What Basic Emotions Really Are: Modularity, Motivation, and Behavioral Variability

Abstract

While there is ongoing debate about the existence of basic emotions and about their status as natural kinds, these debates usually carry on under the assumption that basic emotions are modular and therefore cannot account for behavioral variability in emotional situations. Moreover, both sides of the debate have assumed that these putative features of basic emotions distinguish them as products of evolution rather than products of culture and experience. I argue that these assumptions are unwarranted, that there is empirical evidence against them, and that evolutionary theory itself should not lead us to expect that behavioral invariability and modularity mark the distinction between evolved emotions and higher cognitive emotions. I further suggest that claims about behavioral invariability and modularity have functioned as defeasible conjectures aimed at helping test basic emotion theory. Finally, I draw out the implications of these claims for debates about the existence of basic emotions in humans.

*Keywords:* basic emotions; natural kinds; modularity; behavioral variability; cognitivism; evolutionary psychology

# Introduction

It is widely (though not universally) held among emotion theorists that there is a theoretically interesting distinction between basic and higher cognitive emotions (for a review, see e.g. Clark, 2009). On this picture, basic emotions are primarily structured by evolution whereas higher cognitive emotions are substantially structured by culture or individual experience (perhaps among other things). While there is ongoing debate about the existence of basic emotions and about their status as natural kinds (Barrett, 2017), these debates usually carry on under the assumption that basic emotions are modular in roughly Fodor’s (1983) sense and are therefore inflexible and unable to explain behavioral variability in emotional situations. Moreover, both sides of the debate have assumed that these features of basic emotions distinguish them as products of evolution rather than products of culture and experience. I aim to show that these assumptions are unwarranted, that there is empirical evidence against them, and that evolutionary theory itself should not lead us to expect that behavioral invariability and modularity mark the distinction between evolved emotions and higher cognitive emotions.

To be clear, my concern here is to question the theoretical assumptions that set the stage for empirical testing of theories about emotions. The larger part of this challenge consists in pointing out the actual existence of certain kinds of psychological states in other domains (e.g., the domain of appetites and animal emotions); ones that are products of evolution and modular in some respects, yet highly flexible in terms of their influence on behavior. The dominant theories of human emotions on offer today appear to elide the very possibility of human emotions that fit this description. Yet, if this kind of psychological state exists in the domain of appetites and animal emotions and if instances of this kind are evolutionary adaptations, then there may be similar psychological states in the domain of human emotions. Consequently, theories in this domain should not elide the possibility of emotions that are flexible, but modular adaptations. It is an entirely separate question whether or not such states actually exist and one that I do not take up here. However, I do point out some theoretical benefits along the way, particularly the possibility of explaining certain complex, pan-cultural behavioral phenomena.

# The Basics of Basic Emotions

So what are the core commitments of basic emotion theory (BET)? Basic emotions – including anger, fear, happiness, sadness, disgust, and surprise (for an extended list, see Ekman & Cordaro, 2011) – are supposed to cause species-typical behavioral syndromes in humans. These syndromes include involuntary facial expressions of emotion, physiological changes (e.g. in heart rate, blood pressure, and hormone levels), feelings, cognitive changes (e.g. effects on attention and memory), and changes in bodily posture (including bodily social displays and orienting responses). On this view, feelings are not the primary characteristic of emotions even though feelings may play a role in how people identify or think about their own emotions (whether basic or not).

According to BET, basic emotion syndromes have an evolutionary explanation and are underpinned by similar neural and psychological mechanisms. Perhaps the core theoretical commitment of BET is that each syndrome evolved to address basic life problems or adaptive problems (such as resource competition, avoidance of predators and avoidance of poisons and parasites). In other words, basic emotions are special purpose adaptations. Some of the basic life problems that shaped basic emotions are shared between human and non-human animals, and most basic life problems are thought to involve some aspect of sociality, which is thought to explain the importance of facial expressions as signals (Ekman, 1999; Shariff & Tracy, 2011).

Before I say more about BET, two observations are important for my purposes. First, in addition to their social/expressive functions (i.e., the putative role of facial expressions as signals), basic emotions such as fear and disgust also have distinct individualistic functions (e.g. avoiding bodily damage from falling from heights or avoiding contaminants in food and bodily wastes). Second, even if some basic emotions (e.g. fear, pleasure, disgust/distaste) can be shared between human and non-human animals, there is no reason to suppose that the sets of basic emotions in other animals are identical to the set in humans. Other lineages may have faced distinct basic life problems from our own, or distinct emotional responses may have evolved independently as solutions to the same basic life problem. So, if we identify a basic emotion in rats (as I do in section 4), we need not infer that humans share the same basic emotion.

Now, BET does not only include evolutionary hypotheses about certain emotions but also includes hypotheses about their proximate mechanisms: the subcomponents of basic emotions that explain each basic emotion syndrome. Specifically, the elicitation and production of basic emotion syndromes are supposed to be explained by *automatic appraisal mechanisms* and *affect programs*, respectively (Ekman, 1977, 1999). For instance, the affect program for anger is hypothesized to coordinate the various response components that help to solve the basic life problem of *conspecific challenge*. These components are thought to include facial expressions of anger, increased heart rate and blood pressure, increased blood flow to the upper extremities, and postural changes such as orienting toward the challenger, etc. Importantly, these facial expressions are complex and involve many muscles of the face, some of which are difficult to activate voluntarily (Ekman et al., 1980). The set of affect programs (for each basic emotion) is also supposed to explain the results of experiments that connect facial expressions of emotion (e.g. anger, fear, joy, etc.) with changes in physiological response components (Ekman et al., 1983; Levenson et al., 1990).

Under the assumption that affect programs produce innate, pan cultural and distinctive signals for each basic emotion, one might predict that people across cultures would be sensitive to differences between facial expressions and would be able to attribute these emotions correctly. Thus, affect programs *indirectly* explain the results of some cross cultural experiments on emotion recognition: ones that ask people to distinguish photographs of facial expressions of emotions, connect these expressions with emotion terms, or rate their appropriateness in response to vignettes (e.g. Ekman et al., 1987). To generalize, affect programs are introduced to explain the observed coordination of various response components, including facial expressions, and the cross-cultural production of these various syndromes. This in turn, would explain widespread recognition of facial expressions across cultures (at least, with the additional specification of capacities for emotion recognition).

## 2.1. The Invariability Assumption

Another feature of basic emotions is frequently added to this theoretical picture: what I call the *invariability assumption*. In its weaker form, this is the claim that basic emotion syndromes include defining behaviors, or behavioral *signatures* (L. F. Barrett, 2017), which always occur every time the emotion is elicited. For instance, Ekman and others claim that basic emotion responses *just are* short term, stereotyped responses (Ekman, 1977; Griffiths, 1997, p. 241) that include the symptoms listed above (i.e. involuntary facial expressions of emotion, physiological changes, feelings, effects on memory and attention, and changes in bodily posture). Some of the early experimental results (just described) *seem* to support this conclusion. Moreover, Ekman and colleagues claim that when basic emotions are elicited, their distinctive facial expressions are difficult to consciously suppress or control (Ekman & Friesen, 1969). It is even suggested that habitual suppression of facial expressions (perhaps because of acquisition of cultural “display rules”) cannot completely eliminate “leakage” or residual “microexpressions” of an emotional response (Friesen, 1973; Haggard & Isaacs, 1966). This further supports the *impression* that these facial expressions are emotion “signatures” (cf. Barrett, 2017, Chapter 3).

