Emotional Appraisal is not Memory

Abstract: There is supposedly a distinction in kind between basic and higher cognitive emotions, where basic emotions are products of evolution, whereas higher cognitive emotions are products of culture and experience. This view is supported by putative differences in input/appraisal processes: in basic (but not higher cognitive) emotions, these processes are realized by a proprietary, encapsulated memory system, fine-tuned via Pavlovian learning. Developments in decision science undermine this putative difference in appraisal processes by introducing a new form of Pavlovian learning. Integrated into appraisal models, this makes it possible to collapse the distinction between basic and higher cognitive emotions.

# Low-level Appraisal Processes and the Discontinuity of Basic and Higher Cognitive Emotions

Many have argued or assumed that there is an important discontinuity between basic and higher cognitive emotions. Among the most influential of these is Paul Griffiths (1997). Griffiths followed Ekman (2005) in arguing that basic emotions are a restricted class of emotional responses that have some overlap with folk emotion categories such as anger, fear, joy, sadness, surprise, and disgust. On the output side, affect programs were posited to explain the coordination of various dimensions of these emotional responses including (among other things) brief, stereotyped facial expressions, changes in posture, and activation of the autonomic nervous system (Ekman and Cordaro 2011; Ekman 1971). These programs were thought to appear early in development and were thought to be pancultural and shared with other primates.[[1]](#footnote-2) Finally, these outputs can be elicited without mediation from higher cognitive processes involving explicit beliefs and desires. By contrast, higher cognitive emotions are thought to be highly culturally variable and thought to lack stereotyped physiological responses. Finally, some claim that the occurrence of higher cognitive emotions *requires* phylogenetically recent cognitive inputs (e.g., inferential links with beliefs or desires, causal reasoning, self-awareness, theory of mind). Higher cognitive emotions include not only emotions like shame, guilt, embarrassment and pride (e.g., Clark 2013), but also some instances of folk emotion categories like instances of anger or disgust that are elicited by moral offenses (Griffiths and Scarantino 2009). In such cases, Griffiths and others claim that the basic emotion of anger (or disgust, etc.) is a distinct psychological kind from higher cognitive forms of anger. Importantly, the features of higher cognitive emotions and basic emotions are not definitions but a posteriori identities. Accordingly, these characterizations are subject to revision as we learn more about basic and higher cognitive emotions: one could question whether there are any basic emotions (e.g., Barrett 2017), whether they should be thought of as emotions at all (e.g., as opposed to “survival circuits,” Ledoux RRR), or one could question whether phenomena of higher cognitive emotions need to be explained by mechanisms that require higher cognitive inputs for their activation (as I do toward the end of this paper).

With an impressive synthesis of the scientific evidence available at the time, Griffiths (1997) argued that “there is no reason to suppose that any eventual successful treatment of [higher cognitive emotions] will be an outgrowth of the [basic emotion] approach.” (p. 100) This is one way of stating the *discontinuity hypothesis*, that there is an important discontinuity between the higher cognitive emotions and basic emotions.[[2]](#footnote-3) My aim here is to argue against this hypothesis of discontinuity by attacking the model of emotional appraisal on which it depends in light of recent developments in learning theory and decision science.

Griffiths’ case for this hypothesis is complex, and part of that case concerns the nature of emotional responses as understood on the [basic emotion] approach. In other words, much of his concern is with the output side of emotions. Nevertheless, the larger part of his case rests on the input side of emotions and the nature of appraisal processes, which determine when an emotional response will occur. That is the part of his argument that interests me here.[[3]](#footnote-4)

Here is a summary of his case: “The stimulus appraisal which initiates an affect program reaction is to a large extent informationally encapsulated… [By contrast, higher cognitive emotions] seem more integrated with cognitive activity leading to planned, long-term actions...” (p.100) What does Griffiths mean by informational encapsulation? He means that appraisal processes “…cannot access all the information stored in other cognitive systems and it can store information that contradicts that other information. Conscious beliefs concerning, for example, the harmlessness of earthworms does not get taken into account when the system is deciding upon a response…” (p. 93) Ekman (1999) calls these stimulus appraisal processes the automatic appraisal mechanism (AAM). And aside from informational encapsulation, Griffiths points out that “Ekman’s AAM… possess[es] many of the features which Jerry Fodor (1983) lists as characteristics of modular cognitive systems.” This includes mandatoriness (e.g., “People often respond with fear or anger to a given stimulus whether they choose to or not.” Griffiths 1997, 93) and opacity to central cognition (e.g., “People are aware of [AAM] outputs, which are the emotional responses themselves, but not aware of the processes that lead to them.” Griffiths 1997, 93). So, according to Griffiths, the AAM is modular in these senses.

