emotion review

Emotion Review
1–11
© The Author(s) 2018
ISSN 1754-0739
DOI: 10.1177/1754073918765662
journals.sagepub.com/home/er

Investigating Emotions as Functional States Distinct From Feelings

Ralph Adolphs
Division of Humanities and Social Sciences and Division of Biology, California Institute of Technology, USA

Daniel Andler
Department of Philosophy, Université Paris-Sorbonne, France
Department of Cognitive Studies, Ecole normale supérieure, PSL Research University, France

Abstract

We defend a functionalist approach to emotion that begins by focusing on emotions as central states with causal connections to behavior and to other cognitive states. The approach brackets the conscious experience of emotion, lists plausible features that emotions exhibit, and argues that alternative schemes (e.g., focusing on feelings or on neurobiology as the starting point) are unpromising candidates. We conclude with the benefits of our approach: one can study emotions in animals; one can look in the brain for the implementation of specific features; and one ends up with an architecture of the mind in which emotions are fully accommodated through their relations to the rest of cognition. Our article focuses on arguing for this general approach; as such, it is an essay in the philosophy of emotion rather than in the psychology or neuroscience of emotion.

Keywords
emotion, functionalism, feelings, animal emotions

Introduction

All empirical and theoretical approaches to the study of emotion face a fundamental challenge: where and how does one begin in deciding the subject matter? What counts as relevant, and as part of the phenomenon to be explained, and what should be bracketed or omitted? Our position begins with a subset of all the different phenomena related to emotions that people often talk about; we do not attempt to tackle all of them. At least at the outset, we thus offer what amounts to a promissory note. We are wagering that, if you stick with our approach, you will at the end of the day have a scientifically useful theory of emotions. It may not address all aspects of emotion that one would eventually want to understand, but it addresses enough of them that we can all agree that the wager paid off: that we have a scientific theory that is indeed about emotions (as opposed to having changed the topic to something else), and that accounts for a substantial portion of our commonsense view of emotion. This at least is our conviction.

Our wager can be introduced as follows. Suppose an alien spaceship lands on earth. The aliens, intelligent and skilled in doing science as they are, have none of the prior concepts about the mind that we use, notably also lacking any concept of conscious experience. What science would they now construct in order to explain the complex behaviors of people and animals that they find on earth? We posit that they would come up with two broad classes of explanations: those that do not require latent variables (akin to our concepts of mental states) and those that do. The former would explain many kinds of movements, like reflexes. The latter would explain complex behavior. Among the latter would be a collection of internal states (which could be partitioned in ways rather different from our current science) that are distinguished by particular features or properties that the aliens regularly observe. Whereas we have attention and memory and fear and sadness, the aliens might employ other concepts to do the job. One class of these internal states, a

Author note: The authors are indebted to the French Society for Philosophy of Science, the French Neuroscience Society, and Thomas Pradeu for first bringing us together, and to Andrea Scarantino for helpful detailed critiques of this article. Funded in part by the Caltech Conte Center for the Neurobiology of Social Decision-Making (NIMH) and by the Carver Mead New Adventures Fund (Caltech).

Corresponding author: Ralph Adolphs, Division of Humanities and Social Sciences and Division of Biology, California Institute of Technology, HSS 228-77, Caltech, Pasadena, CA 91125,USA.
Email: radolphs@hss.caltech.edu

class prominently continuous across all animal species including humans, would have a particular set of features or properties (which we elaborate on further in what follows; see Figure 3), and these would roughly correspond to emotions. The moral of this story is that, in our view, emotions emerge as a particular class of latent variables that are required by any science of the complex behaviors observed across higher animals. The particular categories or dimensions by which to distinguish different emotions might vary (indeed, we think they likely will, as we suggest further below), but the generic category would be a requirement of any science of the mind, however mental states are conceived.

To look ahead to the rest of this article, there are three key points: (a) a science of emotion is constructed from observable data (whatever the aliens could measure), but not, at least in the first instance, by how emotions feel (the conscious experience of having one); (b) emotions are internal states (latent variables) of the same broad type as cognitions (i.e., functionally individuated states); (c) but emotions are distinguished from the rest of cognition by specific operating characteristics or features (much the same as attention, memory, and other cognitive state types are distinguished from one another).

As the starting point for a science of emotion, then, we are imagining all the same phenomena that we observe in other people and animals all the time. We are simply asking what kinds of explanations one might come up with if one did not have specific prior beliefs about emotions. We are preserving all of the objective phenomena to be explained (and, again, bracketing conscious experience), and in particular all of the behavioral and physiological observable events that we would normally use to infer emotions: facial expressions, overt motor behaviors, autonomic and endocrine activity, and so forth. In this picture, we find it plausible that emotions would emerge as a class of internal states that provide causal explanations of behaviors at a particular level of control and complexity: a level intermediate to that of reflexes and deliberated behavior. This level of control, in turn, is distinguished by a set of operating features that we can begin to list (and that the aliens would similarly list). Further below, we provide such a preliminary list (Figure 3) and suggest that it can be used to infer computations that characterize emotions.

In what follows we elaborate this view and consider some of its implications. However, we emphatically do not offer a *theory* of emotion in any sense. We are here interested not in arguing for *specific* functional roles that emotions subserve (let alone that specific emotions like fear or anger might subserve), but rather in arguing that a functional account is a promising *type* of account for a science of emotion in general (as an approach). We thus devote most of our article to laying this groundwork, suggesting some specifics only in the penultimate section.

Bracketing the Conscious Experience of Emotions

Our approach is of course a rather bold one, since most people, and many psychological theories of emotion, explicitly (or even exclusively) focus on conscious experiences of emotions. But as we suggest in the next section, there are good reasons to think that feelings cannot provide a principled distinction that carves out emotions from other kinds of conscious content; and feelings create other problems as well, such as methodological ones in studying emotion in animals (or constructing them in robots).

Bracketing feelings means just that: for now, we leave them aside, as a practical matter in order to make scientific progress. We are thus proposing that there exists a level of basic mechanism that can be used to understand emotions without recourse to the feelings caused, or otherwise involved, in emotion states. We think it likely, in fact, that feelings are an important part of the overall picture, a part that eventually needs to be explained. But this does not mean that they need to figure at this stage in a science of emotion. It is even possible that a scientific theory of emotions that begins by bracketing feelings might eventually provide a new explanation of feelings. At a minimum, leaving aside conscious experience would surely seem to make any scientific investigation a more tractable task than including it: there is no agreement on any approach to understanding consciousness, and there remain serious proposals claiming that it cannot be understood at all. Leaving an obviously difficult aspect of emotion aside—again, for the time being and as a practical matter—thus seems a wise strategic choice. It should be noted that many emotion theories in fact do distinguish emotions and feelings (e.g., Damasio, 2003; LeDoux & Brown, 2017), although in general this is because they argue the two are different things. We are not advancing any argument about the possible relation between emotions and feelings—we are simply omitting feelings as one aspect for purely methodological reasons.

Bracketing conscious experience is in fact a common strategy in any project that takes mental processes and attempts a functional, mechanistic, or specifically neuroscientific explanation of them. We take this approach all the time; for instance, in vision science as in memory research (Adolphs, 2017). Visual processes in both humans and animals have usually been studied without reference to conscious visual experience (yet allowing, in principle, for an enrichment that would account for it). Why not do the same thing for emotion?

It might be argued, however, that conscious experience is more easily detached from the central function of vision or memory than from emotion. When we consider conscious visual or mnemonic experience, we are (so the argument might run) generally concerned with what Ned Block (1995) has called “access consciousness”: awareness of seeing or remembering X, and the ability to report this and to make use of it in deliberation. We are often not concerned with the qualitative feel of seeing or remembering X in these cases (what Block calls “phenomenal consciousness,” and what is often referred to as “qualia”). By contrast, when X is the object of an emotion, it seems that the qualitative “feel” that goes with emoting about X seems to take precedence.

Perhaps relatedly, it might also be argued that the aspect of consciousness that is most prominent in the case of vision and memory is a higher order awareness that one is having a visual or mnemonic experience. And again, in the case of emotion experiences, the "raw feel" of the emotion seems to take precedence over the awareness that one is emoting about X (although Joe LeDoux has recently suggested otherwise; LeDoux & Brown, 2017). Moreover, the fact that different emotions seem to carry a distinctive "feel" can be thought to be what differentiates emotional episodes from other kinds of mental episodes, and to differentiate categories of emotion, such as fear, anger, happiness, and so forth, from one another (a view whose utility we argue against further in what follows).

For now, these purported disanalogies between the prominence of conscious experience in the case of emotion, on the one hand, and in the case of vision or memory, on the other, may weaken the appeal to the success of vision and memory research, but do not directly argue against our approach. The mere fact that people tend to include qualitative conscious experiences of emotion in their concept of emotion is certainly no argument. As we will argue further in the following lines, there is every reason to think that conscious experiences of feelings cannot provide a scientific basis for distinguishing emotions from other mental states, or even different types of emotion from one another.

What Are the Alternatives?