These data and conjectures lend support to an evolutionary explanation of basic emotions that is closely tied to the invariability assumption: regular co-occurrence of the emotion with functional behaviors (e.g., baring teeth in response to a challenger) leads to the cooption of the functional behavior as an evolved signal of the emotional state (Shariff & Tracy, 2011). This is called ritualization. If functional behaviors become reliable signals, then it is plausible that the signal will become a signature of the emotion. This ritualization hypothesis would be further supported by pan-cultural recognition of the facial expressions corresponding to basic emotions.[[1]](#footnote-2)

This *weak invariability assumption* is often strengthened by the assumption that the adaptive functions of basic emotions do not explain any other effects on behavior besides these behavioral signatures. This commitment is like a principle of *closure*: that basic emotion syndromes *only* include the symptoms listed above (i.e. involuntary facial expressions of emotion, physiological changes, feelings, effects on memory and attention, and changes in bodily posture) or alternatively that those symptoms exhaust the innate behavioral effects of basic emotions. To be clear, this assumption leaves room for *acquired* responses to emotional situations; ones that are automatically initiated, but not controlled, by an affect program. For example, if someone has experienced predominantly bad outcomes when she gets angry, she may acquire an automatic tendency to try and escape situations in which she experiences anger. This kind of coping behavior may even be directly triggered by the affect program for anger, but her response is clearly not part of the species typical syndrome of basic emotional anger, nor does the affect program plausibly control how such a behavior unfolds once it is triggered.

This *strong invariability* assumption is subtly connected with a range of other commitments of BET. First, if signatures evolved via ritualization, then they are plausibly innate.[[2]](#footnote-3) Moreover, if we are forced to choose whether or not a given emotional behavior is part of the innate, basic emotion signature, then the answer is clear: an involuntary facial expression may be part of the innate package of anger, but punching someone in the face is surely not (Ekman, 1977). Second, Izard assumes that, by definition, basic emotions lack dynamic interaction with cognition (Izard, 2007, p. 265). If this is true, then basic emotions would have limited influence on behavior. Interaction with cognition seems necessary to organize complex or strategic sequences of behavior, such as fighting behavior. (I discuss both of these rationales in greater detail below.)

This whole picture fits nicely with a kind of dual process approach to psychological states, whereby some states are deliberate, controlled, and integrated with conscious beliefs and desires and others are modular or automatic, and thereby insulated from conscious beliefs and desires. For example, Griffiths draws on Fodor’s (Fodor, 1983) account of modularity to describe the emotion systems identified by Ekman (Ekman, 1980, p. 151) and others (emotion & 1984, n.d.; i.e., Zajonc, 1980):

The systems [sic] operation is *mandatory* (Zajonc’s term is “inescapable). People often respond with fear or anger to a given stimulus whether they choose to or not. The system is largely *opaque* to our central cognitive processes. People are aware of its outputs, which are the emotional responses themselves, but not aware of the processes that lead to them. Finally, and most importantly, the system is *informationally encapsulated*. It 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 do not get taken into account when the system is deciding upon a response. Informational encapsulation is one of the most important features of modular systems, since it captures what is meant by their “separation” from other processes, such as those leading to longer-term, planned action. (Griffiths, 1997, p. 93)

Furthermore, Ekman, Griffiths, Zajonc, and many others claim that affect programs are *fast* (i.e., the rapid onset and short duration of their symptoms). It could also be argued that the outputs of affect programs are nonconceptual, which would imply *shallow outputs*.

Griffiths (1997, pp. 88–91) also argues that the input systems for basic emotions, automatic appraisal mechanisms, are *domain specific*, meaning that their operations are sensitive to a restricted range of inputs. Whereas domain specificity is a central feature of modularity (cf. Coltheart, 1999),[[3]](#footnote-4) it will not be a central focus here. This is because my interest is in the variability of emotional behaviors, and domain specificity has not played a significant role in arguments concerning variability. By contrast with encapsulation and opacity, domain specificity does not say much about the processes internal to a system. Thus, it has limited relevance to questions about how an emotion system arrives at its behavioral output and to the features of those outputs. So, in what follows, I will say that a capacity is *modular* if it is both encapsulated and opaque in relation to other systems, primarily the central processes that mediate beliefs, desires, and long term planned action.[[4]](#footnote-5) Moreover, I will assume (as does Fodor) that cognitive integration and modularity are matters of degree and are inversely proportional. Thus, to say that something lacks substantial cognitive integration is to say that it is substantially (though perhaps not entirely) modular.

To sum up, if basic emotions are to be forced into one or the other of the slots in a dual process framework (central/deliberate as opposed to modular/automatic), it is fairly obvious which slot gives the least resistance. On this basis, Griffiths, Ekman, and Izard all seem to suppose that basic emotions have very limited influence on longer term behavior. In other words, they think the defining symptoms of affect programs (e.g., facial expressions and physiological responses) exhaust their innate control over behavior. This is the strong invariability assumption. By contrast, the weak invariability assumption states only that emotions have defining symptoms and behaviors that are present in any manifestation of the emotion, while dropping the closure assumption (i.e., that the defining behaviors are the only aspects of emotional behavior that are controlled by affect programs). Table 1 summarizes the two invariability assumptions and the evidence given to support them.

|  |  |
| --- | --- |
| **Assumption** | **Support** |
| Weak Invariability | Description of emotional response |
| Ritualization explanation |
| Strong Invariability  (Weak Invariability + Closure) | Modularity, automaticity |
| Innateness of response |

Table 1 Reasons supporting assumptions about the behavioral invariability of basic emotions.

These assumptions are connected with more recent additions to basic emotion theory: it is natural to assume that if basic emotion syndromes are strongly invariable and controlled by the central nervous system, there would be distinct and dedicated neural circuits responsible for the implementation of each emotion, which are activated whenever the emotion is elicited (e.g., Panksepp, 2004; Tracy & Randles, 2011). Some neuroscientific data has even been interpreted as evidence of this kind of localization (e.g., Calder, 2003; for a recent review, see also Celeghin et al., 2017).[[5]](#footnote-6)

Accordingly, assumptions about invariability played out well for BET in its early days. Much of the research mentioned above as evidence for the theory appears to be evidence for the invariability of basic emotions: eliciting basic emotions appears to activate specific neural regions, eliciting one response component appears to trigger others, involuntary facial expressions are difficult to control, and facial expressions of emotion are thought to be universally recognizable because they reliably co-occur with the emotion.

Nevertheless, much of this research has not held up under close scrutiny. For example, research suggests that basic emotion facial expressions are not recognized by some cultures (Gendron et al., 2014); recent meta-analyses have shown that the observed coherence of emotional response components is actually quite low in many studies (Reisenzein et al., 2013) and that there is no one-to-one correspondence of neural regions with basic emotions (Lindquist et al., 2012).[[6]](#footnote-7) These data and meta-analyses are often taken as a straightforward falsification of BET, and evolved emotional states more generally. Thus, these invariability assumptions have played a key role in empirical tests of BET, as both proponents and critics of the theory have used this assumption as a way of confirming or falsifying the theory. So, if it were to turn out that invariability is nonessential to basic emotions, then the fault lines of this debate should undergo a momentous shift.

# Weak Evidence for Strong Invariability

And that is the view I defend here: the fault lines should shift because there is actually no good theoretical reason to suppose that the entire class of basic emotions will be invariable in any of their empirical manifestations, any more than are hunger, thirst, and other appetites. Notice that this is not an empirical claim to be substantiated by empirical evidence. Rather, it is a claim about evolutionary theories of the emotions and what predictions those theories ought to make. The question is whether invariability assumptions of the sort just discussed should be used to make empirical predictions about evolved emotions. While I think neither the strong nor weak invariability assumptions should play this role, the focus of sections 3 and 4 is on the stronger version of invariability and in particular the closure assumption: the claim that the defining symptoms of basic emotions exhaust their influence on behavior. If this assumption is false, then it is possible that basic emotions control a range of flexible behaviors (e.g., anger may cause fighting behaviors). As I suggested above, the only evidence provided for this assumption is that basic emotions are innate and modular. In this section, I argue that this is weak evidence for strong invariability. In section 4, I present positive evidence against the strong invariability of animal emotions.