## Evidence for Modularity of the AAM

Why think that basic emotions exhibit these and other components of modularity? Griffiths claims that “The data on the triggering of emotional response…lend themselves very naturally to a modular interpretation…the rules governing the acquisition of information about what is emotionally significant are importantly different from those used to acquire information on other topics. This in turn suggests that a distinct psychological system is involved.” For the most part, Griffiths is referring to data gathered by Seligman (1971) and others on the learning of phobias and disgust elicitors, especially in the domain of classical or Pavlovian conditioning. The fact that phobias are resistant to extinction and can be learned after a single trial suggests that they are best explained by a highly selective, prepared learning mechanism. Ohman’s detailed work on fear conditioning in humans is one important developments of these data (see, e.g., Ohman and Mineka 2001). This work suggests that fear-relevant stimuli, like images of snakes and spiders are more easily associated with shocks and other aversive stimuli, than are fear-irrelevant stimuli like houses or flowers. Snakes and spiders constitute a special category of fear-relevant stimuli, because they represent persistent threats to human life over evolutionarily significant time scales, whereas guns and broken electrical equipment are more recent threats. So, additional evidence of selective association for snakes (etc.) and shocks over electrical outlets and shocks suggests an evolutionary explanation involving a special purpose learning mechanism.[[4]](#footnote-5) As Griffiths puts it, data like these suggest “…that the learning mechanism is biased. It is not a general-purpose mechanism, equally capable of learning any facts that the world may embody.” (p.89) In other words, this suggests the existence of a domain-specific learning mechanism for fear. And so, Griffiths primary case for modularity concerns these phenomena of classical or Pavlovian conditioning. Later, I will point out a substantial revision in our understanding of the mechanisms underlying this kind of conditioning, which should mitigate the inferences Griffiths makes here.

While domain-specificity is central to the idea of modularity (see, e.g. Coltheart RRR), the preceding evidence does not directly demonstrate informational encapsulation or mandatoriness or opacity. More recently, Griffiths has appealed to a much wider range of evidence to support the modularity of the AAM and also the discontinuity hypothesis (Griffiths 2013; Griffiths and Scarantino 2009). For example, he refers to Ledoux’s famous demonstration of the existence of two distinct pathways for fear (or anti-predator responses) in rats (1993), one of which bypasses cortical visual processing. This suggests that fear is fast, mandatory, and inaccessible to other systems (meaning that it can store or respond to information to which other systems do not have access). Additional evidence can be found in backward-masking experiments in humans (e.g., Soares and Öhman 1993), which also demonstrate opacity, since they demonstrate fear conditioning and fear responses to stimuli outside of conscious awareness.[[5]](#footnote-6) As Griffiths and others are keen to note, the very phenomenon of phobias suggests informational encapsulation. For instance, my fear of a spider is not overcome by my knowledge that its bite is not poisonous. Fear persists because it cannot access everything that I know. All things considered, there is modest evidence for the modularity of fear, and much of this evidence concerns the input side of fear. Thus, fear will be a central focus henceforth.