It may help to locate our approach in relation to other views. We do not undertake a detailed review of emotion theories, but sketch the most important alternatives to our view, and the difficulties that we believe they face. Andrea Scarantino has surveyed philosophical theories of emotion as falling largely into three approaches that take different aspects of an emotion as foundational or primary in some sense (Scarantino, 2016). These three are appraisal, feeling, and motivational theories, which, respectively, emphasize the evaluative, experiential, and behavioral components of an emotion. We comment further on appraisal and motivational approaches next, which we find congenial but pegged at a different level of explanation than the present article (see From Emotion Features to Functional Role section). So far, we have argued for bracketing the conscious experience of emotions, but one might well wonder how the story would look if instead of bracketing feelings, one took them to be foundational.

Feelings are taken as the starting point for the layperson's concept of emotion, and also in many theories of emotion. But the attempt to classify emotions in virtue of how they feel immediately brings with it a host of problems about how to measure or characterize feelings. The bulk of the difficulties in finding any kind of consistency between emotional feelings and either physiological patterns or neural activation patterns is focused on the difficulty of distinguishing among different emotional feelings—categorizing happiness, sadness, fear, anger, and so forth as distinct categories (for reviews regarding neural activations and physiological patterns, respectively, see Lindquist, Wager, Kober, Bliss-Moreau, & Feldman Barrett, 2012; Siegel et al., 2018). But there are problems even at a coarser grain: it's also unclear how to distinguish emotions, generically, from other states. Pain and hunger, for instance, also seem to be characterized by feelings that share a lot in common with emotions: they are valenced and they motivate behavior. It is very unclear how to find a principled difference that would distinguish emotions from these other states by feeling only. In the absence of a principled distinction in feelings between these states and emotions, it becomes ad hoc to say they feel different. Indeed, the problem is deeper than that: what is it about emotional feelings that makes them *emotions*, as opposed to some other internal states? Saying that these feelings are accompanied by other phenomena, which in turn serve to individuate them as emotions, of course begs the question since, then, it is these other phenomena or criteria that are doing the work. The burden is on the feelings-first theorist to provide a positive account of some sort.

There is another obvious problem in taking feelings as foundational as well, methodological in nature and with severe consequences for a science of emotion: focusing on feelings makes it currently impossible to study the topic in nonverbal subjects. Infants and aphasic subjects, as well as of course all nonhuman animals, become impossible to study for a science of emotion that puts feelings first because nobody can agree on a reliable dependent measure for feelings other than verbal reports (especially in the case of more finely differentiated feelings that could serve to classify emotions). Worse than that, it becomes quite problematic how to compare measures across people who speak different languages that use different words to describe their feelings.

Indeed, we believe that if feelings are chosen as the place to start, we will be left with a hopelessly anthropocentric, culturally relativistic exercise. Even if we choose for these reasons to constrain our science of emotion just to college-educated, Western, English-speaking human adults (as much of psychology has in fact done), we end up, as a starting point, by choosing to solve what is arguably the most difficult problem that a science of the mind faces: we have chosen to aim for a theory of conscious experience. Why start there? It seems like the last place to get any traction, and the first place to run into large conceptual problems right at the start.

There is another interesting modern view, often curiously intertwined with a feelings-first emphasis: the view that neuroscience can provide the needed conceptual purchase to individuate emotions. This road is of course inspired by the success of neuroscience, and its power to disambiguate emotion theories, two facts we wholeheartedly acknowledge. For instance, neurobiologists like Jaak Panksepp link emotions in a strong sense to the operation (and phylogenetic conservation) of specific brain systems (primarily subcortical systems in his scheme, which implement his list of specific basic emotions; Panksepp, 1998). There may indeed be strong continuity in the neural systems for emotions across species if we stick to mammals, but this approach is unlikely to make sense as foundational if we wish to understand the emotions of octopuses, robots, or aliens.

A different starting point builds on the idea that our current emotion concepts are hopelessly confused and we need a fresh start. Why not look just at data from the brain, and use a more data-driven approach from neuroscience to bootstrap a theory of emotion? Insofar as most neuroscientists do not offer clarification on what they mean by the term “emotion,” they are probably guilty of fueling this approach as well. Perhaps the clearest statement of it from an emotion theorist can be found in Lisa Feldman Barrett’s work. On our reading of Barrett, a psychological theory of emotion should fall out of attention to the neuroscience facts and her recent book (Barrett, 2017) accordingly spends most of its pages detailing anatomical and physiological findings from neuroscience. She is certainly not alone in an emphasis on neuroscience, but we feel that there are a number of problems with this approach as foundational.

First, one might naturally question whether we know enough about the brain to do this. Currently, our knowledge of any brain, and certainly of the human brain, is woefully slim. But even if we knew more about the brain, as surely we eventually will, it remains quite unclear how and whether we could derive any facts about the mind, or about categories of behavior, that would in any way allow us to recover a notion of “emotions” (or, for that matter, any other cognitive states) from neuroscience data alone. The reasons are well known: knowledge of internal processes alone cannot allow us to infer what they evolved for and what they can do for an organism. It is as impossible to figure out the function of invertebrate central pattern generators merely from recording their neurons as it is to figure out a microprocessor that was designed for playing computer games merely from measuring its circuits (Jonas & Kording, 2017). Without some anchor in the world (and, we would argue, in the natural history of interactions with the world), it is either very difficult or outright impossible to make sense of what the brain does (but see below).

Second, there is the old conceptual problem of reference or intentionality. If we restrict ourselves to data from the brain alone, there is no grounding that could link the brain’s internal processing to the world. What makes the brain an *information processing* system cannot be understood without stepping outside the brain in order to see what the information is about. We need to know what stimuli trigger changes in the brain, and in what way behavior is more than mere motion. Ethological data that observe animals and humans in their natural environment are essential not merely for methodological reasons (as one might think of them in providing constraints for the previous paragraph), but serve a deeper need to make representation possible at all.

It should be emphasized that we are of course not saying that neuroscience is irrelevant, or that emotions do not fully supervene on the neurobiological details. But we just have no way to describe emotions as mental states unless we can refer to their function in relation to causes and effects in the environment; indeed, to the natural history of such causes and effects. Our aliens might very quickly acknowledge that all the behaviors that they observe are proximally caused by events in the nervous systems of animals; but we would argue that they would not begin by studying the brain and neglecting ethology. Philosophers have come up with various thought experiments to drive home this point. A simple one from Donald Davidson suffices here (Davidson, 1987). Consider sunburn, which is a state of my skin. The dermatologist may well be incapable of distinguishing true sunburn (caused by the sun) from an identical skin condition caused by a chemical burn. But one is sunburn, and the second is not. The fact that the sunburn is identified as the state that it is (sunburn and not chemical burn) doesn’t make it a different physiological state of the skin, but without this external reference we are unable to individuate it as sunburn—we cannot do it only by looking at a piece of skin. So, in our view, it is perfectly compatible to believe in the functionalism that we will detail further in the next section, yet also to maintain that emotions are just states of the brain. They are individuated by their function (which, in the broad external sense generally requires knowledge of their natural or engineered history) but identical with a particular physiological state of the brain.

Returning to Barrett’s view, we briefly comment further on three of its core features: the primacy of neuroscience data; the confusion of current emotion concepts and terms; and the flexible construction of emotions from other ingredients. We discussed before our reasons for not believing that neuroscience data can be foundational in and of itself. Yet there is one possibility that should be mentioned. It is possible that eventually we know so much about cognitive neuroscience that finding activation of particular brain systems when people are in emotion states may tell us about the functional role of such states because we know so much about the corresponding brain systems. This still raises the question of how to delineate when somebody is in an emotion state in the first place, but one could imagine a very rough characterization, or indeed one based on self-report, that is subsequently refined by the neuroscience data. It is possible that Barrett has something like this in mind in sketching brain systems out of which emotions are constructed: if we know enough about the other functions of such brain systems (appraisal, attention, motivation, etc.) we might get insight into the functions of emotion as well (Barrett & Satpute, 2013). It is possible that other thinkers, such as Luiz Pessoa, have in mind something along these lines as well in their emphasis on brain networks that mediate interactions between emotions and cognition (Pessoa, 2013, XXXX). At this stage, it is unclear to us how useful such a strategy is—we would generally feel that we still know far too little about the brain to use such data in formulating a theory of emotion. It should be the other way around: the theory of emotion should make testable hypotheses, which can then be investigated with neuroscience.

Turning to a second core feature of Barrett’s view, we actually agree with Barrett and others that current emotion concepts, and in particular words and concepts for specific emotions (fear, anger, and so forth) are highly problematic. One good indication of the problem is that most emotion theorists have their own list of emotions. Ekman has his list of basic emotions; Panksepp has his; and there are other accounts as well (Tracy & Randles, 2011). We believe that we do not yet know how to classify emotions into specific emotions, or whether that will even turn out to be the most useful way to characterize their variability (perhaps dimensional approaches will save the day).

Yet our view emphatically disagrees with the third, and perhaps most fundamental, of Barrett’s views: that emotions are constructed. By this she means not only that they are compositional in some sense (a view with which we agree if it just means that their mechanisms can be investigated), but that the construction is highly variable and individual. Indeed, the view that emotions are constructed seems to be Barrett's main reason for arguing that current words and concepts for them are problematic: if they can be constructed so flexibly that anybody can create a new emotion, then views for a scientific taxonomy seem doomed. We believe that there is such a scientific taxonomy (although it may offer dimensions rather than categories), but we also believe that we have not yet found it, and that the current schemes advocated are unlikely to be good candidates. While Barrett believes that we know so much about the neural basis of emotion that we can see that emotions are flexibly constructed in each case, we believe that we know so little about emotions that current schemes are likely wrong—but that there will be a future scheme that science can reveal.