## 3.1. Innateness does not imply closure

First, consider the supposed connection between innateness and invariability. Ekman makes an inference from innateness to closure in the following passage:

Somewhat longer and more elaborated [than facial and physiological responses produced by basic emotions] are the coping behaviors directed at whatever has set off the emotion. Included would be fighting, fleeing, denying, apologizing, etc…Through experience, with sufficient time and learning, habits become established for how to cope with each emotion. I do not believe that such coping behaviors are part of the *given* affect program…Memories, images, expectations associated with one or another emotion are, *like coping, not given but acquired*… (Ekman, 1977, pp. 56–57 emphasis mine)

Undoubtedly, there are many behaviors that can occur when a person encounters a basic life problem. Some will seem incongruous with the central direction of the emotional response itself. Anger, we might suppose, predisposes us to confrontation and attack, so one might be skeptical that a person’s anger could be the motivation for stomping out of a room, effectively fleeing rather than confronting the situation. Ekman takes this skepticism even farther by denying that any “longer and more elaborated” behaviors are controlled by basic emotions (the closure assumption), regardless of how “congruent” they may be with a basic emotion.[[7]](#footnote-8)

Ekman rules this out because the flexible behaviors associated with emotions appear to be acquired rather than given. Ekman categorizes all such flexible behaviors as “coping behaviors,” with which a person learns to cope with the emotional situations that elicit a particular emotion. If we follow Ekman in distinguishing between coping behaviors on the one hand and “The initial skeletal muscle response directed by the affect program” (Ekman 1977, p. 72) on the other, then the coping behaviors seem unlikely to be innate: “When angry, our likelihood of fighting or scratching our face, depends upon what we have learned about how to deal with the particular kind of anger elicitor.” (Ekman 1977, p. 72) Therefore, according to Ekman, the only behaviors that the affect program controls *as part of its innate endowments* are short-term, stereotyped responses.

This conclusion is the culmination of a faulty line of questioning. In the passage above, Ekman is answering the question of whether *specific behaviors* are innate or acquired. For example, if someone fights rather than scratching their face when they become angry, we might ask whether this response is innate or acquired. But this way of approaching the question elides the possibility that there is an innate behavior *tendency* which helps to select from acquired behaviors when angry. Such a tendency would be revealed by the fact that actions under the influence of anger lead to similar outcomes.[[8]](#footnote-9) On this alternative understanding, anger may participate in selecting longer courses of action involving say, retaliation or revenge. Such behaviors would be distinct from coping behaviors that might be indirectly related to anger. Consider again someone who tends to avoid infuriating situations. A coping behavior such as that may be triggered by anger, but anger does not necessarily determine how the response unfolds, as it would for actions of retaliation and revenge. If one avoids staying angry, the avoidance of it is plausibly controlled by other cognitive processes.

While Ekman allows that anger might cause coping behaviors, he simply leaves out the possibility that anger might select or control actions of retaliation or revenge (etc.). If the only possible descriptors of behavioral effects are “acquired behavior” or “innate behavior” without the addition of other possible categories (e.g. “acquired behavior *tendency*” or “innate behavior *tendency*”) then of course, fist fighting looks acquired and only something like “a slight movement forward” will look innate. Under these assumptions, it follows that the innate endowments of a basic emotion could only produce the latter behavioral effects, rather than the former.

To see the problem with this elision and the resulting inference, consider an analogy with hunger. We might guess that hunger has as reasonable a claim to innateness as any other affective or motivational state. Nevertheless, we suppose that when an organism is hungry, this can lead to a wide range of behaviors, which usually end in food consumption. We might capture the same thought by saying that hunger is an innate motivational state or behavior tendency, and one that is present to ensure that organisms will seek food on an empty stomach (among other things). The suggestion concerning innateness is not that these food-seeking behaviors are preprogrammed or innately packaged with the hunger system. It is rather that hunger includes an innate disposition to select behaviors that lead to food consumption by interacting with other cognitive systems (as I argue in more detail in section 3.2.2). One way that hunger could accomplish this is to make certain courses of action appear attractive (e.g. ones that predictably lead to food consumption), making the organism more likely to perform those actions.

If this is a reasonable hypothesis to consider about hunger, why should we not consider a similar hypothesis concerning anger? For instance, one could hypothesize that anger includes an innate motivational state that causes the organism to remove obstacles to their goals through aggressive means.[[9]](#footnote-10) As such, it makes attractive those actions that move against obstacles. For Ekman, this hypothesis is not even on the table for examination, but it clearly should be. Otherwise, theories of evolved emotions would have nothing to offer when it comes to explaining pan-cultural patterns of revenge and retaliation (*inter alia*).[[10]](#footnote-11)

To clarify, the question I am interested in here is not whether anger includes inbuilt, automatic action patterns, as if an angry jab to an opponent’s nose were written into human DNA. The question concerns whether anger can motivate an agent toward a specific, biologically predisposed end state (say, moving against an obstacle) by inclining an agent toward a range of means, some of which may be novel or learned behaviors. These means could include a jab to the nose, a withering glare, a harsh word, or a revenge plot.[[11]](#footnote-12) The inference from innateness only seems plausible if we rule out this possibility at the outset.

There is another important reason why innate responses tend to be viewed as inflexible, which is that they are distinct from goal-directed behaviors (e.g., Dickinson, 1994). To see this, consider the “Pavlovian responses” studied in animal behavior and decision science, which are basically identical to the basic emotional responses that Ekman describes. LeDoux and Daw define these responses as “…innate behaviours that come under the control of novel stimuli through associative learning…” (p.1) Some of these Pavlovian responses are mere reflexes, or “…stimulus–response connections that are automatically and rapidly triggered by an innately programmed stimulus (known as an unconditioned stimulus (US))…” (p.2). Other responses, like freezing and flight in response to threat stimuli as well as involuntary facial expressions, are triggered in the same way but are “less directly related to stimulus intensity, are slower in onset and typically involve complex patterns of coordinated responses…” (p. 3) LeDoux and Daw call the latter category of responses “fixed reaction patterns” (instead of calling them “fixed *action* patterns” as was common in the ethological tradition) “…to convey their automatic nature [and] to respect the recent view that the term ‘actions’ should be reserved for emitted, flexible behaviours rather than elicited, fixed ones…” (p. 3). Moreover, the “recent view” they refer to is the influential work of Dickinson (1994) in distinguishing flexible goal-directed actions from other species of inflexible behaviors (i.e., habits and Pavlovian responses). In sum, even though flight and freezing are complex and coordinated, LeDoux and Daw label them as inflexible and fixed. While they do not use the word “invariable,” one could easily understand (or perhaps misunderstand) them to be making an inference from innateness to invariability.

Similar to Ekman’s inference, this one would only make sense because of an important elision. If the only categories we have are flexible as opposed to fixed behaviors, then flight and freezing are going to look fixed and inflexible. However, this elides the possibility that some fixed reaction patterns are also flexible, though flexible in a different manner than goal-directed behaviors.[[12]](#footnote-13) This possibility may seem incoherent, yet the behavioral pattern of flight (from a predator) is actually quite flexible: even though LeDoux and Daw call it a “fixed reaction pattern,” they admit that it is a complex response. Flexibility is part of that complexity: to successfully avoid predation, flight cannot be a rigid sequence of muscle movements like an involuntary facial expression. Such a predictable escape sequence could easily be exploited by predators. Rather, flight must be dynamically adjusted to maintain a course away from the predator and toward relative safety. Moreover, it must be sensitive to ongoing changes in the predator or predators’ locations and capable of surmounting obstacles. To accomplish all of this, flight must interact with systems that control a wider range of the organism’s behavioral repertoire, and in this way, it is fundamentally distinct from an inflexible reaction like freezing. So, flight is highly flexible, even though it is a “fixed reaction pattern,” which again, merely means that it is elicited by unconditioned stimuli and their predictive cues. As LeDoux and Daw make clear, there should be little doubt that the elicitation of this “fixed reaction pattern” is importantly distinct from the emitting of goal-directed behaviors: by contrast with Pavlovian responses, goal-directed responses have been shown to be remarkably sensitive to the current value of the goal and the goal’s contingency on an instrumental response. Nevertheless, we clearly cannot infer from *this* difference that flight is as inflexible a response as freezing. In the case of flight, in/flexibility in relation to its *elicitors* does nothing to negate the flexibility and variability of the *response* once it is underway.