## The AAM and Memory

Griffiths expresses a critical assumption about the AAM, which I will call the *appraisal as memory* assumption:

The AAM would have some form of memory, storing information about classes of stimuli previously assessed as meriting emotional response. The AAM would receive perceptual information, presumably at quite a basic level of analysis, and compare this to ‘memories’ which would take the form of generalizations about the significance of certain perceptible features. I do not mean to imply a physiological separation between the memory and the decision-making system, only to distinguish two functional aspects of the AAM. The two might be physiologically identical in the manner of some connectionist systems. The effect of the AAM would be to allow a more rapid response to important events than would be possible if it were necessary to evaluate their significance using all available information. It would also cause the organism to be conservative, producing an emotional response more often than was actually necessary. (Griffiths 1997, 92-93)

In other words, appraisal is assumed to include some proprietary memory system (possibly as many as one for each emotion). If we think of the memory system along the lines of a database (as does Ekman and others, e.g. Prinz 2004), then the contents of database would not be cognitively complex. Given Griffiths commitments so far, they would be formatted in terms of sensations, low-level perceptions, or associations thereof. Moreover, the contents of the memory system would be inaccessible to and encapsulated from higher cognitive systems. Cognition can't access database contents, nor are the contents updated by cognition (as depicted in Figure 1). For Griffiths, this assumption is importantly connected to the evidence of modularity described above: “The biased nature of the learning mechanism provides a further reason for thinking of the emotion-triggering system as independent of higher cognitive processes, since the biases seem to be specific to the learning of emotional responses.”

A screenshot of a cell phone

Description automatically generated

Figure The appraisal as memory model of emotion appraisal. Cognition cannot influence appraisal directly. AAM Decision System and Learning System may interact or may be identical on this model.

It follows from these assumptions that if an emotional state can be directly triggered by phylogenetically recent cognitive developments, such as metacognitive or language-dependent concepts, then the emotional state cannot be the same type of emotion as a basic emotion.[[6]](#footnote-7) Rather, if basic emotion fear conditioning is implemented by some kind of proprietary database, then there are limited ways in which higher cognitive systems can elicit basic emotions, for example, via fairly vivid imagery.[[7]](#footnote-8) Griffiths is not alone in drawing these conclusions. Several other philosophers and psychologists have followed his lead, (Matthen 1998; see, e.g., Izard 2007) and mistakenly, I will argue.[[8]](#footnote-9)

# An Alternative Model

Nevertheless, there is a competing model of appraisal that can account for much of the phenomena of modularity and also provide a unified explanation of basic and higher cognitive emotions.[[9]](#footnote-10) Moreover, this model does not imply the discontinuity of basic and higher cognitive emotions. In fact, it shows that some of the phenomena previously attributed to the so-called higher cognitive emotions can be explained by the same appraisal and production processes involved in basic emotions. This model is inspired by recent developments in reinforcement learning.

To understand these developments, it helps to understand the basics of Pavlovian learning mechanisms, which explain the phenomena of classical conditioning. In classical (or Pavlovian) conditioning, the learned relation is between some outcome (e.g., the delivery of food) and some cue that predicts the outcome (e.g., the chiming of a bell). The delivery of food is an important kind of outcome for a dog; so important that dogs do not need to learn of its importance through any other form of reinforcement, which is why it has been called an “unconditioned stimulus” or a “primary reinforcer.” Because food delivery is such an important outcome, it triggers the so-called “unconditioned response” of salivating, which is to say that the dog does not need to learn to salivate in response to the food. In any case, when the delivery of food is frequently paired with a conditioned stimulus (CS) or cue, like the ringing of the bell, the dog learns to expect the food whenever the bell rings. Moreover, it is only because the cue predicts an important outcome that a dog salivates.[[10]](#footnote-11) Learning theorists standardly assume that this kind of learning is the basis for innate responses like affect program responses: Pavlovian learning is likely the basis for so-called *Pavlovian responses*, such as rats freezing in response to predictors of painful stimuli, avoiding substances that are predictors of nausea, and getting ready to fight in response to predictors of territorial conflict. These kinds of responses clearly correspond with components of basic emotion fear, disgust, and anger, respectively.[[11]](#footnote-12)

