Modularity and the Politics of Emotion Categorisation

Raamy Majeed
The University of Manchester
A Tribute to Ronald de Sousa (2022)

ABSTRACT: Empirically-informed approaches to emotion often construe our emotions as modules: systems hardwired into our brains by evolution and purpose-built to generate certain coordinated patterns of expressive, physiological, behavioural and phenomenological responses. In 'Against Modularity' (2008), de Sousa argues that we shouldn't think of our emotions in terms of a limited number of modules because this conflicts with our aspirations for a life of greater emotional richness. My aim in this paper is to defend de Sousa's critique of modular emotion taxonomies from some obvious rejoinders, and to develop his positive proposal as to how we might reconcile the evidence for emotional modularity with an attitude of disapproval towards rigid emotion taxonomies.

'I want to write the moral history of the men of my generation—or, more accurately, the history of their feelings. It's a book about love, about passion; but passion such as can exist nowadays—that is to say, inactive.' Flaubert, *Sentimental Education*.

1. Introduction

Should we think of our emotions in terms of a limited number of emotion modules? That is, should we think of emotions as systems hardwired into our brains by evolution and purpose-built to generate certain coordinated patterns of expressive, physiological, behavioural and phenomenological responses? In his paper, 'Against Emotional Modularity', de Sousa argues that we shouldn't because this conflicts with our aspirations for a life of greater emotional richness.

As I read him, the title of de Sousa's paper is not to be taken literally, as his main concern is not really emotional modularity *per se* — he concedes that there might be such modules — but taxonomies based on such modules. Criticism of modular accounts of emotion isn't new. Nor for that matter is dissatisfaction with the taxonomies they inspire. But I think the paper stands out for the following reasons.

First, it helps disentangle two sorts of worries with emotional modularity that are typically run together: whether emotion modules exist and whether we should categorise our emotions based on such modules. It is the latter which de Sousa pursues by taking a 'political stance' on the issue.

Second, while dissatisfaction with modular emotion taxonomies have been voiced before, the exact source of the dissatisfaction, as far as I can tell, has been hard to locate. (No doubt, this is in part because this dissatisfaction is often run together with scepticism about emotional modularity more generally). For de Sousa, modular taxonomies are problematic because they (i) obscure the roles our emotions play with respect to our individual well-being, and (ii) misrepresent our emotional repertoire by misrepresenting the variety of emotions we can experience.

Third, de Sousa takes some positive steps towards explaining how we might reconcile the evidence for emotional modularity with an attitude of disapproval towards rigid emotion taxonomies. In particular, he offers an 'Aesthetic Model' of emotion taxonomies to rival those that stem from emotional modality; taxonomies that carve emotions into a small set of 'basic' (or universal) emotions, e.g., happiness, sadness, anger, disgust, surprise and fear. By contrast, the new model involves drawing on mediums such as art and literature to appreciate the 'multidimensional field' of emotions it is possible for us to experience.

My aim in this paper is to explain, defend and develop de Sousa's proposal. The paper is structured as follows. First, I give a brief example of a modular emotion taxonomy and explain the two main reasons why de Sousa finds it problematic (Section 2). I then articulate some obvious rejoinders to these problems and explain why they are not so easily overcome (Sections 3 and 4). Finally, I outline de Sousa's positive proposal, and explain how it helps us reconcile the evidence for emotional modularity with an attitude of disapproval towards rigid emotion taxonomies (Section 5).

2. Background

Categorisation is something of a necessary evil. Explanations require categorisations that abstract from detail but such abstractions, by their very nature, oversimplify complex phenomena. Categorisation of emotion proves no exception. To characterize our emotions into categories such as 'anger' or 'love' is to abstract away from the rich detail manifest in instances that fall under these categories. This much is understood. However, the sorts of emotion categories which result from empirically-informed approaches to emotion have proved especially pernicious.

Take the picture of emotion that emerges from evolutionary psychology. On this view, our emotions are, or are the products of,' Darwinian modules': systems/mechanisms/programs hardwired into our brains by evolution and purpose-built to generate certain coordinated patterns of expressive, physiological, behavioural and (perhaps) phenomenological responses.[1] Darwinian modules are typically thought to be 'domain-specific', i.e., they process only certain kinds of information and in specific ways. For instance, a fear module only triggers in response to specific stimuli, namely those perceived as threatening, e.g., loud noises, cliff edges, scary animals etc. Moreover, it elicits characteristic patterns of responses, e.g., when you are fearful, your heartrate accelerates, your muscles tremble, you begin to sweat, your breathing becomes short, you become more alert and poised for defensive action, and so on. Such changes may also be accompanied by certain phenomenological profiles, such as feeling a tightness in your chest, butterflies in your stomach, dizziness etc.[2]

[1] See Tooby and Cosmides (1990), Pinker (1997), Cosmides and Tooby (2000) and Sperber (2002). Note: Darwinian modules are also identified with 'affect programs' and 'basic emotions', e.g. see Tomkins (1962), Ekman (1973, 1999), Izard (1992), Griffiths (1997), Panksepp (1998), Panksepp and Watt (2011), Scarantino and Griffiths (2011), and Scarantino (2015). While these notions are not strictly speaking identical, I will follow de Sousa in treating them as such for the purposes of this paper.

[2] The other major conception of modularity comes from Fodor (1983) although Fodor does not discuss emotions specifically. This account posits several different 'hallmarks' of modality, including domain-specificity. See Faucher and Tappolet (2008) for an overview of the various notions of emotional modularity.

