, R. F. (1995). It's about time: An overview of the dynamical approach to cognition. In R. F. Port, & T. J. van Gelder (Eds.), *Mind as motion* (pp. 1–43). MIT Press.

Westermann, G., Mareschal, D., Johnson, M. H., Sirois, S., Spratling, M. W., & Thomas, M. S. (2007). Neuroconstructivism. *Developmental Science*, *10*(1), 75–83. https://doi.org/10.1111/j.1467-7687.2007.00567.x

Wheeler, M. W. (2005). *Reconstructing the cognitive world: The next step*. MIT Press.