This line of research already suggests a division between appraisal on the one hand and learning and memory on the other. Remember that Griffiths and others encourage us to think of appraisal processes as like a database of stimuli that trigger an emotion. Once we determine that the role of Pavlovian learning is to find cues predictive of outcomes, we see clearly that what is learned is not a new elicitor for an emotion, but rather a cue that is predictive of an *outcome* (as evidenced by blocking experiments, cf. fn. 10) that elicits the Pavlovian response or emotion. This allows the emotion to be more directly elicited by the predictive cue. But then, *the associations that guide learning are not between a stimulus and a Pavlovian response* (e.g., salivation or fear) but between the stimulus and an outcome that elicits the Pavlovian response without any conditioning (e.g., food or pain). So, it is not quite accurate to characterize the process of conditioning as learning a new elicitor for the Pavlovian response, as the *appraisal as memory* view claims. The organism is not salivating or responding fearfully because it remembers doing so when it heard a certain sound. Rather, the elicitation of the response is tied directly to an outcome, and what is learned is a predictor of the outcome. The organism learns to anticipate food or pain upon hearing the sound, and it is the anticipated food or pain that triggers the response.

Even so, a defender of that view could easily reply that this places little strain on *appraisal as memory*. The bottom line remains the same: the system learns a new rule for activating the Pavlovian response; one that could be represented as a rule stored by the proprietary learning/memory system. Perhaps the rule could be expressed as a conditional: if you hear that tone again, respond with salivation (or fear). However, it would be more accurate to express the rule in the following way: if you hear that tone again, expect food (or pain). On my view, the latter rule actually does not sound like learning a new elicitor. Rather, the response is still elicited by a representation of the unconditioned stimulus (i.e. the thing expected).

Regardless, a recent experiment pushes up bigger wrinkles for this way of understanding how new elicitors of Pavlovian responses might be acquired. In it, Robinson and Berridge (2013) conditioned rats to expect a squirt of intensely salty water (3 times as salty as sea water) into their mouths after the insertion of a lever into the cage accompanied by a sound (call this CS1). They conditioned the same rats to expect a squirt of sweet sucrose into their mouths after a distinct CS2, the insertion of a different lever accompanied by a different sound. As expected, the rats learned to avoid the CS1 lever whenever they heard the CS1 sound and to approach the CS2 lever (and lick it, etc.) whenever they heard the CS2 sound. This is Pavlovian conditioning because the rats learned that certain sounds predict certain outcomes; outcomes that elicit pavlovian responses (like a gape response and avoidance of a distasteful stimulus or licking and approach toward a tasty stimulus). After this initial training phase, the rats were injected with drugs that caused their salt levels to plummet. They were again presented with CS1 for the intensely salty water again but this time, in extinction (that is, without the squirt of salt to accompany it). But this time, due to their salt-deprived state, the rats approached the CS1 lever with the same intensity as they had toward the CS2 lever.

Dayan and Berridge (2014) argue that this experiment reveals a type of Pavlovian learning that had not been anticipated or observed. Previously, it was assumed that all Pavlovian learning was *model-free*, meaning among other things that there is a single, or *cached*, representation of the outcome’s value (as liked or disliked) and its *identity*, or sensory characteristics (e.g., its saltiness). If this information were stored separately and used to guide learning, it would involve a model with separable representations to mitigate the relationship between the sensory characteristics and the value of the outcome, perhaps to guide the response more flexibly (when the value of the outcome changes due to a salt deprived bodily state, for example). Instead, on the model-free account, changing the represented value of the outcome (and thus the organism’s response to the stimulus that predict it) would require shifting the cached value gradually through repeated presentations of the stimulus and outcome (with a positive reward value).

If we contrast this with a different kind of learning mechanism, instrumental learning (i.e., operant conditioning), we find that the learned relation in this case (between a behavioral response and an outcome) is represented in terms of a complex model of the state of the organism (e.g., hunger, thirst, low body temperature), the contingency relations between responses and outcomes, the stimulus identity (e.g., saltiness or sweetness), and the value of the outcome for the organism (in a given state). This kind of model-based learning allows an organism to quickly shift its behavioral strategies as the state of the organism changes or as contingency relations change. In the salty water experiment described above, the rats were capable of immediately changing their behavior in relation to the hypersaline solution when they entered a salt-deprived state. In this way, their behavior resembled the behavior produced by model-based systems for instrumental learning. But unlike instrumental learning, the learned relation is between an outcome and a predictive cue (rather than a behavioral response and an outcome). So, this form of learning looks thoroughly Pavlovian, as were the characteristics of the response (i.e., the aversive and appetitive reactions to the lever). Yet, the way that the rats immediately switched their behavioral response from avoidance to approach (upon entering into a salt-deprived state) suggests an ability to represent the value of the stimulus separately from its other characteristics. In other words, this appears to be an instance of *model-based Pavlovian learning*.[[12]](#footnote-13) It also demonstrates a form of learning that is not proprietary to a specific Pavlovian response, since the learned relation between the cue and the hypersaline solution was capable of activating diametrically opposed Pavlovian responses, one appetitive and the other aversive.