Contemporary proponents of this view don't take Darwinian modules to correspond one-to-one with our folk emotion categories. For instance, it is not their claim that there are modules for emotions as complex and culture-bound as Schadenfreude or romantic love. But they do think that there are hardwired emotion modules which we can take as the basis for a limited set of 'basic emotions'. Some of these modules might correspond to our folk categories, e.g., perhaps there is a module for the folk category 'fear'. However, this needn't be the case. Perhaps there is no specific fear module *per se*, but we might find modules for related phenomena, e.g., a module for anxiety and another for panic. What will end up on the final list of basic emotions is up for grabs. What isn't is the type of taxonomies that arise when we model emotions based on Darwinian modules. Regardless of how we specify the details, we end up with emotion taxonomies which treat emotions as a limited set of traits that evolved because they served some adaptive purpose.

While modular taxonomies of the sort mentioned above have proved popular, especially in evolutionary psychology, neuroscience and empirically-informed philosophy of emotion, they have always been controversial. Much of the controversy stems from whether there are actually any emotion modules to begin with. De Sousa identifies two motivations for positing emotion modules. The first concerns the phenomenon of emotional recalcitrance. Recalcitrant emotions are emotions that are in tension with your considered judgements, e.g., you fear your neighbour's toothless old dog despite judging it to be harmless.[3] Since recalcitrant emotions involve a conflict between emotion and judgement, they give us (defeasible) grounds to suppose that there might be systems in the brain specific to emotion-generation which come apart from those that involve cognition.[4] The second motivation is supposed to be more telling. This concerns neurobiological data which (allegedly) reveal specific neural circuits for some basic emotions. For example, LeDoux (1996) is widely recognised as having discovered the neural profile for fear: an amygdala-centred circuit that comprises of activity in the sub-cortical regions of the brain.[5]

---

[3] See Greenspan (1981), D'Arms and Jacobson (2003), Benbaji (2013) and Grzankowski (2016).

[4] Recalcitrant emotions, or related phenomena such as 'irrational emotions', are often cited as evidence for emotional modularity, e.g. see de Sousa (1987), Frijda (1986), Griffiths (1990, 1997), Charland (1995), Goldie (2000), Prinz (2004, 2008), Faucher and Tappolet (2008b), Jones (2008), Tappolet (2016) and Majeed (2019). Note: some of these discussions focus on the Fodorian conception of modularity instead of the Darwinian conception at issue in this paper.

[5] See Griffiths (2002, 2004a). Also see Scarantino (2018) and Majeed (2020) for a discussion. Note: this also gives us evidence for another Fodorian 'hallmark' of modularity, namely they are associated with particular neural structures.

Enthusiasm for emotion modules, however, has been somewhat dampened on account of the fact that meta-analyses of various neurobiological studies of emotion don't seem to reveal any specific 'biological fingerprints' for any of our folk emotion categories (Barrett 2013; 2016). For example, despite early evidence from neuroscience, the data don't reveal any specific neural circuitry for fear. To be clear, the amygdala is implicated in fear, but it is also implicated in various other phenomena, e.g., how we respond to novelty. What is missing is a specific amygdala-centred circuitry that acts as a constant across the various neurobiological studies of fear. Barrett takes this as evidence for the absence of emotion modules, whereas others take this to simply show that there are no modules that correspond neatly with our folk emotion categories. For instance, as Scarantino (2015) points out, Barrett's data don't rule out the existence of multiple emotion-generating mechanisms for our basic emotion category 'fear', some of which might be relevant for unconditioned fear, others for conditioned fear, and so on.

So are there Darwinian emotion modules? And if so, should we categorise emotions based on such modules? Though these two questions are often posed together, we learn from de Sousa that how we answer the first needn't bear on how we answer the second. That is, there might be emotion modules, but this doesn't determine the attitude we should take towards categorising emotions on the basis of such modules. De Sousa himself is neutral as to how we might answer the first question. (I am inclined to agree. For the time being, I think it's prudent to adopt a policy of 'let's wait and see'). However, he thinks we can give a more decisive answer to the second question.

As noted earlier, all taxonomies, by their very nature, oversimplify complex phenomena. Emotion taxonomies based on Darwinian modules prove no exception. But the precise way they do so is especially problematic. For one thing, just because a trait was adaptive in the Environment of Evolutionary Adaptation (EEA) doesn't mean it is useful for us now. For another, even if the trait still serves some adaptive purpose, traits are selected at the species-level, not because they are beneficial for every individual that happens to be a member of that species. These points suggest that to categorise emotions based on which traits were adaptive in the EEA is to lose sight of the roles emotions play with respect to our subjective, individual well-being. What's more -- and this I think is de Sousa's main motivation for raising this worry – they also ignore the non-instrumental value of emotion. Emotions matter for our well-being not just because they satisfy some end (adaptive or

otherwise). They also matter for their own sake. To pick an obvious example, we tend to think that being happy is important in itself, above and beyond whatever else it may bring.

The other major problem is also one of omission. All taxonomies abstract from detail, but most typically don't do so in a way that ignore the sorts of detail which pose a threat to the very taxonomies on offer. The trouble with emotion taxonomies based on Darwinian modules is that they ignore the richness of our emotional experiences — feature that, arguably, call into question the simple bifurcation of emotions into a rigid set of limited categories. The phenomenal profiles of feeling heartbroken on account of being romantically shunned and feeling grief for a loved one who has passed away, for example, are arguably much too disparate and varied to be lumped under the common banner of 'sadness'. Simple modular taxonomies, thereby, misrepresent our emotional repertoire because they ignore the sheer variety of emotions we experience.