## 3.2. Modularity does not imply closure

Now consider the inference from modularity to closure in greater detail. Izard argues that emotions are modular by pointing out non-cognitive causes of emotions:

Three other types of evidence suggest that [basic] emotion processes can operate independently of cognition. Emotions have been induced by unanticipated pain…, manipulation of facial expressions…, and changing the temperature of cerebral blood… In all these conditions the immediate cause of the emotion was noncognitive. (Izard, 1992, p. 563, 2007)

The idea is that basic emotions can operate independently of cognition because they can be triggered by low-level stimuli. This suggests that emotional states are opaque, since they respond to (and also store) information that is not integrated with (or filtered through) other cognitive processes. In addition, Ekman claims to have observed involuntary “micro-expressions” of basic emotions that “leak” through, despite attempts to inhibit them (Ekman & Friesen, 1969).[[13]](#footnote-14) If true, this suggests that basic emotions are mandatory and also encapsulated, in the sense that their operations are not influenced by information outside the system (or by information from central processes in the case of cognitive penetrability). Overall, basic emotions appear to be modular in some respects.

Griffiths is one of the few who explicitly draws out the implications of this kind of modularity. As discussed above (in section 2.1), he argues that basic emotions are opaque and informationally encapsulated. Since they have these and other marks of modularity, Griffiths then claims that basic emotions have “limited involvement” with higher cognitive processes, which are “…the processes in which people use the information of the sort they verbally assent to (traditional beliefs) and the goals they can be brought to recognize (traditional desires) to guide relatively long-term action and to solve theoretical problems” (Griffiths, 1997, p. 92). From this, Griffiths appears to draw a broader conclusion: that basic emotions are not “flexible [or] integrated with long-term, planned action” and are instead “restricted to short-term, stereotyped responses” (Griffiths, 1997, p. 241). In other words, Griffiths concludes that the defining symptoms of basic emotions (e.g., physiological arousal and involuntary facial expressions of emotion) exhaust the influence of basic emotions on behavior (the closure assumption).

The problem with this inference is that a system can be modular in one respect (e.g., its control over facial expressions) while also controlling highly flexible behaviors. To see this, consider again the analogy with hunger. Kent Berridge’s well known work on food reward (e.g. Berridge, 1996; Castro et al., 2015; Seeley & Berridge, 2015) casts hunger and other appetites such as thirst and salt appetite as homeostatic signals that systematically influence two distinct systems of affect and motivation. On this model, hunger is modular in many respects. One of its subsystems, the affective “liking” system, organizes fixed affective elements such as sensory pleasure and displeasure and stereotyped facial expressions thereof. Moreover, hunger appears to be modular in exactly the same ways as basic emotions. Like basic emotions, hunger is triggered by low-level stimuli (e.g. low-level detection of changes in blood sugar); and it can store information (e.g. about which foods are more calorically dense) that other systems cannot directly access (cf. Goldstone et al., 2009). Additionally, everyone has fairly direct evidence for the encapsulation of hunger: when one feels hungry, one cannot interfere with the feeling of hunger by merely thinking about it (e.g., by noticing that the amount of energy one’s body has stored in fat deposits is more than enough to sustain oneself).

Yet, hunger’s influence on behavior depends radically on context. In this respect hunger is akin to desire. Just as a desire for beer can produce radically different behaviors depending on context (e.g. walking to the fridge, driving to a convenience store, or even drinking a cider as a substitute), so can hunger. Hunger can even produce novel behaviors, as when rodents take previously unavailable “short cuts” to get to a food box in a maze (Olton, 1979; Tolman, 1948). This relationship with novel behavior is observed quite generally in research on behavioral innovation in animals. According to a recent review, “in most experimental tasks used to date, innovation is driven by hunger” (Tebbich et al., 2016, p. 9).[[14]](#footnote-15)

This should be little mystery given the importance of food-seeking for survival and reproduction. When an organism is hungry, its food-seeking behaviors need to be informed by its ability to register many different contextual cues. For one, food sources are not always easy to track down. In many cases, the food sources are themselves concerned with their own survival and thus have good reason to make themselves difficult for other organisms to kill and eat.[[15]](#footnote-16) In other cases, food sources vary across many dimensions to which a forager should be sensitive, including density of food, distance from home, and exposure to predators or parasites, each of which may change over time.[[16]](#footnote-17) This requires food-seeking behaviors to be highly flexible across these various changes in context. If parasite density increases in an area that has the highest relative density of food, then an organism may be better off if it avoids such an area in favor of another foraging ground. So, it should be no surprise that hunger can control behaviors that vary widely according to context.

Since hunger and basic emotions are modular in similar ways, observed forms of modularity in basic emotions cannot be evidence of closure. This is because the analogous closure assumption for hunger does not seem to follow from the kind of modularity observed in basic emotions. Quite the opposite: in spite of being modular in exactly the same respects, hunger has effects on behavior that transcend its more rigid symptoms (e.g., involuntary facial expressions). Why is it possible for a motivational system to have this kind of influence on behavior despite being modular in some respects? There are two reasons, which apply equally to appetites and basic emotions.

### 3.2.1. Reason 1: Modularity of information and motivation are distinct

First, the realm of the cognitive picks out a broad range of internal states that function as causal intermediates between stimulus and response, perception and action. Cognitive states so understood include not only informational states (such as beliefs) but also motivational states (such as desires). Moreover, questions about cognitive integration may be asked about either informational or motivational states. Whereas informational modularity concerns encapsulation and opacity *with respect to representing facts*, motivational modularity would concern encapsulation and opacity *with respect to selecting behavior* (or perhaps *changing facts*).[[17]](#footnote-18) The informational modularity of hunger (e.g., that it responds to noncognitive inputs and processes them independently of central processes) does not obviously conflict with it being motivationally integrated (i.e., capable of interacting with central processes to select different behaviors across different contexts). Wholesale modularity (of informational *and* motivational states), therefore, does not obviously follow from informational modularity. If this is correct, then having noncognitive inputs (etc.) is not a reason to think that emotions on the whole operate independently of cognition. They might very well be integrated with cognition as motivational states.

So, can basic emotions really interact with cognition to select behavior? One might suspect that straightforward empirical descriptions of basic emotions close off this possibility, but this suspicion is faulty. If Ekman and his collaborators are right (Ekman et al., 1980; Ekman & Friesen, 1969), then affect programs are opaque and encapsulated with respect to behavior selection, but this only concerns the expressive function of basic emotions. This does not show that basic emotions cannot interact with cognition to select adaptive behaviors independent of their signaling/expressive function. For all the evidence so far presented, it is still possible that basic emotions have other, non-modular influences on behavior.

### 3.2.2. Reason 2: Basic emotions may have both modular and non-modular influences on behavior

In other words, it is possible for motivational systems to have multiple channels of influence over behavior, some modular and others nonmodular. Hunger appears to be a prime example of such a system. As mentioned above, work on food reward suggests at least two subsystems controlled by homeostatic signals (see figure 1). One is for “liking” and the other for “wanting.” Like basic emotions, the liking system may very well be modular with respect to behavior selection. That is, facial signals of sensory pleasure may be difficult to suppress or control. Nevertheless, it would not follow from this that hunger (or any other homeostatic signal) is modular wholesale.

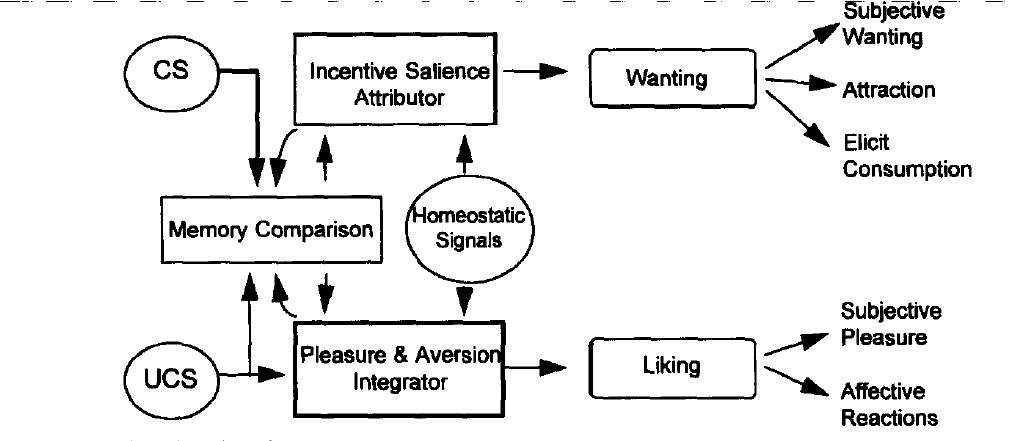


Figure 1 Berridge’s model of food reward. Hunger and other appetites are homeostatic signals that systematically influence an affective (“liking”) system and a motivational (“wanting”) system. The latter controls “attraction,” which flexibly guides behavior to some end. From Berridge (1996), p. 11. © Elsevier. Originally published in Neuroscience and Biobehavioral Reviews.