This suggests that there may be two separate learning systems for Pavlovian responses: one model-free and perhaps also domain-specific; the other model-based and relatively domain-general. Nonetheless, it also drives a wedge between the model-based learning system on the one hand and appraisal systems for appetitive or emotional responses on the other. In particular, the model-based learning system finds cues predictive of outcomes, but the fact that an outcome has triggered a Pavlovian response previously (e.g., distaste in response to the hypersaline solution) does not mean that it will always trigger the same response in the future. The database metaphor suggests that it is the *stimulus* that elicits the emotional response, but model-based Pavlovian learning involves learning a relation between a stimulus and an *outcome*, where the outcome could have different significance for the organism depending on the state of the organism (e.g., osmolyte concentrations, blood sugar, reproductive state, etc.). If we define appraisal as a determination of significance for the organism, then this process appears to be clearly separate from memory processes that help predict the outcome. This suggests a slightly different model of how conditioning occurs for emotional responses, as depicted in figure 2.

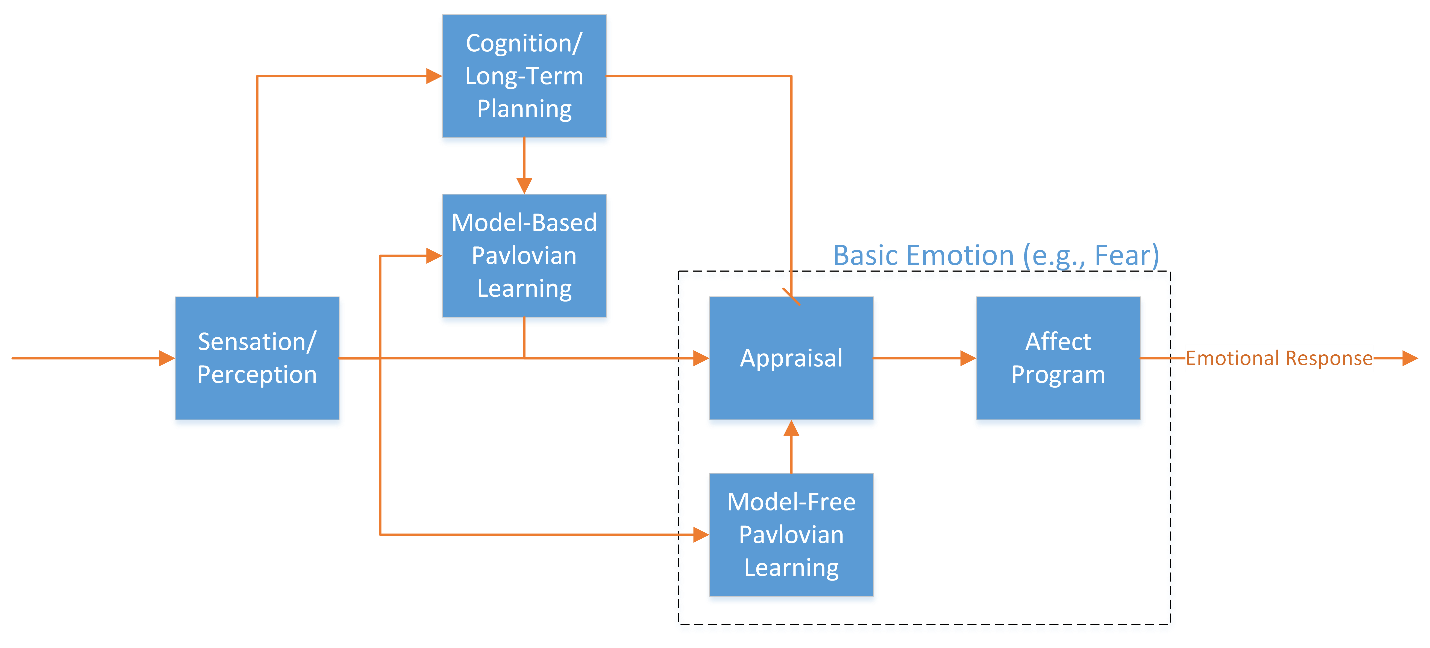


Figure On this model, appraisal is separate from memory. There is a domain-general and a domain-specific memory process that informs appraisal. Cognition can influence model-based Pavlovian learning fairly directly, but not automatic appraisal.

This model allows that an emotion such as fear could be triggered by more cognitively complex representations (e.g. metacognitive representations). We can easily imagine how the model-based system could have been modified in our lineage to represent complex cues and use them to predict situations with adaptive significance for our lineage.

Since a proprietary model-free Pavlovian learning system is retained, the phenomena of selective association remain well-explained, as do many of the phenomena of modularity. Evidence that basic emotions *can* respond quickly and prior to conscious evaluation is obviously not evidence that they *must* respond quickly and prior to conscious evaluations (or cortical visual processing, etc.). Moreover, being able to store information that other systems cannot access in the Pavlovian learning system and emotional appraisal system does not prevent the influence of other systems through the model-based Pavlovian learning system. Finally, opacity to conscious awareness is preserved, since higher cognitive systems are still not capable of accessing information from the system at levels of processing between the inputs and outputs.