Together, these two problems motivate an altogether different take on the modularity debate: even if emotion modules exist, we shouldn't categorise our emotions based on such modules. So what are our options? One is to resist these problems by demonstrating that modular accounts of emotion can, despite appearances, accommodate our individual goals as well as the richness of our emotional experiences. The other is to offer a way to categorise emotions that doesn't rely on emotion modules and yet accommodates any potential evidence for their existence. In what follows, I raise some worries with the first option and motivate the second.

3. Modularity & Individual Goals

On the face of it, there are obvious rejoinders to de Sousa's criticisms of modular emotion taxonomies. Take the first problem: (i) such taxonomies obscure the roles our emotions play with respect to our individual well-being. Surely evolutionary psychologists have something to say about how evolved modules can be coopted to meet individual ends I think they do, but I don't think what they have to say will completely address de Sousa's worry.

One of the main problems with Darwinian modules is that such modules are supposed to be incapable of undergoing any social or cultural learning.[6] Darwinian modules are supposed to be relatively fixed systems which trigger automatically in response to a certain class of innately-specified stimuli. The existence of such modules, however, is called into question because there aren't many stimuli we respond to in this fashion. As Deonna and Teroni note, 'Trotting out shopworn examples of snakes and spiders to delineate the class of stimuli relevant in the case of the fear affect program is often the best we can do in employing this strategy' (2012, 26). I think that's right, but we should be clear on the distinction between modules that are input-rigid and those that are input-flexible. An input-rigid system only triggers in response to a select class of innately-specified properties, whereas an input-flexible system is capable of learning which stimuli are relevant on the basis of environmental cues.[7]

To be clear, input-flexible systems, despite exhibiting plasticity, still operate under certain innate principles. In particular, they come equipped with certain learning biases that make it more likely that they will 'learn' to respond to some stimuli rather than others, a phenomenon called 'learning preparedness' (Seligman 1970). For instance, a fear of snakes mightn't be innate, but it is easier to condition and harder to extinguish than a fear of arbitrary stimuli such as shapes (Seligman 1971). However, even if our emotion systems actually manifest learning preparedness, this doesn't preclude them from learning to respond to various sorts of stimuli they weren't selected for, including those that weren't around in the EEA. This is important because it means an input-flexible modular system could, in theory, learn to respond to cues that align with an individual's own goals, their own sense of self. For example, an evolved disgust module most likely won't be innately-specified to trigger in response to the sight of a used syringe lying around a park, but such a module might, through social conditioning, be 'recalibrated' to track such socially-salient stimuli (Prinz 2004).

This sort of response raises two sorts of questions. First, how does the process of recalibration actually work? And second, assuming it does work, do we still end up with a mechanism that can reasonably

[6] See Karmiloff-Smith (1994), Jones (2008) and Tomasello (2019). See Majeed (2020, forthcoming) for a discussion.

[7] E.g. see Mayr (1974), Griffiths (1997), Ekman and Cordaro (2011) and Scarantino (2015).

be classified as a module?[8] For present purposes, let us suppose the recalibration explanation works and in a way that conserves modularity. If that's right, *pace* de Sousa, evolved modules can promote our individual well-being. But crucially, the ways they can do so will still be constrained in various ways.

For example, the outputs of Darwinian modules are usually thought of as short-term impulses and thereby might not help with our long-term goals. Put another way, the modules need to be not only input-flexible, they need to be output-flexible as well, including in their temporal duration. This idea is nicely summed up by Solomon when he writes, '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' (2007, 19). To give this some legs, consider what it is to be deeply angry at someone. Say you are angry at the president. This could result in a tendency to have various sorts of expressive, physiological, behavioural and phenomenological responses whenever you think of him or see him on the news. But it needn't. You needn't feel a hot-flash every time you see him on TV. What's more, your anger far outlasts the sorts momentary responses we take to form the output side of an emotion module. Your anger at the president could last days, months or years. Moreover, your anger doesn't just prime you for instantaneous action, but can affect your long-term goals themselves, as well as how you might seek to satisfy them. For instance, you might be impelled to speak out against the president whenever the situation arises, go canvasing for his opponent, boycott businesses that fund his campaign etc. The existence of an input-flexible anger module might explain how you come to have certain automatic short-lived 'anger' responses towards the president, but it won't suffice to explain the way your anger towards him situates you in the world in Solomon's sense.[9]

Here it would be remiss not to mention that some proponents of emotional modularity are happy to grant that modules can be output-flexible. For example, as Scarantino observes, 'The learning history of the individual, and his or her cognitive capacities and personality traits affect both the input and

[8] Jones (2008), for example, argues that emotional recalibration is in tension with emotions being informationally encapsulated. Elsewhere I argue the existence of encapsulated modules needn't necessary rule out learning. In particular, modules might be synchronically encapsulated but still be open to diachronic updating. See Majeed (2020).

[9] Note: I think we can make this point without accepting Solomon's (1973, 1976) wider existentialist approach to emotion, which de Sousa (1987) rejects.

the output sides of the core system' (2015, 362). This means being angry needn't be associated with a distinct pattern of responses. But regardless of the patterns concerned, modules are supposed to trigger responses that are short-lived. Such situations, I think, could still be of relevance for our long-term goals. Being angry at the president, for instance, could involve being disposed to having certain short-lived anger responses whenever you see the president on the news. Such responses could have behavioural consequences. But significantly, while such dispositions can be motivating, they won't explain the myriad of ways your anger towards the president situates you in the world in Solomon's sense. In other words, they won't explain how your anger can affect your long-term goals themselves, as well as how you might seek to satisfy them. Joining the local branch of the opposition party, donating money to targeted causes, reading more political material etc., all by way of ousting the president, aren't the sorts of things that can be neatly captured by the kinds of stimulus-response patterns typical of modular systems.