Quite the opposite: hunger is plausibly integrated with cognition via the “wanting” system. Above, I argued that the “wanting” system produces flexible behavior. At this point, we need to ask whether hunger could produce this kind of behavioral flexibility if it were completely modular. It clearly could not: hunger selects from a wider range of an organism’s behavioral repertoire than a module could plausibly store. So, it is unlikely to be both opaque and encapsulated *with respect to behavior selection* (even if the “liking” aspect is modular in this way). If wanting were fully modular in this respect, then it would almost certainly be unable to generate the novel food-seeking behaviors we observe, nor would it reliably solve the evolutionary problems surrounding food-seeking (described above).

This raises the question of whether any basic emotions are likely to have the same kind of motivational structure as hunger. This remains an open question. The evidence so far only suggests that basic emotions (if they exist) are modular with respect to *some* of their putative functions: preparatory functions and signaling functions.[[18]](#footnote-19) Nevertheless, this does not show that basic emotions are modular with respect to *all* of their functions. Which other functions might basic emotions include? If we are sticking with the central theoretical commitment of BET, then the answer is whichever functions are required to solve a given basic life problem.

Might one of these functions require the selection of different, contextually appropriate behaviors across different scenarios? I have seen no evidence against this possibility. Much of the evidence that appears to contradict it vanishes when we appreciate the moral of this section: we cannot answer questions about the cognitive integration of a multi-component system by pointing at its modularity with respect to a subset of its components and their sub-functions (e.g., the expressive function of hunger’s “liking” subsystem). Modularity of some components may be required to solve the basic life problem, whereas cognitive integration may be required for other components.

## 3.3 Theoretical conservatism does not support closure

Now consider one final line of support for the strong invariability assumption. Paul Griffiths suggests that dropping this and other assumptions about basic emotions would widen the reference class of basic emotions to include a motley of psychological entities.

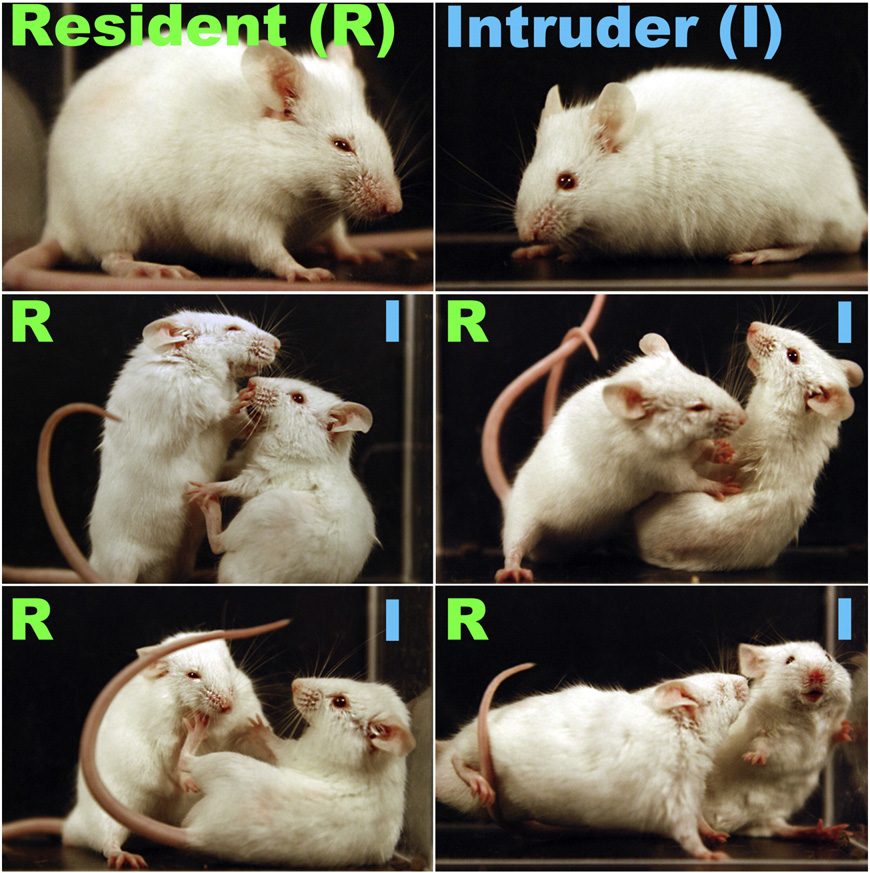
Steven Gaulin and Donald McBurney argue that it is inappropriate to demand that an emotion have a distinctive facial expression, since it may be more adaptive to keep emotions secret… Finally, they urge that the recognition of new emotional adaptations should not be prevented by the inability of current measurement techniques to identify any distinctive physiology associated with that adaptation (Gaulin & McBurney, 2001: 264-7)… these arguments threaten to extend the meaning of ‘basic emotion’ to cover just about any phenomenon in the general domain of motivation and emotion for which a plausible evolutionary rationale can be suggested. Ekman’s approach has the advantage that it identifies a range of broadly comparable and individually well-characterized psychological states… [The] methodological value of a list of basic emotions is to have a list of states of more or less *the same kind*, so that we can look for psychological and neurological principles about states of that kind. (Griffiths, 2003, p. 48)

The conservative impulse here is to set apart the phenomena discovered by Ekman, Friesen, Izard, and others and to thereby avoid running together psychological states that may differ substantially from one another.[[19]](#footnote-20) Yet, as a guide for research, this conservative impulse would almost certainly lead researchers to ignore or dismiss the possibility of cognitively integrated motivational states associated with basic emotions (e.g., of the sort that could explain highly variable but pan-cultural patterns of revenge, or avoidance of certain disease vectors, or patterns of response to dangerous animals like snakes, spiders, and large cats). Moreover, this line of reasoning only justifies invariability assumptions if behavioral invariability really is a “well-characterized” feature of basic emotions. The arguments above suggest that it is not. If we were to follow Griffiths’ approach regardless, then we would be making an educated guess about the nature of entities that cause the specific set of phenomena observed by Ekman and others (e.g. involving involuntary facial expressions and physiological responses): that these states are all behaviorally invariable in the strong sense. Nevertheless, this educated guess should be jettisoned if there is even modest evidence against it.

# Modest Evidence against Strong Invariability

In fact, when we consider basic emotional responses of some nonhuman animals, we do find some evidence against behavioral invariability and modest evidence for cognitive integration. To see this, consider the instinctive patterns of territorial behavior in rodents. These behaviors have been investigated in great detail using a resident-intruder experimental paradigm (for an overview, see Adams, 1980, 2006; D. C. Blanchard & Blanchard, 1984, 2003) in which resident (who have occupied a cage or colony for a few weeks) will attack unfamiliar male intruders introduced into their cage. The attacks of the resident and the defensive maneuvers of the intruder comprise sets of stereotyped behaviors. Each attack behavior of the resident is paired with a matching defensive maneuver of the intruder (see figure 2). The resident adopts a set of stereotyped postures and attacks aimed at biting the dorsal surfaces of the intruder. On the other hand, the intruder adopts a distinctive set of stereotyped behaviors aimed at avoiding or blocking the resident’s attempts to bite its back.

While these behaviors are certainly stereotyped, they are not brittle or reflexive. For instance, attacks of residents vary depending on the defensive strategy adopted by the intruder, and they seem to be governed by a motive to approach and attack that persists the entire time that the intruder is present. By contrast, the intruder rodent’s whole suite of behaviors seems to be governed by a persistent motive to escape and avoid.



**Figure 2** Confrontation and avoidance behaviors (e.g. facial expressions, postures and maneuvers) of resident and intruder mice (respectively). From Defensor and Corley (2012), p. 683. © Elsevier. Originally published in Physiology and Behavior.