To summarize this section: very broadly, we see three approaches to providing foundations for emotion. The most obvious one is to anchor emotions in conscious experience; we have chosen instead to bracket feelings (again, not to eliminate them or necessarily exclude them forever). Another currently popular approach is to hope that neuroscience can give us all the answers, especially if we also believe that current concepts are infused with a hopeless folk psychology and just need to be eliminated. We have argued that this is just impossible—but we are in agreement that our current terminology will likely require substantial revision. Instead of looking to neuroscience as the foundation, we advocate looking to functionalism, and subsequently using neuroscience data to test more refined functional hypotheses. Finally, a third broad approach indeed uses functional criteria: appraisal theories and motivational theories both fall into this category. However, these are functional role theories of emotion that characterize emotions as having specific functional features (and, in general, do so for specific emotions like fear, anger, and so forth). We are not proposing any such theory here, although towards the end of this article we will suggest an approach for how to further develop a specific functional-role theory for emotions. There we will argue that one could do so by listing reasonable properties of emotions (including aspects of appraisal and motivation; see Figure 3 for our provisional list). But we will also argue that it is likely premature to commit to a strong and specific theory that chooses only one or a small number of emotion properties as foundational (e.g., appraisal, motivation, valence). We think that there will be numerous features that emotions exhibit by which they achieve their functions (most or all of the ones listed in Figure 3) and the challenge will be how to combine them into a computational architecture. Again, we feel this is a large and difficult project for the future, and it is not our current project here.

Functionalism and Emotions

Our current project, then, is to provide a broadly foundational approach that is methodological. *How* would you go about creating a theory of emotion, how would you interpret empirical observations? (How would our aliens do it, observing behaviors amongst species on earth?). We see emotions as functional mental states in the sense intended by classical philosophy of mind: as relational, that is, individuated by the web of causal relations they have to all other internal states, as well as to stimuli and behaviors. We thus shun any views that try to locate emotions essentially as physical properties at a particular level ("essentialism," a term we avoid here since it can be confusing). In a nutshell: emotions are individuated by what they do, not by how they are physically constituted.

We also refrain from any metaphysical claims about what emotions "really are": we are not committed to claiming that they *really are* nothing but place holders in the causal web. We are not defining or characterizing emotions (again, we are not providing a theory of emotion), but describing an approach to investigating them and theorizing about them. Our stance is methodological, sparing us the need to counter the objections that have been raised against psycho-functionalism as a metaphysical thesis. Thus, for example, arguments purporting to show that a functionalist theory of mental states cannot for metaphysical reasons account for the "feel" that some of those states (such as emotions) carry, do not compromise our position. There may be features of emotions that don't directly fall out of their functional role; we think it wise at this point to refrain from taking a bet one way or another. On the other hand, we certainly allow for the possibility that the same functionally defined emotion state could occur in animals or robots with very different hardware, as long as the structure of cause–effect relationships that defines the emotion state is preserved.

This classic view has an immediate and attractive consequence for the scientist: it tells us how to infer an emotion, namely, from observing its causes and effects. Thus, physiological dependent measures, including neurophysiological ones, as well as facial expressions and overt behaviors can all serve as evidence on the basis of which an emotion state can be inferred. Similarly, particular circumstances and environments can all serve as part of the evidence on the basis of which an emotion state can be inferred. And this is exactly what we do all the time in everyday life: we regularly infer emotions in other people and in animals from their observed behavior, or from observing the situations in which they find themselves, or both. This skeleton schematic is shown in Figure 1. Needless to say, the inference of emotions from observable measures is shared with many emotion theories, even though they may not be functionalist in nature.

Of course the story is not always so simple, and there are at least three complicating factors. First, the behaviors caused by an emotion can, at the next time step, become stimuli in their own right; so there are feedback loops to incorporate into the picture—emotions are temporally extended processes. Second, there is substantial regulation, at least in adult humans, that complicates how emotions are caused by stimuli. (It is a large and interesting further question to characterize how such regulation might also be found in other animals, such as trained dogs, and how it might differ from the volitional regulation of emotion seen in adult humans.) Third, the links from stimuli to emotions, and from emotions to behaviors, are extremely diverse: many kinds of stimuli can cause a certain type of emotion state, and a given type of emotion can cause many different behaviors. This

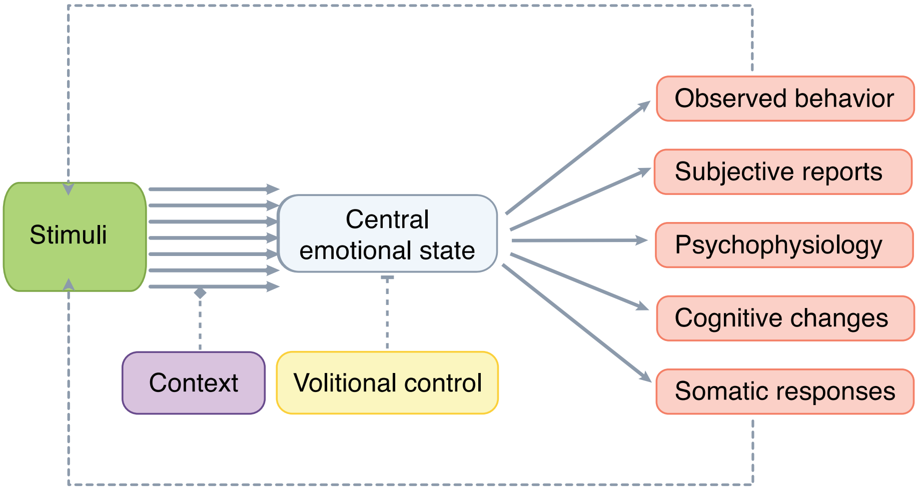


Figure 1. Skeleton schematic for a functional role of emotion states. The figure underemphasizes the full functional role of emotions, which would include causal connections (both as inputs and outputs) not only to many other cognitive states, but also to other emotion states.
Note. Modified from (Adolphs & Anderson, 2018).

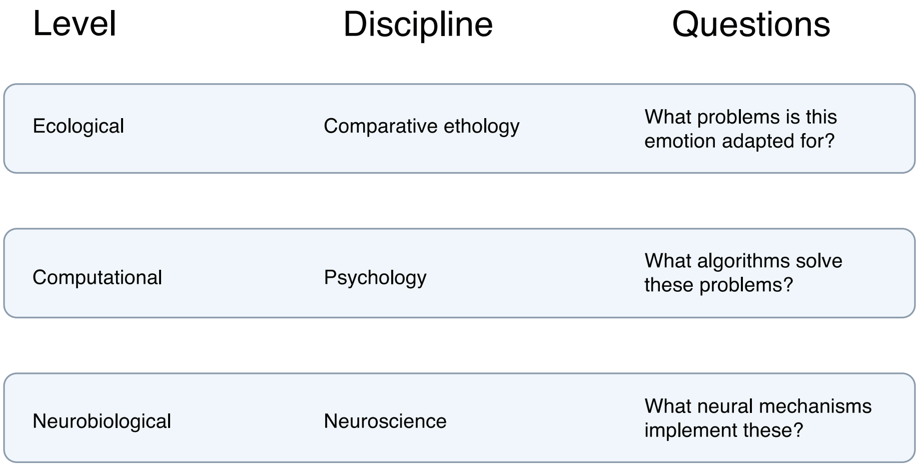
"fan-in, fan-out" architecture is exactly the reason why behaviorism was unable to solve stimulus–response relations, and exactly why science (and evolution) needed internal states such as emotions that could serve to bind together a large range of different stimuli and a large range of different behaviors.

Psycho-functionalism,[1] then, provides a place to start, but it does not yet say anything useful about how exactly to distinguish emotions from any other cognitive state, nor how to distinguish different types of emotions. So the next question is: *what* is the functional role of emotions, and *what* is the functional role of, say, fear, that makes it a different emotion than, say, disgust? We do not answer these questions here but sketch how one might go about answering them. One would need to focus on the function within the general scheme of the creature's life and ecology. In other words, we need to understand the *biological significance* of the emotion viewed as a set of relations in the web of mental states, stimuli, and behavior. Our usage of the term "biological significance" here is also functional, but in a different sense than we have been using "functional" so far. Biological significance requires a functional account that is broad and external: what functional role does the emotion, operating in an organism, subserve in order to produce behavior that is adapted to that organism's environment? Evolutionary psychology has attempted some such characterizations, but these are difficult questions for which we desperately need more empirical data. As so often is the case in biology, the analysis of proximal causes, in Ernst Mayr's terminology (Mayr, 1961), needs to be complemented by the consideration of distal or ultimate causes, those of evolutionary history (these are also what could explain mistakes or errors in how an emotion functions in any given instance). Whatever functions it is that emotions carry out in this broad and external sense, they likely evolved, they show considerable phylogenetic continuity, and they were mostly adaptive in the ancestral environment in which they evolved.

There has to be a story to tell about how simpler organisms like worms and flies first evolved states that might be considered the precursors to emotions, and how these evolved into the emotion states that we see, with remarkable homology, across many mammals and other higher animals today. A functional role from the perspective of evolution, then, is what careful comparative and ethological work would bring to the table. Using such data to construct psychological theories is extremely difficult, a hurdle with which the entire field of evolutionary psychology has needed to battle. But the fact that it is difficult should not detract from its necessary programmatic role in a theory of emotion. It is only by gleaning an account of the natural history of emotions, however incomplete and inaccurate it might be, that we could begin to answer the broad functional question about emotions, what they are "for."