The other constraint on the way evolved modules can promote our individual well-being is perhaps more obvious. If our emotion taxonomies comprise of only a limited set of modular or basic emotions, the ways our emotions can promote our well-being will be limited to these emotions. What about the emotions which seem crucial for individual well-being but don't appear in typical lists of basic emotions? What about, say, love? According to some proponents of modular emotion taxonomies, e.g., Griffiths (1997), love is a social pretence. This may be, but love, regardless of its developmental roots, clearly plays a significant role in many of our lives. The issue, of course, is partially semantic. Someone like Griffiths can grant that love is significant for our individual well-being whilst denying that love does so in virtue of being a basic emotion. But the issue is also more than just semantic, for something is lost in making this qualification. It means love doesn't get to be important for our well-being in the same ways things included in our emotion taxonomies do.

To summarise, I think these constraints go some way towards pre-empting some obvious rejoinders to de Sousa's first problem: (i) that modular taxonomies obscure the roles our emotions play with respect to our individual well-being. However, I am mindful that we have so far left unaddressed one of his main motivations for raising this problem, namely Darwinian modules are supposed to obscure the non-instrumental value of emotion. That is, emotions are supposed to matter for their own sake. For de Sousa, this has to do with their phenomenology. Certain kinds of emotional experiences can promote our well-being while others might diminish them. They do so simply in virtue of feeling a certain way, i.e., irrespective of any instrumental bearing they might also have on our choices, actions

etc. This point is familiar, but it isn't typically raised as a worry for modular accounts. Why should a commitment to emotion modules preclude emotions from having intrinsic value? The answer, I take it, lies in de Sousa's second problem: (ii) modular taxonomies misrepresent our emotional repertoire by misrepresenting the variety of emotions we can experience. If that's right, it's not that modular taxonomies undermine the intrinsic value of emotion completely, but they will be seen to leave out important ways our emotional experiences contribute to our well-being. Let us now take a closer look at this problem.

4. Modularity & Emotional Experiences

There are obvious rejoinders to de Sousa's second problem as well. According to de Sousa, simple emotion schemas are useful by way of predicting and explaining other people's behaviour, but this 'falsely instills the conviction that the representational scheme simply represents reality, and that (to put it excessively simply) the number of our emotion words is a sure guide to the number of emotions it is possible to experience' (pg. 47).

At a first glance, this objection is misguided. Although modular schemas, by necessity, abstract from detail, they are compatible with there being rich, varied emotion-tokens that fall under the broader emotion-types that form such schemas. For instance, our modular taxonomy might only recognise a general 'anger' category, but this does not rule out considerable phenomenological variability amongst instances that fall under this category. Moreover, adopting a modular taxonomy doesn't commit us to the fictitious view that we only experience one basic emotion at a time. Our emotional lives are complex, and it is common practice in psychology to recognise that we often undergo several emotions at once. Seeing your lover in the arms of someone else might induce jealousy, but it could also make you angry, anxious and sad, all at the same time. The emotions you experience at that very moment might be difficult to capture in words, however, this does not mean that what you are experiencing is anything more than a blend of various basic emotions.

I think the question of whether simple emotion taxonomies misrepresent our emotional repertoire turns on whether we miss something important when we categorise emotions along these lines. Take moral indignation. Is something lost when we classify it as a type of anger? I think it is. Moral

indignation doesn't seem reducible to anger. Someone who has felt angry but never experienced moral indignation, arguably, doesn't know what it is like to experience such indignation. The obvious retort is that moral indignation comprises of a modular response, say that which is involved in anger, in conjunction with a cognitive component, something along the lines of a judgement that something morally repugnant has taken place. To claim that such emotions involve a mix of modular responses and cognitive processes, however, is to concede the point that not all emotions can be categorised along modular emotion taxonomies. This is because in order to capture moral indignation, it turns out that we need something more than what is just there in such taxonomies; we need a specific kind of cognitive component. So although some other components of the emotion might coincide with something in our emotion taxonomies, namely anger, the emotion itself, i.e., moral indignation, will be left out of such taxonomies.[10] The other response is to insist that any features we think are lost when we classify moral indignation as a species of anger can be accounted for by the blending of various basic emotions. Maybe moral indignation involves a blend of surprise, anger and disgust.

To my mind, there are blends of basic emotions in the sense that we can experience more than one basic emotion at a given time. But whether such blends can account for the full gamut of our emotional experiences, including emotions as complex as love, jealousy and schadenfreude, is an altogether different matter. Perhaps surprisingly, some of the strongest criticism of the idea that complex emotions are mere blends of basic ones come from an advocate of modular emotion taxonomies, i.e., Griffiths (2002). While conceding that basic emotions might form some of the components of complex emotions, Griffiths is sceptical that blends of basic emotions will suffice to constitute complex emotions. He offers numerous arguments but let us just consider two by way of example.

First, the appraisal criteria for complex emotions are too sophisticated to be obtained by merely adding together the criteria for various basic emotions. In particular, emotions like jealously or moral indignation require a sensitivity to complex properties of the stimulus situation. For example, jealousy is triggered not merely by the smell of unfamiliar perfume on your lover, or the recognition that they are once again late from work, but when these features are interpreted as signs of infidelity (Jones

[10] Deonna and Teroni (2012: 24) offer a rival view. According to them, we could think of non-basic emotions (e.g., moral indignation) as basic emotions (e.g., anger) which are caused by certain kinds of judgements. However, as they themselves note, such a view is unlikely to work for all non-basic emotions.