However, it is fair to ask whether the evolution of an emotion like fear would predict its integration with a model-based Pavlovian learning system. There is at least one clear rationale for such a prediction. Griffiths points out that there is likely to be an evolutionary advantage for a bias in favor of false-positives over false-negatives, and there is a clear sense in which a model-based learning system supports this kind of bias: If some of the predictors of basic life problems are complex (e.g., observable signs of big cat predators could be different next to water versus in tall grass), then only an organism with model-based Pavlovian fear learning will be capable of noticing their predictive value. Such an organism would fare better than an organism without it. The latter organism will experience more false negatives (e.g., big cats not detected). It will not become afraid in situations where fear is a valuable response, and it will pay fitness costs for this inability.

Finally, according to this model, there is less reason to drive a sharp wedge between complex, culture-dependent fears (e.g., fear of flying in a plane) and phylogenetically ancient fears (e.g., fear of predators or heights). Model-based Pavlovian learning may very well explain why the shaking of a plane’s cabin can be acquired as a predictor of an unsafe situation, which triggers fear on that basis. If so, then perhaps many of the phenomena attributed to so-called higher cognitive emotions may in fact be explained by the same processes as basic emotions.[[13]](#footnote-14) In that case, these phenomena would obviously not arise from mechanisms that *require* higher cognitive inputs for their activation. Rather, they would arise from basic emotions. So, if I am right about the distinction between appraisal and memory, then the category of basic emotions may subsume a fairly wide range of folk emotion categories. In other words, the discontinuity hypothesis is undermined. There are fewer reasons to think that there is some theoretically interesting discontinuity between basic and higher cognitive emotions.

# References

Arcediano, Francisco, Helena Matute, and Ralph R Miller. 1997. “Blocking of Pavlovian Conditioning in Humans.” *LEARNING AND MOTIVATION*. Vol. 28.

Barrett, Lisa Feldman. 2017. *How Emotions Are Made: The Secret Life of the Brain.* *How Emotions Are Made: The Secret Life of the Brain.*

Chevalier-Skolnikoff, Suzanne. 2006. “Facial Expression of Emotion in Nonhuman Primates.” In *Darwin and Facial Expression: A Century of Research in Review*, 273.

Clark, Jason. 2013. “Integrating Basic and Higher-Cognitive Emotions within a Common Evolutionary Framework: Lessons from the Transformation of Primate Dominance into Human Pride.” *Philosophical Psychology* 26 (3): 437–60. https://doi.org/10.1080/09515089.2012.659168.