To date, there are merely a handful of functional accounts along these lines that have been fleshed out in any detail. For instance, there are models of disgust and models of fear, each of which are grounded in specific aspects of adaptive function (avoiding toxins, evading predators), and that can then be elaborated as a complex web of subprocesses. Indeed, for each of these emotions, the models typically fractionate the initial emotion into subtypes: there are varieties of disgust (for food, for other people, for moral acts; Tybur, Lieberman, Kurzban, & DeScioli, 2013) and varieties of fear (such as anxiety, fear, and panic in threat imminence theory; Mobbs, Hagan, Dalgleish, Silston, & Prevost, 2015). There are also broader accounts that cut across emotions, such as motivational and appraisal theories, which differentiate emotions according to their focus on these properties.

Is it a requirement of our view that all emotions must have evolved? Of course not; we are merely saying that this will be one large aspect of understanding the functional role of emotions in humans and animals. If we want to understand the emotions of robots (which, being functionalists, we certainly believe



Level	Discipline	Questions
Ecological	Comparative ethology	What problems is this emotion adapted for?
Computational	Psychology	What algorithms solve these problems?
Neurobiological	Neuroscience	What neural mechanisms implement these?

Figure 2. Levels of abstraction in a functional account of emotion.
Note. Note that it is possible to view all of these as functional through and through. Even the implementation level could be construed as functional: it is solely through their causal properties that any neurobiological events implement algorithms. Modified from Adolphs and Anderson (2018).

are in principle possible) we would not look to evolution—but the kind of explanation to the question of what the function is "for" would have the same flavor. Need all functional roles of biological emotions have evolved? No, even this is not necessary—we explicitly note that culture and learning are a very major feature of emotions. But even here, the learning that takes place, and the constraints on it, link back to an evolutionary story that provides the grounding.

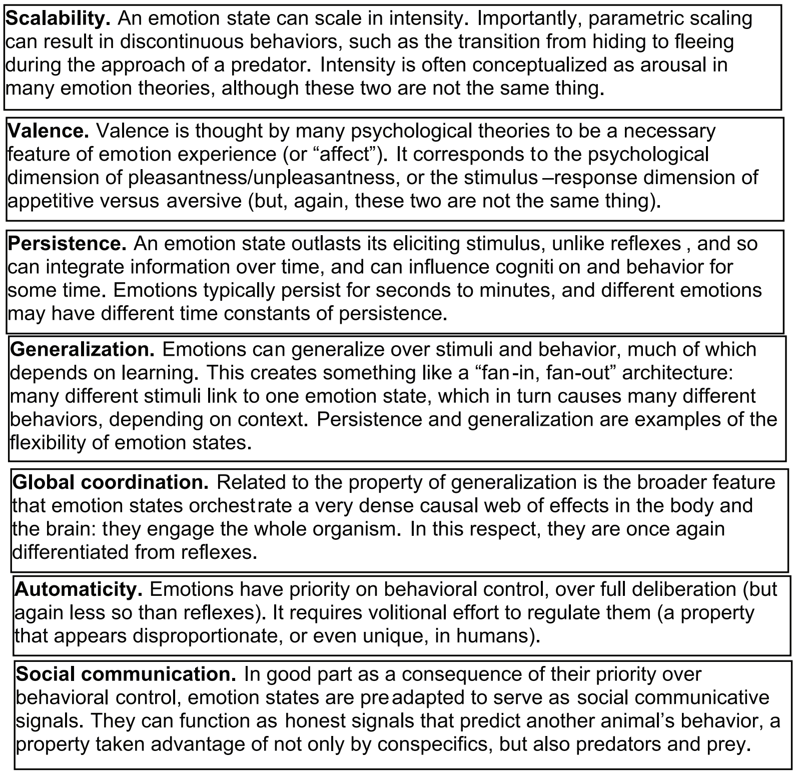
An account of the provenance of the functional role of an emotion, whether evolutionary, evolutionary + learned, or by human engineering, is required to know what that functional role is. Consequently, a molecule-by-molecule duplicate of one of us would not have emotions because there is no history whose reference could provide an answer to the question of what any functionally individuated state is for. Our molecular duplicate would be indistinguishable in all behavioral and physiological respects, and would of course claim to have emotions, to remember things, and to recognize things. But we would be unable to tell if it is remembering or recognizing anything for the same reasons that we could not tell if it is emoting: it never encoded or cognized anything in the past, and its emotional behaviors never had any past stimuli to cause them or past consequences (this situation might quickly change once our duplicate becomes embedded in a history of causal interaction with the world, however brief).

Finally, we again follow the tradition of classical cognitive science in thinking of mental states in general, and emotions in particular, as computational states, so that we can help ourselves to a multilevel description along the lines proposed by David Marr (Marr, 1982): at the highest level of abstraction, the function of an emotion is the biological role it plays within the web of mental states, stimuli, and behaviors ("broad" or "external" functional role; what we called "biological significance"); at the next level, that function is computationally realized by a certain algorithm or set of algorithms (our focus on a functional approach to emotion); and further down, that algorithm is actually implemented by some physical process that very much depends on the physical substrate in which the process is taking place (in most cases, neuroscience; see Figure 2). Each of these levels can be further decomposed into sublevels of increasing concreteness and so this scale is more continuous than tripartite.

Finally, it is important to note that emotions are states of an entire organism. Single brain structures, such as the amygdala, cannot possibly implement an emotion, since that is not their function. The functions of the amygdala can only be defined in relation to the inputs to, and outputs from, this structure—and those inputs and outputs are not sensory and behavioral, but rather pertain to other brain regions. A neural system that processes emotion would then be analyzed compositionally, with each neural component contributing a particular functional, algorithmic, and implementational part that, in the whole system, functions to instantiate an emotion. Thus the Marr-like scheme shown in Figure 2 can be applied across all levels of scale, from organisms (or indeed societies) to brain networks to brain structures to microcircuits.

Challenges

There are a number of challenges faced by the approach we have sketched. One common misconception is that a theory of emotion should be sufficient, by itself, to explain how stimuli result in behaviors. That is, if our aliens came up with a theory of emotions in their science, following the functional approach we just discussed, should they now be able to explain emotional behaviors *solely with this theory of emotion*? Clearly not, and no theory of emotion could, or should, provide that kind of explanation. We do not require this of theories of memory, perception, attention, or even of decision-making. None of these processes, by themselves, are sufficient to explain behavior: they each contribute a part. The same thing holds for emotions.



Scalability. An emotion state can scale in intensity. Importantly, parametric scaling can result in discontinuous behaviors, such as the transition from hiding to fleeing during the approach of a predator. Intensity is often conceptualized as arousal in many emotion theories, although these two are not the same thing.

Valence. Valence is thought by many psychological theories to be a necessary feature of emotion experience (or "affect"). It corresponds to the psychological dimension of pleasantness/unpleasantness, or the stimulus –response dimension of appetitive versus aversive (but, again, these two are not the same thing).

Persistence. An emotion state outlasts its eliciting stimulus, unlike reflexes , and so can integrate information over time, and can influence cogniti on and behavior for some time. Emotions typically persist for seconds to minutes, and different emotions may have different time constants of persistence.

Generalization. Emotions can generalize over stimuli and behavior, much of which depends on learning. This creates something like a "fan -in, fan-out" architecture: many different stimuli link to one emotion state, which in turn causes many different behaviors, depending on context. Persistence and generalization are examples of the flexibility of emotion states.

Global coordination. Related to the property of generalization is the broader feature that emotion states orchest rate a very dense causal web of effects in the body and the brain: they engage the whole organism. In this respect, they are once again differentiated from reflexes.

Automaticity. Emotions have priority on behavioral control, over full deliberation (but again less so than reflexes). It requires volitional effort to regulate them (a property that appears disproportionate, or even unique, in humans).

Social communication. In good part as a consequence of their priority over behavioral control, emotion states are pre adapted to serve as social communicative signals. They can function as honest signals that predict another animal's behavior, a property taken advantage of not only by conspecifics, but also predators and prey.

Figure 3. Table of preliminary list of the features of emotions.

There are a host of enabling conditions, and a host of other cognitive processes with which emotions interact, that make the links from stimuli to behavior possible. Only a full mental architecture, complete with many other cognitive and physiological processes, could produce a full mechanism for the behavior.

This point may also speak to our disagreement with Lisa Barrett's view of emotions as constructed from many other cognitive processes. Arguably, if one feels the need to have an emotion to be a state so comprehensive that it is sufficient to explain behavior, one would need to incorporate all other relevant processes. But this seems counterproductive to us: of course, emotions *interact* with all other cognitive states (see Pessoa, XXXX). But they are not constituted by these processes—we want to individuate them as distinct from other cognitive states. Making distinctions is essential if we want to understand mechanisms.

Consider for instance the fleeing from a predator caused by a state of fear. It makes sense to explain an animal's (or a person's) fleeing as caused by the fear, and it makes sense to look in the brain for the neural mechanisms of fear processing. But the state of fear merely causes a motivation to get away from the predator (amongst other effects). It does not itself provide the detailed description of how this should be done. Depending on the terrain, the lighting, obstacles in the way, I may turn left or right, go fast or slow, and the details of how I flee will be offloaded to other sensorimotor processes and are not a part of what the emotion state needs to explain.