What scientists have discovered about these behaviors (the flexibility of these behaviors and their coherent aims) indicates that they are produced by two separate motivational systems, what I call the *confrontation* and *avoidance* systems (Wiegman, 2016).[[20]](#footnote-21) The confrontation system is tuned to bring about a specific end state, repeated back-biting. Moreover, this motive does not depend on learning: rats which have been socially isolated from birth will still attempt to bite the back of an intruder (Eibl-Eibesfeldt, 1961). So far, the focus has been on cases in which a given rodent is purely motivated by confrontation or avoidance, but aggressive encounters in the wild usually involve a mix of offensive and defensive postures. This suggests that these motivational systems can be activated simultaneously or in close succession to produce mixed patterns of behavior.

These systems have many of the characteristics of affect programs in humans. They are posited to explain a coordinated suite of behaviors and physiological changes that probably include facial expressions, cardiovascular changes, and endocrine responses (Defensor et al., 2012; Fokkema et al., 1995). Moreover, these systems are tailored to solve two distinct basic life problems. Specifically, the confrontation system solves the problem of defending territories from other males for breeding purposes (and without fatally injuring kin in the process)[[21]](#footnote-22), whereas the avoidance system solves the problem of avoiding occupied territories and failing that, defending against the attacks of residents. Thus, if we have any reasons to postulate basic emotions in humans, we have the same reasons to postulate these basic emotions in rodent.[[22]](#footnote-23) The behavioral and motivational distinctions between the two (distinct interaction strategies for each role and distinct motivations for each role) are sufficient to distinguish them as two separate behavior syndromes that serve two very different purposes. Let us suppose then that the confrontation and avoidance systems are basic emotions in rodents.[[23]](#footnote-24)

Importantly for my purposes, the presence of an intruder can produce highly flexible and novel behaviors in the resident rat under certain conditions. In the bound-intruder task, an intruder is tied down on a Plexiglas plate with only its ventral surfaces (belly-side) exposed and placed in the cage of a resident, so that the resident cannot easily bite the back of the intruder. As a result, the resident will sometimes bite at the bands that tie down the intruder or dig under the intruder so that the resident can bite the intruder’s back (Blanchard et al., 1977). In contrast, none of these behaviors are adopted when the intruder is tied down with his back exposed.

These instrumental behaviors are clearly not stereotyped forms of attack, rather they are forms of flexible behavior adjustment to achieve the aim of biting the intruder’s back. In other words, the same end (back biting) can be achieved by several, novel means (digging, biting the bands, etc.); the confrontation system seems to include a motivational state that interacts with other systems to select from a range of behaviors instrumental for back-biting. In other words, the confrontation system is not modular with respect to behavior selection. This is something we would actually expect if this system is a solution to the basic life problem of defending a territory from intruders: flexibility is required to successfully repel an intruder because it is not in the intruder’s best interest to be repelled easily or to act predictably. For instance, the intruder would be sure to fare poorly if it acted in a way that accommodates the attacks of the resident. So, a single rigid movement or even a whole suite of rigid movements on the part of the resident would not tend to be successful against the most likely strategy of the intruder.[[24]](#footnote-25) It is more adaptive to have a flexible motivational state that leads to repeated back biting across a wide range of strategies or postures that the intruder might adopt. Rather than leading only to inflexible, stereotyped responses, it appears that a solution to this basic life problem demands some degree of integration with respect to action selection.

# The Architecture of Basic Emotions

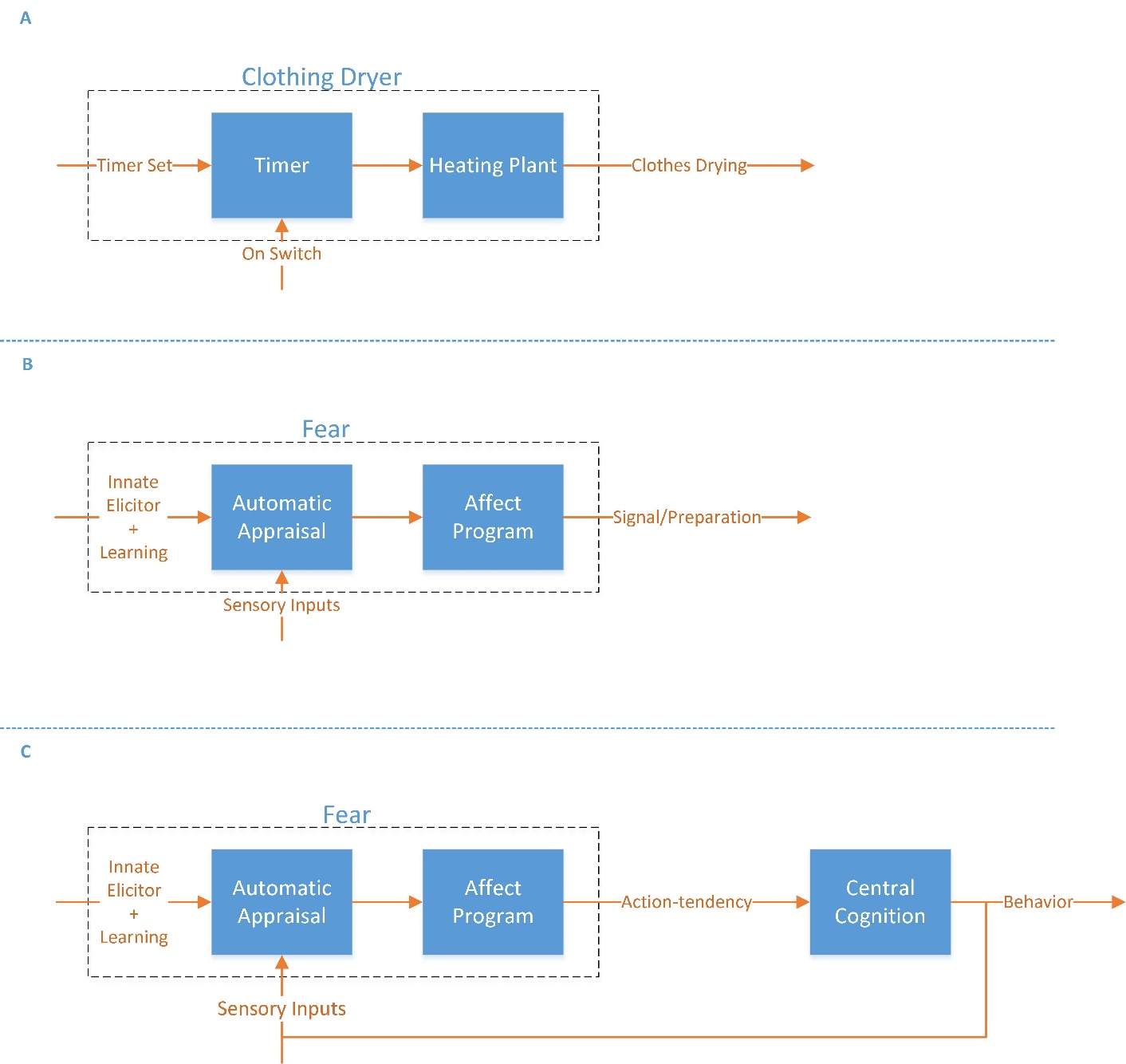
To some, this may seem like a dramatic revision to basic emotion theory, but it is actually quite modest. We can see this by considering a systems theoretic sketch of the traditional theory and the kind of modifications I propose. The distinction between the old and new version of basic emotions maps nicely onto the difference between open loop and closed loop systems in systems theory. An example of a closed loop system is clothing dryer that operates on a timer (see Figure 3A). The on switch starts the dryer which dries the clothing until the time goes off and is thus insensitive to how dry the clothing gets. According to the old conception of basic emotions (as depicted in Figure 3B), fear is like the clothes dryer, in that when triggered, it directly brings about its influence on behavior, relatively insensitive to feedback from the environment or whether it has achieved its function. On my proposed conception of basic emotions, fear is instead integrated with central processing (though it may also include some open loop signals, not depicted in Figure 3C) and is closed loop, receiving dynamic updates from the environment that modulate its activities to flexibly achieve its function (as when fleeing a predator). Even if a more phylogenetically ancient version of fear was open loop, it is easy to see that with very minor modifications to its structure, its behavioral outputs could be coopted and modified to serve as motivational inputs (or perhaps, action-tendencies in the sense of Frijda, 1986, p. 70) to cognition in a way that allows it to close the loop with the environment. 