Daly, M, and M Wilson. 1988. *Homicide*.

Dayan, Peter, and Kent C. Berridge. 2014. “Model-Based and Model-Free Pavlovian Reward Learning: Revaluation, Revision, and Revelation.” *Cognitive, Affective and Behavioral Neuroscience*. Springer New York LLC. https://doi.org/10.3758/s13415-014-0277-8.

Ekman, Paul. 1971. “Universals and Cultural Differences in Facial Expressions of Emotion.” *Nebraska Symposium on Motivation* 19.

———. 1999. “Basic Emotions.” In *Handbook of Cognition and Emotion*, edited by T. Dalgleish and M. J. Power, 45–60. Chichester, UK: John Wiley and Sons. https://doi.org/10.1016/s0005-7967(00)00008-5.

———. 2005. “Basic Emotions.” In *Handbook of Cognition and Emotion*, 45–60. https://doi.org/10.1002/0470013494.ch3.

Ekman, Paul, and Daniel Cordaro. 2011. “What Is Meant by Calling Emotions Basic.” *Emotion Review* 3 (4): 364–70. https://doi.org/10.1177/1754073911410740.

Fodor, Jerry A. 1983. *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: The MIT Press. https://doi.org/10.1017/cbo9780511814273.046.

Griffiths, Paul E. 1997. *What Emotions Really Are: The Problem of Psychological Categories.* *The Philosophical Review*. Vol. 108. Chicago: University of Chicago Press.

———. 2013. “Current Emotion Research in Philosophy.” *Emotion Review* 5 (2): 215–22. https://doi.org/10.1177/1754073912468299.

Griffiths, Paul E, and Andrea Scarantino. 2009. “Emotions in the Wild: The Situated Perspective on Emotion.” In *The Cambridge Handbook of Situated Cognition*, 437–53.

Hamlin, J. Kiley. 2013. “Moral Judgment and Action in Preverbal Infants and Toddlers: Evidence for an Innate Moral Core.” *Current Directions in Psychological Science* 22 (3): 186–93. https://doi.org/10.1177/0963721412470687.

Izard, Carroll E. 2007. “Emotion Feelings Stem from Evolution and Neurobiological Development, Not From Conceptual Acts: Corrections for Barrett et Al. (2007).” *Perspectives on Psychological Science* 2 (4): 404–5. https://doi.org/10.1111/j.1745-6916.2007.00053.x.

LeDoux, Joseph E. 1993. “Emotional Networks in the Brain.” In *Handbook of Emotions*, edited by Michael Lewis and J. M. Haviland-Jones. New York: Guilford Press.

Matthen, Mohan. 1998. “Biological Universals and the Nature of Fear.” *The Journal of Philosophy* 95 (3): 105. https://doi.org/10.2307/2564712.

Ohman, Arne, and Susan Mineka. 2001. “Fears, Phobias, and Preparedness: Toward an Evolved Module of Fear and Fear Learning.” *Psychological Review* 108 (3): 483–522. https://doi.org/10.1037/0033-295X.108.3.483.

Prinz, JJ. 2004. *Gut Reactions: A Perceptual Theory of Emotion*.

Robinson, Mike J.F., and Kent C. Berridge. 2013. “Instant Transformation of Learned Repulsion into Motivational ‘Wanting.’” *Current Biology* 23 (4): 282–89. https://doi.org/10.1016/j.cub.2013.01.016.

Scarantino, A. 2015. “Basic Emotions, Psychological Construction, and the Problem of Variability.”

Seligman, Martin E.P. 1971. “Phobias and Preparedness.” *Behavior Therapy* 2 (3): 307–20. https://doi.org/10.1016/S0005-7894(71)80064-3.

Soares, Joaquim J. F., and Arne Öhman. 1993. “Backward Masking and Skin Conductance Responses after Conditioning to Nonfeared but Fear‐relevant Stimuli in Fearful Subjects.” *Psychophysiology* 30 (5): 460–66. https://doi.org/10.1111/j.1469-8986.1993.tb02069.x.