Another challenge is generalizability to other emotions. The good specific examples we alluded to in this article seem restricted to a few emotions like fear and disgust. What about love, embarrassment, pride, awe, and many others? Does our functionalist approach work for all emotions? This is a challenge only in the sense of saying that there's more work to do, but does not seem to us to provide an actual argument for why it should not work (again, bracketing conscious experience). In Figure 3, we outline features of emotions generically, not with any specific emotion in mind. Are there clear counterexamples, a state that everybody agrees is an emotion but which does not have most or any of the properties listed in Figure 3?

To be sure, there are unclear cases, such as spontaneous moods or the feelings induced, say, by listening to music that moves us. These may well lack some of the features listed in Figure 3, and they certainly seem to lack functional descriptions from an evolutionary perspective. It is hard to say what adaptive role they serve, for example. In our view, it is quite unclear whether to call such states emotions; but we are also not strictly wedded to the requirement of a functional role that needs to be explicable in terms of evolution (see previous lines). If they do not exhibit the functional roles in Figure 3, and if they do not play an adaptive functional role that evolution, or our alien visitors, would have seen—well, then the only basis for calling them emotions would seem to be how they feel to us, and we are back to the problems of conscious experience. One could conclude from this that our approach failed to include an important set of emotions, because it bracketed conscious experiences. Or one could conclude that feelings are not a good basis for determining category membership for emotion states, and that, in the absence of other evidence, a feeling of awe induced by music may simply not be an indicator of an emotion state. We are inclined to believe the latter, but more empirical work is needed.

From Emotion Features to Functional Role

The task ahead for a science of emotion is progressively to unpack the picture we sketched here. Needless to say, most of the work remains to be done. While most functionalist theories of emotion in fact already propose specific functional roles for emotions (or for specific emotions), we have not done so here. Here we briefly want to sketch a possible approach to this question. We return to our thought experiment of the aliens observing animal behavior on earth. How might they begin to approach the question? If they located emotions as functional states that explain behavior, without mention of conscious feelings, what properties would these functional states exhibit? In particular, what properties of emotions would distinguish them from all other mental states?

The broadest way to frame this question is to ask how emotions carry out the ecological roles that contribute to the survival of an animal; that is, their "broad" or "external" functional role that gives them biological significance. As we noted, answers to this question are extraordinarily difficult to come by. Another way to frame the question is to ask what properties emotions as commonly conceived show (i.e., start with some clear, ordinary examples). There are of course many properties of emotions that they share with cognitive states in general, but the idea would be that one could come up with a list that would serve as a sort of package that could help to individuate emotions as distinct from other cognitive states. Ideally, this list would span some of the levels shown in Figure 2: they would speak to the ecological (broad) function of emotions, they would suggest algorithms by which that function could be accomplished, and they would motivate hypotheses that could be tested with neuroscience methods.

In Figure 3, we have proposed a preliminary list of features or properties that are intended to apply generically to all emotions (although the particular values some of these parameters take may serve to distinguish different emotions). This list of features is intended to distinguish emotions, generically, from other cognitive states. The list is likely incomplete, and not every instance of an emotion needs to exhibit all of the features in the list (but it should exhibit most of them; more detailed discussion can be found in Adolphs & Anderson, 2018; Anderson & Adolphs, 2014). We generated this list by searching through examples of what we would normally consider bona fide emotions, and asking what properties they exhibit empirically. Consequently, the features are a rather heterogeneous set of attributes. Some of them are probably higher order properties that emerge from the functional architecture in which an emotion state is embedded. Some arise from physical constraints at the implementation level—from how neurobiological systems can actually operate to implement a computation.

This set of features motivates specific neurobiological hypotheses: they suggest ways of implementing the features. Thus, we could search in the brain for specific features, such as persistence. And we could see if we can distinguish different types of emotions on the basis of their implemented features. For instance, we might expect neural systems that process fear to have shorter time constants than those that process sadness, and we might search for emotions that are hypothesized to differ in their degree of coordination and generalization.

Our list of features and their emphases bear some resemblance to other conceptual work on emotion. For instance, Andrea Scarantino's view, the motivational theory of emotion (Scarantino, 2014), derives partly from earlier work by the psychologist Nico Frijda (Frijda, 1986) and sees emotions as tendencies or dispositions for action. This view is not uncommon, and has a lot of prima facie appeal. Like the list of features we provided, Scarantino locates emotions as motivating behaviors while exhibiting certain properties: flexibility, impulsivity, and bodily underpinnings. The motivational theory of emotion, like appraisal theories of emotion, fleshes out detailed stories that serve to distinguish different emotion categories. The particular relational goals towards which emotions motivate behavior serve a similar purpose in Scarantino's theory as the core relational themes do in Richard Lazarus's appraisal theory (Lazarus, 1991; see also Sander, Grandjean, & Scherer, XXXX).

However, unlike these other thinkers, we do not take our list of emotion features to be foundational. Generically, we take the functional approach we described before to be foundational for emotion. Specifically, we take the list of emotion features as helping to point us in the direction of a theory of emotion, and of specific emotions. Particular functional roles and particular computations will correspond to the entries in Figure 3.

We briefly elaborate on one broad feature that can be abstracted from Figure 3 and that may be the most diagnostic of emotions, that helps to situate them at a level of behavioral control between that of reflexes and volitional behavior, and that motivates neurobiological hypotheses about what to look for in the brain. This broad feature might be called "semiflexibility" and to some degree subsumes the features of persistence, generalization, and automaticity in Figure 3. One might also call it "weak modularity" under an updated version of Fodor's notion of modularity (Spunt & Adolphs, 2017). (Although we do not elaborate on it here, the rough characterization would be that emotions are strongly domain-specific, weakly cognitively penetrable, and not at all informationally encapsulated.)

Emotion states are semiflexible in the sense that they accept a limited variety of inputs innately (but a much larger variety of inputs through learning), span a limited time interval, and produce a limited variety of outputs. As we stated, this is in contrast with both reflexes (one kind of input, one kind of output, the entire process occupying a fixed, very brief interval that dispenses with the need for an internal state altogether) and fully reflective behavior (no preset limit on kinds of inputs or outputs, no bound on the duration of the process). It is a large open question, requiring considerably more empirical data, to understand what sets the permissible range of these parameters for emotions (and hence, what, at least in part, helps to distinguish emotions from either reflexes or fully flexible cognition).

We can distil some further points from Figure 3. First, the range of permissible values that these features can take is not rigidly preset, but is itself subject to some flexibility, as learning from experience, education, and volitional effort and training can, and do, modify the initial values to some degree. Second, the set of permissible values depends on the specific emotion and is subject to interindividual differences. This second point offers substantial purchase for neuroscience: we can, in part, distinguish different types of emotion states on the basis of the kinds of inputs they can take or how long they last (e.g., sadness lasts a long time, whereas surprise or fear are more ephemeral), and we may, in part, be able to characterize mood disorders by their abnormal parameter values (they might accept inputs that they should not and last longer than they should, for example). Third, as a first approximation, we may think of a type of emotion state as something akin to the "scripts" or "scenarios" postulated by early work in artificial intelligence and cognitive psychology, which were computational schemes in which certain typical inputs generated certain typical outputs, in a typical time sequence (Shank & Abelson, 1977). For instance, such scripts comprised rules for what to do when going to a restaurant. "Typical" allows for quite a lot of variety: there are indeed many kinds of restaurants and many ways of complying with the explicit and implicit rules of going to a restaurant. But that variety is restricted—this was in fact held against the idea of using such scripts as a reasonable model of cognition; but it is exactly what we think characterizes emotions. One might thus think of these scripts as similar, at the algorithmic level, to the "core relational themes" that classical appraisal theories have put forth (Lazarus, 1991).

In summary, while provisional, a list of emotion features derived from observation (like in Figure 3) could be used to begin assembling a theory of emotion that explicates the functional role of emotions. Such a theory might add or jettison specific entries in Figure 3 as more data are accumulated, and would aim to sculpt a core set of features, a package, that can serve to individuate emotions.

The Task Ahead for Philosophy and for Cognitive Neuroscience

The science of emotion has followed a somewhat erratic path over the last decades, leading to an inordinate amount of confusion even by the standards of an emerging discipline. Can we hope now to make progress towards a shared framework from which to undertake the empirical work? From the perspective we have suggested here, it seems, first, that philosophy of emotion should take the naturalistic turn more resolutely. While it has directed much of its attention to the phenomenology of emotion experience (for a summary, see Solomon, 2009) and to the conceptual analysis of the structure of emotions (for a recent lucid analysis and illustration, see Tappolet, 2016), it would profit from a deeper engagement with the relevant sciences, on a host of issues. These include the provenance of emotions, at both the developmental and evolutionary scale; an account of how errors might arise, for example, in psychopathology; and a taxonomy of emotions based on distinctions in functional role. Perhaps most importantly, we need to aim for a theory that explains not only the distinction between, but also the interaction among emotion and cognition, types of mental states that, we suggest, are both functionally individuated (see Pessoa, XXXX).