Figure 3 Systems theoretic sketches of (A) an open loop clothing dryer, (B) basic emotion fear as a closed loop response, and (C) a proposed modification of basic emotion theory on which fear is closed loop (at least in relation to some of its outputs). From Wiegman (2020). © Springer. Originally published in Erkenntnis.

It may even be possible to chart this kind of transition over the course of develepment, as the outputs of an emotion become more flexible over time. For example, angry behaviors in small children are fairly stereotyped, stiffening limbs and arching the back is quite commonly observed in tantrums along with a number of other symptoms such as crying or hitting (Potegal & Davidson, 2003). Moreover, angry tantrums carry on to completion in a way that is fairly insensitive to input from the environment, even when they consistently do nothing to advance the child’s goals (aggressive or otherwise). By contrast, in adults angry aggression has been measured by a much wider range of behaviors that are much more sensitive to background knowledge about how to hurt other people: administering electrical shocks or uncomfortable air blasts to another person, writing negative performance evaluations, or even putting spicy salsa on a cracker that someone else has to eat (see, e.g., DeWall et al., 2007).

# Implications for Basic Emotion Theory

If basic emotion can control the latter kind of actions (via their action-tendencies and ongoing feedback from the environment), this changes the background assumptions of an ongoing debate concerning the existence of basic emotions in humans because it dramatically changes what should count as evidence for their existence.

In the past, this debate has carried on under the assumption that if an emotion is biologically basic, then one should expect a one-to-one correspondence between emotions and their “defining behaviors.” For example, Barrett (2006) argues that basic emotions such as fear are not good explanations of animal behaviors:

The most straightforward evidence for whether each emotion category has specific, defining instrumental behaviors comes from the animal learning literature. The simple answer seems to be that behavioral responses correspond to situational demands… rather than to specific emotion categories per se…A fear situation, defined by the presence of threat (e.g., a predator), will be associated with different behaviors (e.g., vigilance, freezing, flight, attack) depending on the functional demands of that specific situation. (p. 42)

Barrett takes this to be an argument against BET (and makes similar arguments about emotions in humans), but it only counts against BET under the following assumption (in the form of a conditional): if situational demands explain variation in anti-predator behaviors across different contexts, then the relevant behavior is not under the control of basic emotion fear. The result would be that anti-predator behaviors are to be explained by other cognitive abilities, perhaps even ones that are not functionally specific to dealing with predators.

Nevertheless, this conditional is true only if fear has a small set of specific defining behaviors *and no other influence on behavior*. In other words, it is only under the strong invariability assumption that BET has a problem explaining contextual variability. Given the arguments above, one can defend the existence of the basic emotion fear against Barrett’s criticism by reassessing what patterns of emotional response are actually predicted by BET. As we saw in the section above, the motivational component of a basic emotion may very well select novel, instrumental behaviors. This kind of sensitivity to “functional demands” is exactly what one should expect of an adaptation for dealing with the basic life problem of predation. If so, it is more difficult to rule out the influence of basic emotions on more complex patterns of behavior in animals and adult humans alike. Consequently, there is another source of potential evidence to support basic emotion theory, aside from so-called emotion signature (evidence of universal facial expressions, etc.). For example, despite their irrationality according to the cannons of rational choice theory, there is evidence that retribution and revenge are nearly universal across cultures (see, e.g., Daly & Wilson, 1988; Henrich et al., 2006). Similarly, there are important similarities across cultures in how people avoid the disease vectors in their environment (Kelly, 2011). If the arguments above are correct, then these phenomena may ultimately be explained by basic emotions like anger and disgust, respectively. Even though revenge and disease avoidance behaviors can be complex and temporally extended, they may very well result from species typical emotions that were shaped by evolution for survival and reproduction. The strong invariability assumption leaves these evolutionary explanations off the table at worst or, at best, permanently disjointed from our understanding of other emotional phenomena (e.g., facial expressions of emotion).

A related theoretical possibility casts doubt on the weak invariability assumption as well: it is possible that for some basic life problems, motivational states are more consistently required than other dimensions of an emotional response (e.g. involuntary facial expressions). Suppose for example, that anger in humans is a solution to basic life problems of deterring harms, challenges, and insults from conspecifics. If so, it may be that the only reliable requirement of successful deterrence (at least in our lineage) is a flexible motivation to retaliate against perceived wrongs (McCullough et al., 2012). For instance, a reliable disposition to garner a reputation for revenge (e.g. by avenging personal offenses) appears to be a highly reliable strategy for deterrence (Daly & Wilson, 1988; Frank, 1988), perhaps more so than any facial expression or physiological responses.[[25]](#footnote-26) But if revenge can be served cold, long after the initial offense, then anger may not always require anything more than a motivation to avenge. If so, then we might expect that the *only* reliably occurring component of anger is the relevant motivational state. While this is a just-so story that may or may not end up being true, it shows that the expected level of coherence in a basic emotion (by which we generate testable predictions) depends critically on which basic life problem shaped that emotion.[[26]](#footnote-27) In some cases, we might expect the motivational state to be the only component that does not significantly vary across the situations in which these problems arise. If so, then contextually variable responses will be the rule rather than an exception to it. This would cast doubt on the weak invariability assumption, too. The upshot is that evidence about response variability by itself is not a good test of BET, nor is it a reliable way to arbitrate between BET and its competitors.[[27]](#footnote-28)

How then can we develop empirical tests among these theories of emotions? The problem is that it is not always clear or obvious how evolution constrains the structure of basic emotions. The arguments above only make headway by telling us that strong invariability is not an important constraint. What more can be said then to develop a positive picture of how evolution constrains the structure of basic emotions? I am skeptical that there are any interesting constraints that apply to *all* the emotional adaptations humans may possess and that would allow us to test the existence of basic emotions as a class.[[28]](#footnote-29) I would even argue that the opposite assumption was the source of many of the shortcomings of earlier expressions of basic emotion theory. With very little theoretical scaffolding, it was merely assumed that a certain class of six or more emotions had defining facial expressions that served as ritualized signals (e.g., Shariff & Tracy, 2011), but why think that there would have been selection pressures on *all* basic emotions to establish a signaling function for each of them?

Theoretical tools like evolutionary game theory have been developed to answer exactly these kinds of questions, but only seldom have they been applied to develop predictions about basic emotions (cf. Frank, 1988; O’Connor, 2016; Sell, 2005). In the case of fear and disgust, there is a straightforward basis for the stability of emotion signals. For example, in the context of close living arrangements characteristic of the last 10,000 plus years of human evolution, transmitting information about disease vectors via expressions of disgust would benefit the sender and receiver of those signals (Kelly, 2011). By contrast, emotions like anger, pride, and shame deal with problems of resource competition, dominance, or prestige. But these are all cases where the interests of potential signalers may conflict, and conflicting interests generate familiar problems: if it pays off to send dishonest signals (e.g., to foil one’s opponent), then a signaling system may never get off the ground. The upshot is that distinct evolutionary problems and pressures may very well have created emotions with very different functional profiles and behavioral symptoms. Therefore, the only sensible way to proceed is by using theoretical tools like evolutionary game theory to generate testable predictions *for each emotion in turn*. We have little reason to think that any theory of evolved emotional responses will be a unified theory (as Griffiths and Ekman seem to suppose).

If this is correct, then the proponents and critics of BET alike have failed to appreciate the boundary between its essential theoretical commitments (evolutionary influence, etc.) and the auxiliary assumptions that help to test them (e.g., invariability assumptions). One result is that proponents of basic emotions have resisted evidence that casts doubt on these auxiliary assumptions, perhaps beyond what is reasonable. Another is that critics of BET have attempted to dismiss the theory when it is the auxiliary assumptions that should be called into question. In particular, Barrett and colleagues (L. F. Barrett, 2006, 2017; Lindquist et al., 2013) have treated data concerning the contextual variability of emotions as critical tests of BET. If the arguments presented here are on track, this may be a mistake.[[29]](#footnote-30) In other words, the reasons given for rejecting BET are not good reasons to reject evolutionary psychological theories of emotional phenomena more generally (e.g., Cosmides & Tooby, 2000; McCullough, Kurzban, & Tabak, 2010; Sell, 2011; Tooby & Cosmides, 2008).