Wiegman, Isaac. 2020. “Emotional Actions Without Goals.” *Erkenntnis*. https://doi.org/10.1007/s10670-019-00200-8.

1. For example, Griffiths draws on evidence for homology of various facial expressions in support of his thesis (e.g., Chevalier-Skolnikoff 2006). [↑](#footnote-ref-2)
2. I say “produce” instead of “experience” because on Griffiths view, emotions are not simply feelings that we experience, but more like a naturally occurring syndrome, of which basic emotions are the putative cause. I side with Griffiths in thinking that feelings may be part of emotion syndromes, but that we should not assume that feelings are their essences. [↑](#footnote-ref-3)
3. For a discussion of his arguments concerning the output side, see Wiegman [RRR]. [↑](#footnote-ref-4)
4. Extensive work on observational conditioning of snake phobia in rhesus macaques provides further support for this conclusion (RRR). [↑](#footnote-ref-5)
5. While there are numerous studies that purport to find unconscious fear responses, there is some indication of publication bias, and one systematic review of this literature concludes that there is no convincing evidence of fear conditioning where subjects are unaware of contingencies between presentation of certain images and aversive stimuli such as electrical shocks (RRR). [↑](#footnote-ref-6)
6. Griffiths is somewhat inconsistent about this inference. Griffiths (Griffiths 2013) claims that it is uncontroversial “that there are some emotion episodes where an affect program response is triggered by a high-level appraisal process which makes use of concepts that could not plausibly be attributed to lower level processes (or to simpler organisms).” (p. 218) Yet, this claim is strikingly inconsistent with claims in his earlier work: that if Othello’s sexual jealousy were the output of a basic emotion, “he would have had to catch Desdemona in bed with Cassio, or at least have seen the handkerchief, before his jealousy was initiated.” (p. 117) Hard to see why Othello would have to see anything in particular if a basic emotion could be triggered by a high-level appraisal process (etc.). Maybe he would admit to getting this specific case wrong, but in that case, another problem arises. If it is uncontroversial that basic emotions can be triggered by high-level appraisal processes (etc.), then it is unclear why he would remain insistent upon the discontinuity of basic and higher cognitive emotions. My best guess is that he thinks there are phenomena of higher cognitive emotions that cannot be explained except by processes that *require* higher cognitive capacities for their activation or ongoing processing/production. In that case, we are owed some description of these phenomena together with an argument that they can only be explained in that particular way. There are descriptions for emotions such as shame or pride (for a review, see Clark RRR), which seem to invoke a full-fledged self-concept. However, I have yet to find such an account of higher cognitive emotions such as anger or fear, which require higher cognitive capacities for their activation or production. [↑](#footnote-ref-7)
7. Imagery can be understood as an offline activation of lower-level perceptual systems, which presumably *can* trigger the AAM by matching the low-level representations in its database. See, e.g., Griffiths 1997, p. 97 note. [↑](#footnote-ref-8)
8. Among other things, it is not always clear in these other cases whether this is because of the assumption that appraisal includes memory processes. [↑](#footnote-ref-9)
9. At least for some emotion categories. Shame and pride and other self-conscious emotions are possible exceptions to this rule. Cf. fn. 9 [↑](#footnote-ref-10)
10. The evidence for this claim is in the details of so-called “blocking” experiments (see, e.g., Arcediano, Matute, and Miller 1997). [↑](#footnote-ref-11)
11. This is because they appear to solve the “basic life problems” that these basic emotions were adapted to solve. Specifically, problems of avoiding danger, avoiding poisons, and dealing with conspecific challenges (Ekman 1999). Seligman’s work on prepared learning (RRR), to which Griffiths refers in discussing the modularity of basic emotions (RRR), encompasses standard examples of Pavlovian learning. [↑](#footnote-ref-12)
12. For subsequent work on model-based Pavlovian learning, see RRR and RRR. [↑](#footnote-ref-13)
13. This becomes much more plausible in light of more recent explications of that concept, some of which focus on the output side of basic emotions (Wiegman 2020; Scarantino 2015). [↑](#footnote-ref-14)