2008). Unlike Griffiths, I don't think we should divvy up emotions into basic and complex emotion categories. For instance, anger is supposed to be a basic emotion but the appraisal criteria for anger can also be more sophisticated than Griffiths lets on. Your anger at the president, for instance, most likely involves various sophisticated cognitive appraisals of him, his character, his service, how he compares to previous presidents etc. But I think Griffiths's point stands if we think of emotions along modular vs. non-modular lines. Darwinian modules are supposed to function independent of other systems, such as those that concern cognition. Thus, the eliciting-conditions for an anger module needn't be very complex. You needn't cognitively appraise a situation to feel angry. To pick an example from Griffiths (2002), you feel instantly angry when someone sharply pokes you in the back as you wait in line to get into a nightclub. However, there are other kinds of things that we call anger which behave in a more complex manner. Your anger at the president, for example, most likely involves numerous sophisticated cognitive appraisals. The point is, such appraisals are unlikely to be obtained by simply adding together the sorts of appraisal criteria involved in eliciting modular emotional responses.[11]

One thing a sceptic might say here is that this line of criticism involves not emotions *per se* but the appraisal criteria which determine when they are triggered. This brings us to Griffith's second objection, namely complex emotions endure for a lot longer than basic ones. Once again, I think we can sidestep the basic vs. complex emotion distinction and focus instead on its modular vs. non-modular counterpart. Modular anger responses, though perhaps varied, are supposed to be short-lived. By contrast, non-modular anger, such as being angry at the president, can last days, months, even years. These sorts of non-modular responses, according to Griffiths, 'endure as real psychological processes, not mere dispositions. When a woman's feeling of guilt explains her behavior through a long session of negotiation with her husband and their lawyers, it does more than dispose her to intermittently display affect-program sadness and affect program fear.' (2002, 239). If that's right, an emotional episode like being angry at the president cannot be reduced to short-lived modular anger responses or even to dispositions to have such short-lived responses. They involve the sorts of long-standing states which help us situate ourselves in the world.

[11] This also helps bolster our previous worry. The appraisal criteria for moral indignation is distinct from those that trigger our short-lived anger responses. Therefore, insofar as such criteria matter for how we categorise emotions, moral indignation would not be easily classified as a species of (basic) anger.

It is also worth noting that though I have used familiar emotions, like jealousy and moral indignation, de Sousa's point is that there are actually vastly more emotions than for which we currently have words. His claim, recall, is that although it may be useful to employ a limited set of emotion terms by way of explaining and predicting the behaviour of others, we err when we use such terms to think about our own emotions; the emotions it is possible for us to experience. This, I think, is an intriguing possibility. Perhaps there are more emotions that those that can be captured even by our most liberal emotion taxonomies. Deonna (2011), for example, argues that 'being moved' is its own unique emotional category. De Sousa, however, goes further in supposing that there are 'literally *innumerable* emotions' (pg. 44). If that's right, modular emotion taxonomies misrepresent our emotional experiences in two ways. As we have seen, they are too simplistic; a small, finite set of evolved modules won't account for the sheer number of emotions we can experience. But such taxonomies might also be too rigid.

Darwinian modules are those that evolved because they satisfied some adaptive purpose in the EEA. This means the kinds of emotion modules that exist, and thereby the kinds of emotions that will appear in our modular emotion taxonomies, will not only be finite in scope, they will also be relatively fixed. Emotions like moral indignation, or even romantic love for that matter, are fairly recent in terms of the evolutionary history of our species. Put more precisely, they are phylogenetically novel; much too novel to be traits which evolved in accordance with the Darwinian principles of natural selection. So just as we are able to experience fundamentally different emotions to our ancestors, what's not to say that future generations will also experience fundamentally different emotions from us. If phylogenetic emotional novelty is a distinct possibility, which I think it is, simple emotion taxonomies based on Darwinian modules will be too rigid to represent them in any meaningful way.

5. Beyond Modular Taxonomies

Given everything I have said so far, it may seem as if de Sousa is against emotion taxonomies. But that's not right. For starters, he accepts that simple emotion taxonomies are useful by way of explaining and predicting the behaviour of others. As we have seen, his criticism is that such taxonomies aren't up to the task of representing the variety of emotions that we are capable of

experiencing. So how should we categorise emotions in a way that better captures our rich emotional repertoire?

De Sousa's answer is the 'Aesthetic Model', which involves drawing on mediums such as art and literature to appreciate the 'multidimensional field' of emotions which we are capable of experiencing. That art, literature, music etc. can facilitate emotional learning is a familiar idea in aesthetics.[12] Likewise, the idea of multidimensional emotional scales, though not commonplace, are to be found in psychology and cognitive neuroscience.[13] But combining these ideas, by way of creating more flexible emotion taxonomies, I think, is fairly novel. What's more, I think de Sousa does in fact succeed in pointing the way for future taxonomies; ones which can avoid the usual pitfalls of simple emotion schemas. What is less clear, however, is precisely how these more flexible taxonomies are supposed to be compatible with the evidence for emotional modularity. Let me end by explaining how we might fill this gap.