Neuroscience, in turn, should focus on how the processing features of emotions listed in Figure 3 are implemented (and perhaps help to revise the list). The cognitive neuroscience of emotion needs to link broad functional accounts with algorithmic ones, aiming to provide computational accounts of what emotions do and how they do it (see Bach & Dayan, 2017). Such accounts can then be used to propose neurobiologically testable hypotheses; and of course, the neuroscience findings will help constrain the accounts. However, we resist the idea that neuroscience data, in and of themselves, and especially at this stage of our meager knowledge of brain function, can provide a foundation for a theory of emotion. The foundation has to come from elsewhere—although, of course, neuroscience should be in the dialogue. Success in forging a science of emotion will require tapping the resources of ethology, evolutionary biology, and paleo-anthropology, so as to pinpoint homologies across species and construct evolutionary scenarios. While the foundation for a theory of emotion should lie in a science of behavior, and thus psychology, our (and the aliens') version of this has a different emphasis from most current versions, in particular in ignoring conscious experience.

We thus advocate a thoroughly interdisciplinary science of emotion that includes philosophy, ethology, psychology, neurobiology, and cognitive science. Calls for interdisciplinary work are notoriously easier to issue than to follow, and this is where philosophy can help, this time in the guise of philosophy of science rather than philosophy of mind. Its role here is not, as previously stated, to propose a theoretical framework in the tradition of philosophical psychology, but rather to inspect the actual procedures and results of the research programs relevant to the study of emotions, and bring to light the structure of the field, its tenets, its inferential structure, its conceptual repertoires, its real or apparent disagreements. Philosophy of science can contribute in this way to a dissolution of misunderstandings and a gradual convergence, and clear articulation, of the research programs involved, as it has in other scientific areas. Nor is this task the exclusive province of professional philosophers: the entire community can and should contribute. Naturalism also consists in fostering the direct involvement of philosophers in the scientific process, and of scientists in the philosophical search for intelligibility.

Declaration of Conflicting Interests

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

Note

[1] There are many possible functionalist theories of the mind. Psychofunctionalism is the functionalist theory that accords with the best future scientific explanation of human behavior (i.e., what our aliens would come up with). As such, a psycho-functional theory may diverge from commonsense or folk psychological views; it also requires all the ingredients of the best science—careful observation and arbitration between competing theories.

References

Adolphs, R. (2017). How should neuroscience study emotions? By distinguishing emotion states, concepts, and experiences. *Social Cognitive and Affective Neuroscience*, *12*, 24–31.

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

Anderson, D. J., & Adolphs, R. (2014). A framework for investigating emotion across species. *Cell*, *157*, 187–200.

Bach, D. R., & Dayan, P. (2017). Algorithms for survival: A comparative perspective on emotions. *Nature Reviews. Neuroscience*, *18*, 311–319.

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

Barrett, L. F., & Satpute, A. B. (2013). Large-scale brain networks in affective and social neuroscience: Towards an integrative functional architecture of the brain. *Current Opinion in Neurobiology*, *23*, 361–372.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*, 227–287.

Damasio, A. R. (2003). *Looking for Spinoza: Joy, sorrow, and the feeling brain*. New York, NY: Harcourt.

Davidson, D. (1987). Knowing one's own mind. *Proceedings and Addresses of the American Philosophical Association*, *60*, 441–458.

Frijda, N. H. (1986). *The emotions*. New York, NY: Cambridge University Press.

Jonas, E., & Kording, K. P. (2017). Could a neuroscientist understand a microprocessor? *PLoS. Computational Biology*, *13*, e1005268.

Lazarus, R. S. (1991). *Emotion and adaptation*. New York, NY: Oxford University Press.

LeDoux, J., & Brown, R. (2017). A higher-order theory of emotional consciousness. *Proceedings of the National Academy of Sciences USA*, *114*, E2016–E2025.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Feldma Barrett, L. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences*, *35*, 121–143.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York, NY: W. H. Freeman and Co.

Mayr, E. (1961). Cause and effect in biology: Kind of causes, predictability, and teleology are viewed by a practicing biologist. *Science*, *134*, 1501–1506.

Mobbs, D., Hagan, C., Dalgleish, T., Stilson, B., & Prevost, C. (2015). The ecology of human fear: Survival optimization and the nervous system. *Frontiers in Neuroscience: Evolutionary Psychology and Neuroscience*, *9*, 55. doi:10.3389/fnins.2015.00055

Panksepp, J. (1998). *Affective neuroscience*. New York, NY: Oxford University Press.

Pessoa, L. (2013). *The cognitive-emotional brain: From interactions to integration*. Cambridge, MA: MIT Press.

Pessoa, L. (XXXX). Emotion and the interactive brain. *Emotion Review, X*, XXX–XXX.

Sander, D., Grandjean, D., & Scherer, K. R. (XXXX). An appraisal-driven componential approach to the emotional brain. *Emotion Review, X*, XXX–XXX.

Scarantino, A. (2014). The motivational theory of emotions. In D. Jacobson & J. D'Arms (Eds.), *Moral psychology and human agency* (pp. 156–185). New York, NY: Oxford University Press.

Scarantino, A. (2016). The philosophy of emotions. In L. F. Barrett, M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (pp. 3–48). New York, NY: Guilford Press.

Shank, R., & Abelson, R. (1977). *Scripts, plans, goals and understanding*. Hillsdale, NJ: Lawrence Erlbaum.

Siegel, E. H., Sands, M. K., van den Noortgate, W., Condon, P., Chang, Y., Dy, J., . . . Barrett, L. F. (2018). Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories. *Psychological Bulletin*, *144*(4), 343–393. doi:10.1037/bul0000128

Solomon, R. C. (2009). Emotions in phenomenology and existentialism. In H. L. Dreyfus & M. A. Wrathall (Eds.), *A companion to phenomenology and existentialism* (pp. 291–309). Chichester, UK: Blackwell.

Spunt, R., & Adolphs, R. (2017). A new look at domain-specificity: Insights from social neuroscience. *Nature Reviews. Neuroscience*, *18*, 559–567.

Tappolet, C. (2016). *Emotions, values, and agency*. Oxford, UK: Oxford University Press.

Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review*, *3*, 397–405.

Tybur, J. M., Lieberman, D., Kurzban, R., & DeScioli, P. (2013). Disgust: Evolved function and structure. *Psychological Review*, *120*, 65–84.

emotion review

Emotion Review
Vol. 10 No. 3 (July 2018) 1–2
© The Author(s) 2018
ISSN 1754-0739
DOI: 10.1177/1754073918772090
journals.sagepub.com/home/er

Comment: Two Challenges for Adolphs and Andler's Functionalist Theory of Emotions

Andrea Scarantino
Department of Philosophy and Neuroscience Institute, Georgia State University, USA

Abstract

Adolphs and Andler's methodological functionalism recommends that affective science focuses on what emotions do rather than on what emotions are physically constituted by or how emotions feel. In addition, it is suggested that the functional roles of emotions should be extrapolated from a set of "features" emotions intuitively appear to have. In this brief commentary, I discuss both prescriptions, focusing on the concept of function and on the role folk psychological platitudes should play in a functionalist theory of emotions.

Keywords
emotions, feelings, functionalism, functions, teleosemantics

Adolphs and Andler's (A&A; XXXX) *methodological functionalism* boils down to two prescriptions:

(a) In constructing your scientific theory of emotions, focus on what emotions *do* rather than on what emotions are *physically constituted by* or how emotions *feel*.
(b) To figure out the *functional roles* of emotions, begin from a set of "features" emotions intuitively appear to have and bootstrap from there.

Prescription (a) pits *functionalism* against the *identity theory of mind*, its primary historical antagonist, and the *feeling theory of emotions*, a still common approach in the study of emotions, recommending that we conceptualize emotions as functional states rather than brain states or states of consciousness.

I agree that emotions can be multiply physically realized, contra the identity theory. If so, neuroscience can discover realizers of functional roles but not identity conditions for emotions. To wit, human fear is *not* identical to the activation of any neural circuit N, although N may realize the fear role in humans (such role may be realized differently in other species/robots).

Furthermore, emotions are not *by definition* states of phenomenal consciousness, although they typically involve such states (Scarantino, 2014, 2016). The connection between emotions and feelings is a piece of folk psychology, not an ironclad constraint on scientific theorizing. Consequently, functionalist accounts of emotions are not fatally crippled by functionalism's alleged inability to capture phenomenal consciousness.

But what notion of function should functionalists endorse? Garden-variety functionalism relies on a *causal-role notion*: the function of a mental state is the causal contribution it makes to the capacities of the organism. A&A presuppose this notion, which they refer to as the "broad" functional role of emotions, when they write that the "function of an emotion is the biological role it plays within the web of mental states, stimuli, and behaviors" (XXXX, p. XXX), a role defined by the causal relations in which emotions are currently embedded.

However, in other parts of their article, A&A (XXXX) endorse an *etiological notion* of function, according to which the function of a mental state X amounts to the beneficial effects X had in the past that explain why X was selected for (see Garson, 2016, for an overview of function concepts). This would make A&A *teleofunctionalists* rather than *garden-variety functionalists*. They write, for instance, that "a molecule-by-molecule duplicate of one of us would not have emotions . . . because there is no history whose reference could provide an answer to the question of what any functionally individuated state is for" (XXXX, p. XXX).