To my mind, one of the most thorough defences of the idea that engaging with literature can facilitate emotional learning, or a 'sentimental education', is to be found in Robinson's seminal work, *Deeper than Reason: Emotion and its Role in Literature, Music, and Art*. Like de Sousa, Robinson is sceptical that the power of art consists in the simple reinforcement ideas about emotion most of us know already and can be captured in our everyday emotional terminology. As she writes, novels 'introduce both characters and readers to emotional states for which there are no one-word descriptions in folk psychology' (2007, 159). Using a close reading of Edith Wharton's novel *The Reef*, Robinson argues that novels can foster a sentimental education in two ways. First, they can describe, in detail, not just the emotional states of various characters, but how these characters themselves are educated by their emotions. For instance, they draw our attention not merely to the emotional states of the characters, but also how the focus of their attention, and ultimately their points of view, are maintained and shaped by these emotions. Second, novels can also trigger various emotions in us as we read such that how we view characters and situations in these novels are also gradually shaped by these emotions. In effect, as readers, we come out not only having viewed the sentimental education of

[12] Robinson (2005), Gaut (2007) and Matravers (2001).

[13] These accounts typically focus on the affective component of emotion, e.g. see Barrett and Russell (1999), Salzman and Fusi (2010), and Anderson and Adolphs (2014).

various fictitious characters but having undergone one of our own. For Robinson, both are means of expanding our emotional horizons beyond our folk emotion categories, but why think this? In other words, while literature can tell us about (real or imagined) emotional transformations and how these in turn colour our view of the world, why think this goes anything beyond our more nuanced emotion taxonomies?

Much of the role of emotion in art, according to Robinson, involves 'non-cognitive affective appraisals': quick, automatic and pre-conscious appraisals of environmental features (both internal and external) that are relevant to the organism; appraisals that are non-cognitive because they occur prior to any cognitive evaluation. To be clear, Robinson does not think that emotions are identical to such appraisals. As she notes, 'An episode of emotion is best thought of as a process, consisting minimally in a non-cognitive affective appraisal succeeded by physiological changes of specific sorts, action tendencies, and cognitive monitoring of the non-cognitive appraisal and the other elements of the response' (157). However, as evident, she does think that they form an integral part of emotion. What's more, she thinks of these appraisals as something akin to basic emotions; 'innate' emotional responses that are 'universal' and stem from emotion-specific systems in the brain. This creates a potential problem: if basic emotions form the basis of our emotion taxonomies, it would appear that all we can learn from literature are truths about our standard emotional repertoire. In other words, it might be that we learn certain things about our emotions, e.g., we learn new things about what might elicit them, but the emotions we consider on this picture are still the very same emotions that belong to our rigid emotion taxonomies.

Robinson doesn't have much to say about emotion categorisation, but we know de Sousa's position, i.e., even if there are basic emotions-cum-Darwinian modules, we should resist our tendency to categorise emotions on their basis. In this context, the trick is to treat emotional appraisals — be they cognitive or not — as points in a multidimensional state-space. To elaborate, we can represent the affective appraisals relevant for emotion on a multidimensional model consisting of several parameters. For example, emotions could vary according to their valence (from positive to negative; pleasant to unpleasant), their intensity (from high to low), their persistence (from short to long) etc.[14]

[14] De Sousa uses Scherer's (1993, 2005) multi-dimensional appraisal theory by way of example, but here I draw on Anderson and Adolphs (2014).

Crucially, conceptualising emotions in this manner provides a sharp contrast to 'digital' models which represent emotions along a finite set of pre-defined possibilities. Moreover, it also enables us to see more clearly why engaging with art can, literally, expand our emotional horizons. Literature, for instance, is well-positioned to describe the grey areas; locations in the state-space which don't correspond neatly with any of the emotional schema we use to explain and predict the behaviours of others. Likewise, music and cinema, for example, can arguably arouse emotions outside the range of those we encounter in our day-to-day lives and thereby give us a clearer picture of the vast range of emotions it is possible for us to experience.

One final piece of the puzzle, however, remains. How is this aesthetic model supposed to accommodate the evidence for emotional modularity? According to de Sousa, thinking of emotions along a multidimensional state-space is not to deny the existence of privileged 'hot spots': locations of the state-space that attract more clusters than others. If there are such hot spots, de Sousa is willing to grant they may arise because of Darwinian modules. As he notes, 'it is entirely compatible with the evolutionary psychology hypothesis that those hot spots are actually wired in, or wired-to-be-learned' (pg. 34). I think that's right. The existence of hot spots within a multidimensional state-space allows us to resist rigid emotion taxonomies all the while accommodating any possible emotional modularity.

We can see this more clearly by looking at what constitutes our best available evidence for emotional modularity, *viz.* recalcitrant emotions: emotions that are in tension with our considered judgements. I am of the view that such emotions can only provide us with defeasible evidence for emotional modularity.[15] However, even if they provide us with robust evidence for modularity, the existence of modular recalcitrant episodes is something the aesthetic model can readily accommodate. Suppose some recalcitrant emotions are innate or 'pancultural', i.e., they can be found in most cultures. This is not entirely farfetched. An irrational fear of heights or disgust at drinking from a sterilised cup that once housed a cockroach might stem from responses that were adaptive in the EEA. If these sorts of emotional responses are really innate, they will most likely cluster at various regions in our affective

[15] Recalcitrant emotions are testimony to the fact that our emotions can function independent of conscious thought, however, they don't, in themselves, show that they function independent of all cognition. E.g. your fear of Fido might run counter to your considered judgement that he is harmless, but it might be grounded in a more deeply held, perhaps unconscious, conviction that even docile dogs can be unpredictable and potentially dangerous.

state-space. But crucially, such clustering won't preclude the possibility that there are instances of fear and disgust that fall outside these hot spots. For instance, someone's fear of the long-term consequences of Brexit is unlikely to be found in the same region of the state-space as the more pancultural recalcitrant fear of heights. (The physiological and phenomenological aspects of these two kinds of fear are very different, which suggests a difference in their affective profiles). It would appear, then, that our aesthetic model can show us how to have it both ways; it can represent a limited set of modular emotional responses without representing our entire emotional repertoire as being limited in this manner.