What gives an emotion its (teleo)functional role is not what the emotion disposes its bearer to do, but what the emotion ought to do for its bearer in light of a selection history. Note that it is irrelevant how recent the history is: whether we are focusing on evolutionary history, or cultural history, or learning history, having an etiological function does not entail contributing to any current capacities (but: the shorter the history, the more likely it is that the etiological function = causal-role function).

Despite the claim that an "account of the provenance of the functional role of an emotion . . . is required to know what that functional role is" (Adolphs & Andler, XXXX, p. XXX), prescription (b) makes it clear that A&A are not true-blue teleosemanticists, since they believe the functional role of emotions can be inferred from what emotions currently do. But why do

Corresponding author: Andrea Scarantino, Department of Philosophy and Neuroscience Institute, Georgia State University, 34 Peachtree Street, Atlanta, GA 30302-4089, USA.
Email: ascarantino@gsu.edu

they then deny to molecule-by-molecule duplicates of humans not just a selection history, but emotions themselves?

A&A cannot have it both ways: either the relevant functional roles emerge from an observation of current capacities and duplicates have our emotions since they share such capacities, or the relevant functional roles emerge exclusively from a past history of selection, but then the current capacities we observe cannot shed light on functional roles.

A&A's proposal (XXXX) is motivated by a thought experiment which has aliens ascribing emotions to humans on the basis of observed capacities. At the same time, the account emphasizes the importance of selection history for understanding functional roles. But why would the aliens care about such history, since it is orthogonal with respect to the complex behaviors they want to explain—an emotion can currently have a causal-role function F* that is not its etiological function, and an emotion can currently lack a causal-role function F** that is its etiological function?

My diagnosis is that A&A (XXXX) have conflated two of Tinbergen's (1963) four whys: to explain an emotion, we need to understand its evolutionary origin, its current function (or biological significance), its proximate implementation, and its ontogenetic development. But to understand its current function, we do not need to know about its origin, evolutionary or otherwise.

If methodological functionalism is to treat emotions as causally efficacious latent variables, it must focus on causal-role functions. Understanding selection history is of course vital to the theoretician, but not as a means to the end of explaining complex behaviors in terms of emotions. This is great news, because it is much harder to understand selection history than it is to infer latent variables on the basis of observed behaviors.

Prescription (b) is that we collect a set of "features" of emotions and use them to figure out what causal-role functions they suggest, regardless of historical provenance. The features, which are said to be "plausible" and "derived from observation" (Adolphs & Andler, XXXX), include the following: emotions have different intensities, they can have positive and negative valence, they require an appraisal of the stimulus, they persist from seconds to minutes, they can be caused by different stimuli and they can cause different responses, they engage the whole organism unlike reflexes, they are irruptive but allow for regulation, and their expressions can communicate.

These strike me as folk psychological platitudes about emotions. Deriving a functional role from them would make emotions functional correlates of a folk psychological theory. I take it that this is *not* what A&A (XXXX) recommend—their choice of the label *psychofunctionalism* to designate their theory suggests as much, because *psychofunctionalists* identify mental states with the functional correlates of *scientific* rather than *folk* psychology.

What I am less clear about is how A&A understand the relation between folk psychological and scientific emotion concepts. I have argued that the folk psychological notion of emotion is too heterogeneous to allow for any scientifically interesting generalizations to apply to all or even most of its members (Scarantino, 2012). To develop enhanced conceptual frameworks that best serve our scientific needs, (psycho)functionalist theories of emotions must focus on subvolumes of the hyperspace of emotional continua captured by the platitudes.

We need a theory T_1 of emotions that are, say, highly intense, short-lived, caused by pattern-matching primitive appraisals, present across species, endowed with subcortical neural circuits, and so on; another theory T_2 of emotions that are less intense, long-lived, caused by language-dependent central appraisals, only present in humans, lacking dedicated neural circuits, and so on.

I call this *methodological pluralism*—the view that there is no single psychological kind (let alone neurobiological kind) individuated by our folk psychological platitudes, even granting some "massaging" of folk intuitions. On this view, a (psycho) functionalist theory of emotions should treat the folk psychological platitudes about emotions as scientific astronomy treats the platitudes about celestial bodies of folk astronomy. They are just a preliminary way to individuate the phenomena to be scientifically investigated, and preserving as many of them as possible should *not* be an objective of scientific analysis.

Is this also A&A's (XXXX) view? Is methodological pluralism what ultimately drives their view that the current emotion concepts of folk psychology are problematic? Do they have any suggestions for distinguishing between acceptable and nonacceptable discrepancies between scientific and folk psychological emotion concepts, and for arbitrating between (psycho) functionalist accounts that may differ in terms of how well they accommodate the folk psychological platitudes?

References

Adolphs, R., & Andler, D. (XXXX). Investigating emotions as functional states distinct from feelings. *Emotion Review, X*, XXX–XXX.

Garson, J. (2016). *A critical overview of biological functions*. Dordrecht, The Netherlands: Springer.

Scarantino, A. (2012). How to define emotions scientifically. *Emotion Review*, *4*, 358–368.

Scarantino, A. (2014). The motivational theory of emotions. In D. Jacobson & J. D'Arms (Eds.), *Moral psychology and human agency* (pp. 156–185). Oxford, UK: Oxford University Press.

Scarantino, A. (2016). *The philosophy of emotions and its impact on affective science*. In M. Lewis, J. Haviland-Jones & L. Feldman Barrett (Eds.), *The handbook of emotions* (4th ed., pp. 3–48). New York, NY: The Guilford Press.

Tinbergen, N. (1963). On the aims and methods of ethology. *Zeitschrift für Tierpsychologie*, *20*, 410–433.

emotion review

Emotion Review
1–3
© The Author(s) 2018
ISSN 1754-0739
DOI: 10.1177/1754073918772092
journals.sagepub.com/home/er

Author Reply: We Don't Yet Know What Emotions Are (But Need to Develop the Methods to Find Out)

Ralph Adolphs
Division of Humanities and Social Sciences and Division of Biology, California Institute of Technology Pasadena, USA

Daniel Andler
Department of Philosophy, Université Paris-Sorbonne, France
Department of Cognitive Studies, Ecole Normale Supérieure, PSL Research University, France

Abstract

Our approach to emotion emphasized three key ingredients. (a) We do not yet have a mature science of emotion, or even a consensus view—in this respect we are more hesitant than Sander, Grandjean, and Scherer (henceforth "SGS") or Luiz Pessoa (henceforth "LP"). Relatedly, a science of emotion needs to be highly interdisciplinary, including ecology, psychology, neuroscience, and philosophy. (b) We recommend a functionalist view that brackets conscious experiences and that essentially treats emotions as latent variables inferred from a number of measures. (c) But our version of functionalism is not definitional or ontological. It is resolutely methodological, in good part because it is too early to attempt definitions.

Keywords
emotion, feelings, functionalism

Response to Scarantino

We thank Scarantino (XXXX) for having accurately summarized our position (Adolphs & Andler, XXXX) in his opening sentences. We also appreciate his attempt to relate our views to other positions in the philosophy of mind, but these comparisons are sometimes ill-posed. In particular, our view is emphatically methodological, and we consequently do not share many of the premises or metaphysical commitments of others. More than that, we feel that much of the conceptual entrenchment that most theories of emotion bring with them should be put aside. In a science of emotion we are studying a natural phenomenon that depends on empirical observation (although philosophy plays an important role as well, but more as philosophy of science than as metaphysics).

For instance, we do not see feeling theories of emotion as necessarily incompatible with our functionalist perspective—as we note in our article, we have chosen to bracket feelings only for methodological reasons. We see no reason that feelings could not be functionally explained in principle (or at least, functionalism seems to us about as promising in this regard as any other theory of consciousness). Perhaps some organisms can have emotion states yet be incapable of conscious experiences; perhaps others require conscious experiences to accompany emotions; and perhaps yet others have emotions that are sometimes conscious and sometimes not. We suspect that we are in agreement with Scarantino (XXXX) that these are empirical questions—we just do not yet have a good methodological approach to answer them.

Scarantino (XXXX) also remarks on psychophysical identity theories that equate brain states with mental states. Identity theories of the mind cover too great a variety to discuss in detail here, but in general we also do not see them as a serious challenge to functionalism, for several reasons. One reason is that identity theories often seem to be about conscious experience in one way or another—just the topic we decided to bracket anyway. A second reason is that, as both SGS (Sander, Grandjean, & Scherer, XXXX) and LP (Pessoa, XXXX) point out, it is nowadays (but not yet in the days when much of the philosophy of identity theory was developed) an empirical fact that emotions cannot simply be related to a single brain structure or neurotransmitter. Emotions are system-level phenomena that depend on dynamic network interactions in the brain. That fact alone would seem to substantially weaken any theory that type-identifies emotions with brain states, because those brain states would have to be individuated by criteria that begin to look quite functional. Brain systems and networks are relevant to

Author note: We thank Andrea Scarantino for helpful comments on a draft of this response, and Frederick Eberhardt and Christopher Hitchcock for stimulating discussion. Funded in part by a Conte Center grant from NIMH.

Corresponding author: Ralph Adolphs, Division of Humanities and Social Sciences and Division of Biology, California Institute of Technology, HSS 228-77, Caltech, Pasadena, CA 91125, USA. *Email*: radolphs@hss.caltech.edu

cognition and behavior in light of what they do; asking what color they are or how much they weigh is besides the point (once again, we think we are in agreement with Scarantino here). LP seems to have something similar in mind when he notes that "the boundary between anatomy and function becomes blurred," and that functional connectivity will be the right metric to delineate networks (Pessoa, XXXX, p. XXX).