6. Conclusion

In this paper I have defended de Sousa's charges against modular emotion taxonomies. Such taxonomies obscure the roles our emotions play with respect to our individual well-being. For instance, they ignore some aspects of the intrinsic value of emotion, as well as how our emotions can contribute to our long-term goals. Moreover, such taxonomies also misrepresent our emotional repertoire because they ignore the variety of emotions we can experience. We can see this clearly by looking at how we respond emotionally to art and music, as well as what we can learn about emotions secondhand through literature, poetry, cinema etc. In relation to this, I have also defended de Sousa's alternative to such taxonomies, *viz.,* the Aesthetic Model. I argued that such a module can represent the existence of modular emotions in the form of certain clustering in a multidimensional state-space. But crucially, it can do so without presupposing that all emotions will be represented in such clusters. To categorise emotions solely based on modules, in effect, is to suppose that regions in the state-space that don't harbour 'hot spots' remain empty. Music, art, literature, film and poetry, according to de Sousa, show us that this is simply not the case. If he's right, our emotions, contrary to modular emotion taxonomies, contain multitudes.[16]

[16] Many thanks to Julien Deonna for helpful comments on the draft.

References

Anderson, David. J and Ralph Adolphs. 2014. "A framework for studying emotions across species". Cell 157(1):187-200. doi: 10.1016/j.cell.2014.03.003.

Barrett, L.F and James A. Russell. 1999. "The structure of current affect: controversies and emerging consensus." Curr. Dir. Psychol. Sci. 8: 10–15.

Benbaji, Hagit. 2013. "How is recalcitrant emotion possible?" *Australasian Journal of Philosophy* 91: 1–23.

Charland, Louis. 1995. "Emotion as a natural kind: Towards a computational foundation for emotion theory". *Philosophical Psychology* 8: 59–84.

Cosmides, Leda and John Tooby. 2000. "Evolutionary Psychology and the Emotions". In Michael Lewis and Jeannette Haviland-Jones (eds.), *Handbook of Emotions* (2nd ed.), 91-115. New York: The Guilford Press.

D'Arms, Justin and Daniel Jacobson. 2003. "The significance of recalcitrant emotion". In Anthony Hatzimoysis (ed.), *Philosophy and the Emotions*, 127–145. Cambridge: Cambridge University Press.

Deonna, Julien. 2011. "Etre ému [Being moved]". In Christine Tappolet, Fabrice Teroni, and Anita Konzelmann Ziv (eds.), *Les ombres de l'âme: Penser les émotions négatives*. Geneva: Markus Haller.

de Sousa, Ronald B. 1987. *The Rationality of Emotion*. Cambridge and London: MIT Press.

de Sousa, Ronald B. 2008. "Against Emotional Modularity". In Luc Faucher and Christine Tappolet (eds.), *The Modularity of Emotions*, 29-50. Alberta: University of Calgary Press.

Deonna, Julien and Fabrice Teroni. 2012. *The Emotions. A Philosophical Introduction*. London/New York: Routledge.

Ekman, Paul. 2003. *Emotions Revealed: Understanding Faces and Feelings to Improve Communication and Emotional Life*. New York: Henry Holt and Company.

Ekman, Paul. 1999. “Basic Emotions”. In Tim Dalgleish and M. Power (eds.), *Handbook of Cognition and Emotion*, 45-60. Chichester: John Wiley and Sons, Co.

Ekman, Paul., and Daniel Cordaro. 2011. “What is meant by calling emotions basic?” *Emotion* Review, 3(4): 364–370.

Faucher, Luc and Christine Tappolet. 2008. “Introduction: Modularity in the Nature of Emotions”. In Luc Faucher and Christine Tappolet (eds.), *The Modularity of Emotions*, vii-xxxi. Alberta: University of Calgary Press.

Fodor, Jerry A. 1983. *The Modularity of Mind*. Cambridge and London: MIT Press.

Frijda, Nico H. 1986. *The Emotions: Studies in Emotion and Social Interaction*. Cambridge, Paris: Cambridge University Press, editions de la maison des sciences de l’homme.

Gaut, Berys 2007. *Art, Emotion and Ethics*. Oxford: Oxford University Press

Goldie, Peter. 2000. *The Emotions. A Philosophical Exploration*. Oxford: Oxford University Press.

Greenspan, Patricia S. 1981. “Emotions as Evaluations”. *Pacific Philosophical Quarterly* 62*:* 158–169.

Griffiths, Paul Edmund 1990. “Modularity, and the psychoevolutionary theory of emotion”. *Biology and Philosophy* 5(2): 175-196.

Griffiths, Paul Edmund. 1997. *What Emotions Really Are*. Chicago and London: Chicago University Press.

Griffiths, Paul Edmund. 2002. “Is Emotion a Natural Kind?” In Robert C. Solomon (ed.), *Thinking About Feeling: Contemporary Philosophers on Emotions*, 223-249. New York: Oxford University Press.

Griffiths, Paul Edmund. 2004. “Emotions as natural and normative kinds”. *Philosophy of Science* 71(5): 901-911.

Grzankowski, Aex. 2016. “The Real Trouble with Recalcitrant Emotions”. *Erkenntnis* 82 (3):641-651.

Izard, C. E. 1992. “Basic Emotions, Relations Amongst Emotions and Emotion–Cognition Relations”. *Psychological Review* 99: 561–565.

Jones, Karen. 2008. “Quick and Smart? Modularity and the Pro-Emotion Consensus". In Luc Faucher and Christine Tappolet (eds.), *The Modularity of Emotions*, 3-27. Alberta: University of Calgary Press.

Karmiloff-Smitch, Annette. 1994. “Précis of Beyond Modularity: A Developmental Perspective On Cognitive Science”. *Behavioral and Brain Sciences* 17(4):693-707.

LeDoux, Joseph. 1996. *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Simon & Schuster.

Matravers, Derek. 1998. *Art and Emotion*. Oxford: Oxford University Press.

Majeed, Raamy. 2019. “What can information encapsulation tell us about emotional rationality?” In L. Candiotto (ed.), *The Value of Emotions for Knowledge*, 51–69. Basingstoke: Palgrave Macmillan.

Majeed, Raamy. 2020. “Does modularity undermine the pro-emotion consensus?” *Mind & Language* 35: 277– 292. https://doi.org/10.1111/mila.12241

Majeed, Raamy. (forthcoming). “The “Puzzle” of Emotional Plasticity”. *Philosophical Psychology.*

Mayr, Ernst. 1974. “Behavior programs and evolutionary strategies”. *American Scientist* 62: 650–659.

Panksepp, Jaak. 1998. *Affective Neuroscience: The Foundations of Human and Animal Emotions.* New York: Oxford University Press.

Panskepp, Jaak., and Watt, D. 2011. “What is basic about basic emotions?: Lasting lessons from affective neuroscience”. *Emotion Review* 3(4): 387–396.

Pinker, Steven. 1997. *The Way the Mind Works*. London: Penguin Press.

Prinz, Jesse. 2004. *Gut Reactions: A Perceptual Theory of Emotion.* Oxford: Oxford University Press.

Prinz, Jesse. 2008. “Is Emotion a Form of Perception?” In Luc Faucher and Christine Tappolet (eds.), *The Modularity of Emotions*, 137-160. Alberta: University of Calgary Press.

Robinson, Jenefer. 2005. *Deeper Than Reason: Emotion and its Role in Literature, Music, and Art*. Clarendon Press.

Salzman, C.D., and Stefano Fusi. 2010. “Emotion, cognition, and mental state representation in amygdala and prefrontal cortex”. Annu. Rev. Neurosci. 33: 173–202.

Scarantino, Andrea. 2015. “Basic Emotions, Psychological Construction and the Problem of Variability”. In James Russell and Lisa Feldman Barrett (eds.), *The Psychological Construction of Emotion*, 334-376. New York and London: Guilford Press.

Scarantino, Andrea. 2018. “Are LeDoux’s Survival Circuits Basic Emotions Under a Different Name?”*Current Opinion in Behavioral Sciences* 24: 75-82.

Scarantino, Andrea, and Paul Edmund Griffiths. 2011. “Don’t give up on basic emotions”. *Emotion Review,* 3(4): 444–454.

Scherer, Klaus R. 1993. “Studying emotion-antecedent appraisal process: An expert system approach”. *Cognition and Emotion* 7(3-4): 325-55.

Scherer, Klaus R. 2005. “What are emotions? And how can they be measured?” *Social Science Information* 44(4): 695-729.

Seligman, M. E. P.: 1970. “On the Generality of the Laws of Learning”. *Psychological Review* 77: 407-418.

Seligman, M. E. P. 1971. “Phobias and Preparedness”. *BehaviourTherapy* 2: 307-320.

Sperber, Dan 2002. “In Defense of Massive Modularity”. In *Language, Brain and Cognitive Development: Essays in Honor of Jacques Melher*, ed. E. Dupoux, 47-57. Cambridge: Cambridge University Press.

Solomon, Robert C. 1973. “Emotions and Choice”. *Review of Metaphysics* 27 (1):20 - 41.

Solomon, Robert C. 1976. *The Passions. The Myth and Nature of Human Emotions*. Doubleday.

Solomon, Robert C. 2007. *True to Our Feelings: What Our Emotions Are Really Telling Us.* Oxford University Press.

Tappolet, Christine. 2016. *Emotions, Value, and Agency*. Oxford: Oxford University Press.

Tomasello, Michael. 2019. *Becoming Human: A Theory of Ontogeny*. Cambridge and London: Harvard University Press.

Tomkins, Silvian S. 1962. *Affect, Imagery and Consciousness*. New York: Springer.

Tooby, John and Leda Cosmides. 1990. "The Past Explains the Present: Emotional Adaptations and the Structure of Ancestral Environments". *Ethology and Sociobiology* 11: 375–424.

West-Eberhard, M. J. 2003. *Developmental Plasticity and Evolution*. Oxford: Oxford University Press.

RAAMY MAJEED (raamy.majeed@manchester.ac.uk) is Lecturer in Philosophy at the University of Manchester. He works primarily in the philosophy of mind and cognitive science. His recent publications include 'The "Puzzle" of Emotional Plasticity' (*Philosophical Psychology*, forthcoming), 'On Biologising Racism' (*British Journal for the Philosophy of Science*, forthcoming), 'The New LeDoux: Survival Circuits and the Surplus Meaning of 'fear''(*Philosophical Quarterly*, 2020) and 'Does Modularity Undermine the Pro-Emotion Consensus?' (*Mind and Language*, 2020).