An important consequence of the fact that our functionalism is methodological (rather than ontological) is that it acknowledges a multilevel view of functional role. We do not choose between etiological functionalism and causal-role functionalism but think that both are useful. At the "broadest" functional level, emotions, according to our view, are grounded in descriptions from behavioral ecology. These are then realized in more "narrow" functional roles at the level of algorithms and causal mechanisms (see Figure 2 in Adolphs & Andler, XXXX). The broad, etiological functional level is essential to account for *mal*functions: something needs to be known about the natural or engineered history of the organism (teleofunctionalism), without which there is simply no fact of the matter about the proper function of an emotion (since we cannot distinguish proper functions from errors). It is for this reason that we deny a spontaneously created molecular duplicate those states that we can individuate as emotions.

Scientists are not faced with spontaneous molecular duplicates, which have no natural history at all, and instead are faced with organisms whose natural history is simply exceedingly hard to figure out. Hence our methodological recommendation to focus on the features that we listed in our Figure 3 (Adolphs & Andler, XXXX; see also Adolphs & Anderson, 2018). They happen to reflect functional roles in the real world. Currently observed emotion features broadly reflect a history based on functional criteria under normal circumstances. Aliens observing life on earth would draw similar inferences. Of course, things are more complicated: new functions evolve, and the current function of an emotion needs to be interpreted dynamically and relative to both historical and current context. But an instantaneously created molecular duplicate has no recent or remote history at all; the only possible grounding for an emotion would indeed be a physical identity theory. As soon as we begin observing the causal interactions between the molecular duplicate and its environment, we can of course begin to make sense of its functional states—but that requires a history of interaction with the world, however brief.

We can only make sense of the broad function of emotions if we assume that there is some continuity with the past. That is how we can attribute fear to a rat in a Pavlovian fear-conditioning experiment in the lab: there is no advantage to the survival of the rat in the experiment, but we can tell a plausible story about the proper function of the fear state—it is just the state that played such a role in the natural environment of the rat in the past. Without such a presumption, there is no telling what the function is (it is certainly not, for example, "to help the experimenter publish an article in *Nature*").

Folk psychology begins to catalogue some of the features of emotions, current science helps refine this list, and the future best science would provide further improvements to the list. We need to start somewhere, and so the list we produced in Figure 3 of our article (Adolphs & Andler, XXXX; see also Adolphs & Anderson, 2018) was intended as a preliminary starting point. It is one that has considerable scientific support, but that will surely be revised. This raises an interesting question about the contrast between our folk psychological concepts of emotions and of propositional attitudes like beliefs and desires. Although much more work would be needed on the topic, it seems plausible to us that while there are many folk psychological platitudes about propositional attitudes ("if I believe it is going to rain and I desire to stay dry, I take an umbrella"), there seem to be few comparable folk psychological platitudes about emotions. If we are right about this intuition, it would provide a justification for a more empirical, discovery-based approach to emotions than in the case of propositional attitudes. Functionalism about beliefs and desires may well be of a definitional kind, whereas functionalism about emotions should be more methodological.

Affective Neuroscience

SGS (Sander et al., XXXX) and LP (Pessoa, XXXX) both stress brain networks as an important level of analysis and description in understanding emotion. Whereas LP develops a more holistic view in which the interaction between emotion and other cognitive processes needs to be explained by a dynamical systems approach, SGS provide five specific networks that implement particular components of an emotion. There is much to like about the attempt by SGS to map emotion components onto specific brain networks, offering a more hypothesis-driven approach for affective neuroscience. Whether or not this will actually work, or whether, as LP seems to suggest, complex cross-network interactions will instead need to be analyzed, remains to be seen.

We are in full agreement with LP (XXXX) that a network-level account will be needed for the neuroscience of emotions, as it is needed for all of cognition. The very fact that there is context dependency, and that emotion interacts with other aspects of cognition, requires this. On the other hand, as noted before, we feel that pretty much all of the hard work remains to be done: we know almost nothing about the particular architectural constraints, or specific implementations, by which network-level neuroscience explains much about emotions. This is good news: neuroscientists can get to work.

We agree with LP (XXXX) that a simple division between cortical and subcortical regions needs to be replaced with a more integrated view, and that such integration also accounts for much of the context sensitivity that animals exhibit. However, we suspect that a stronger notion of "context dependency," together with a somewhat clearer partition of functional roles to cortical and subcortical circuits, will also be useful. While the effects of context are pervasive, they are more restricted in the case of emotions than in the case of beliefs. Absolutely anything can contribute to my beliefs about the world, but only certain kinds of contextual information, or contextual information with certain

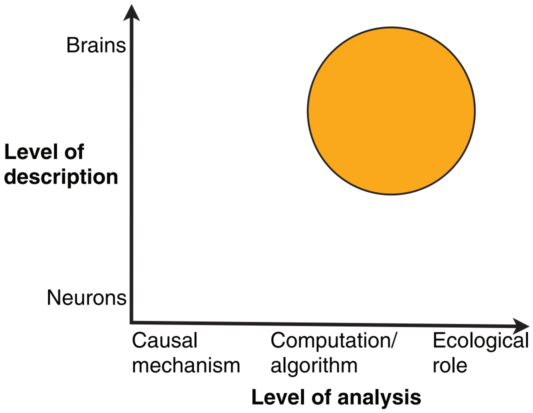


Figure 1. Locating emotions by level of analysis and description.

features, can contribute to certain emotions. We would speculate that there is a narrower, and more rigidly specified set of contextual effects that are implemented through subcortical structures alone, and that there is a broader, and more flexibly specified set of contextual effects that are implemented through the interaction between cortical and subcortical structures. The former would have the widest phylogenetic continuity, whereas the latter might be most apparent in humans.

Two important challenges are raised by SGS' (XXXX) emphasis on "relevance." (a) Is it current relevance that matters, or is it relevance viewed over evolutionary history? We think both matter, but that in order to understand the former, we need to know the latter (allowing us to distinguish true adaptations from exaptations or errors). And (b) all adaptations are about relevance or survival (ignoring spandrels), failing to provide anything that would distinguish emotion in this respect from any other cognitive state. It is here perhaps that appraisal theory offers a key contribution, in fleshing out just exactly what "relevance" means, and how different types of relevance might serve to distinguish different emotion categories. Our own view borrows the idea of an "interrupt" mechanism from Herbert Simon (1967): emotions are engaged specifically when normal ongoing cognitive processes need to be interrupted, because a novel challenge has been encountered. An antelope grazing in the savannah is engaging a host of cognitive processes in the service of survival, but it is the sight of a lion that interrupts this activity through the elicitation of an emotion state. An important part of the function of emotions is thus to take over control when business as usual is no longer adaptive. This feature, together with a narrower notion of context dependency, was part of what we called "semiflexibility" in our article. There are times when maximal flexibility is not a good strategy and a more automatic level of control needs to take over, as emotions do.

Concluding Thoughts

Given the issues of multiple realizability, our general lack of knowledge about the nervous system, and challenging questions about the most useful taxonomy for emotions, what role can neuroscience play? We comment on two roles: figuring out the level of description and analysis that is most useful; and providing data for categorizing emotions according to their similarity to one another. With respect to the first, all levels of description (micro to macro) and abstraction (Marr-like levels) are important, but they are not equally important. We feel that a somewhat more abstract, and somewhat more macroscopic, region of this space is currently the best suited for emotion science (see Figure 1). This would correspond to computational models that are brought to bear on neuroscience data at the systems level. This region of the space, at least in the human brain, can be fruitfully investigated with techniques like fMRI. However, the sheer complexity of the systems, nicely summarized by LP (XXXX), as well as multiple functions subserved by every brain system, make it difficult to link specific brain networks with specific emotions.

Instead, a more data-driven approach could begin to examine similarity relationships, an approach we have stressed elsewhere (Adolphs & Anderson, 2018). Similarity spaces can also correspond to dimensional ways of characterizing emotions. Indeed, the precise dimensions may not be so important as long as we have enough to capture relevant variability. Contra Scarantino (XXXX), we do not see that this requires "methodological pluralism." Some emotions are short-lived, some longer lasting, some more or less intense—but persistence and intensity are continuous variables, not discrete categories per se (emotions may turn out to cluster in such a dimensional space empirically, but that's a fact about how they occur in nature, not about the methods we choose to characterize them).

References

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

Adolphs, R., & Andler, D. (XXXX). Investigating emotions as functional states distinct from feelings. *Emotion Review*, *X*, XXX–XXX.

Pessoa, L. (XXXX). Emotion and the interactive brain. *Emotion Review*, *X*, XXX–XXX.

Sander, D., Grandjean, D., & Scherer, K. R. (XXXX). An appraisal-driven componential approach to the emotional brain. *Emotion Review*, *X*, XXX–XXX.

Scarantino, A. (XXXX). Comment: Two challenges for a functionalist theory of emotions: Functions and the role of folk psychology. *Emotion Review*, *X*, XXX–XXX.

Simon, H. A. (1967). Motivational and emotional controls of cognition. *Psychological Review*, *74*, 29–39.