# Conclusion

I have argued against prevailing assumptions that basic emotions are invariable in their behavioral manifestations and that they lack cognitive integration. These features do not mark any important distinction between the products of evolution and the products of culture and experience. This shifts the debate about basic emotion theory: the existence of basic emotions (or at least evolved emotions) does not hinge on whether there exist emotion-related syndromes with contextually invariable effects on behavior, behavioral signatures.

Cases like this reveal the loose connection between the central commitments of theories in the behavioral sciences and the (real or fictional) entities to which they refer. At present, the hypothesis that some emotions are evolutionary products provides very loose constraints on what those products would look like (e.g., whether they would include behavioral signatures). Thus, it remains a challenge to determine the nature of emotions *qua* evolutionary products, and we will remain uncertain whether such emotions exist until this challenge is met. This much is clear: specifying the nature of evolved emotions requires more theoretical work than has been done up to this point.

# References

Adams, D. B. (1980). Motivational systems of agonistic behavior in muroid rodents: A comparative review and neural model. *Aggressive Behavior*, *6*.

Adams, D. B. (2006). Brain mechanisms of aggressive behavior: an updated review. *Neuroscience and Biobehavioral Reviews*, *30*(3), 304–318. https://doi.org/10.1016/j.neubiorev.2005.09.004

Adams, D. B., & Schoel, W. M. (1982). A statistical analysis of the social behavior of the male stumptail macaque (Macaca arctoides). *American Journal of Primatology*.

Barrett, H. C., & Kurzban, R. (2006). Modularity in cognition: framing the debate. *Psychological Review*, *113*(3), 628–647. https://doi.org/10.1037/0033-295X.113.3.628

Barrett, L. F. (2006). Are Emotions Natural Kinds? *Perspectives on Psychological Science*, *1*(1), 28–58. https://doi.org/10.1111/j.1745-6916.2006.00003.x

Barrett, L. F. (2017). How emotions are made: The secret life of the brain. In *How emotions are made: The secret life of the brain.* www.hmhco.com

Berg, D., & Baenninger, R. (1974). Predation: Separation of aggressive and hunger motivation by conditioned aversion. *Journal of Comparative and Physiological Psychology*, *86*(4), 601–606. https://doi.org/10.1037/h0036174

Berridge, K. C. (1996). Food Reward: Brain Substrates of Wanting and Liking. *Neuroscience and Biobehavioral Reviews*, *20*(1), 1–25.

Blanchard, D. C., & Blanchard, R. J. (1984). Affect and aggression: An animal model applied to human behavior. In R. J. Blanchard & D. C. Blanchard (Eds.), *Advances in the Study of Aggression* (Vol. 1, pp. 1–62).

Blanchard, D. C., & Blanchard, R. J. (2003). What can animal aggression research tell us about human aggression? *Hormones and Behavior*, *44*(3), 171–177. https://doi.org/10.1016/S0018-506X(03)00133-8

Blanchard, D. C., Litvin, Y., Pentkowski, N. S., & Blanchard, R. J. (2009). Defense and Aggression. In G. G. Berntson & J. T. Cacioppo (Eds.), *Handbook of Neuroscience for the Behavioral Sciences* (pp. 958–974). Wiley.

Blanchard, R. J., Blanchard, D. C., Takahashi, T., & Kelley, M. J. (1977). Attack and defensive behaviour in the albino rat. *Animal Behaviour*, *25*, 622–634.

Browne, D. (2005). Konrad Lorenz on Instinct and Phylogenetic Information. *The Rutherford Journal*, *1*.

Calder, A. J. (2003). Disgust discussed. *Annals of Neurology*, *53*(4), 427–428. https://doi.org/10.1002/ana.10565

Carruthers, P. (2006). *The architecture of the mind: Massive modularity and the flexibility of thought: Massive modularity and the flexibility of thought*.

Caryl, P. G. (1979). Communication By Agonistic Displays: What Can Games Theory Contribute To Ethology ? *Behaviour*, *68*(1), 136–169. https://doi.org/10.1163/156853979X00287

Castro, D. C., Cole, S. L., & Berridge, K. C. (2015). Lateral hypothalamus, nucleus accumbens, and ventral pallidum roles in eating and hunger: interactions between homeostatic and reward circuitry. *Frontiers in Systems Neuroscience*, *9*(June), 90. https://doi.org/10.3389/fnsys.2015.00090

Celeghin, A., Diano, M., Bagnis, A., Viola, M., & Tamietto, M. (2017). Basic emotions in human neuroscience: Neuroimaging and beyond. In *Frontiers in Psychology* (Vol. 8, Issue AUG, p. 1432). Frontiers Media S.A. https://doi.org/10.3389/fpsyg.2017.01432

Clark, J. (2009). Relations of homology between higher cognitive emotions and basic emotions. *Biology & Philosophy*, *25*(1), 75–94. https://doi.org/10.1007/s10539-009-9170-1

Coltheart, M. (1999). Modularity and cognition. In *Trends in Cognitive Sciences* (Vol. 3, Issue 3, pp. 115–120). Elsevier Ltd. https://doi.org/10.1016/S1364-6613(99)01289-9

Daly, M., & Wilson, M. (1988). *Homicide*. Transaction Publishers.

Defensor, E. B., Corley, M. J., Blanchard, R. J., & Blanchard, D. C. (2012). Facial expressions of mice in aggressive and fearful contexts. *Physiology & Behavior*, *107*(5), 680–685. https://doi.org/10.1016/j.physbeh.2012.03.024

DeWall, C. N., Baumeister, R. F., Stillman, T. F., & Gailliot, M. T. (2007). Violence restrained: Effects of self-regulation and its depletion on aggression. *Journal of Experimental Social Psychology*, *43*(1), 62–76. https://doi.org/10.1016/j.jesp.2005.12.005

Dickinson, A. (1994). Reinforcement and choice. In N. J. Mackintosh (Ed.), *Animal Learning and Cognition* (pp. 45–78). Academic Press.

Eibl-Eibesfeldt, I. (1961). The Fighting Behavior of Animals. *Scientific American*, *205*, 112–122.

Ekman, P. (1977). Biological and cultural contributions to body and facial movement. In J. Blacking (Ed.), *Anthropology of the body* (pp. 34–84).

Ekman, P. (1980). *The Face of Man*. Garland.

Ekman, P. (1999). Basic Emotions. In T. Dalgleish & M. J. Power (Eds.), *Handbook of Cognition and Emotion* (pp. 45–60). John Wiley and Sons. https://doi.org/10.1016/s0005-7967(00)00008-5

Ekman, P. (2006). Darwin, Deception, and Facial Expression. *Annals of the New York Academy of Sciences*, *1000*(1), 205–221. https://doi.org/10.1196/annals.1280.010

Ekman, P., & Cordaro, D. (2011). What is Meant by Calling Emotions Basic. *Emotion Review*, *3*(4), 364–370. https://doi.org/10.1177/1754073911410740

Ekman, P., & Friesen, W. V. (1969). Nonverbal Leakage and Clues to Deception. *Psychiatry*, *31*(1), 88–106. https://doi.org/10.1080/00332747.1969.11023575

Ekman, P., Friesen, W. V, O’Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., & others. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology; Journal of Personality and Social Psychology*, *53*(4), 712.

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, *221*(4616), 1208–1210.

Ekman, P., Roper, G., & Hager, J. C. (1980). Deliberate Facial Movement. *Child Development*, *51*(3), 886–891. https://doi.org/10.2307/1129478

emotion, R. Z.-A. to, & 1984, undefined. (n.d.). The interaction of affect and cognition. *Books.Google.Com*. Retrieved August 9, 2020, from https://books.google.com/books?hl=en&lr=&id=k0mhAwAAQBAJ&oi=fnd&pg=PA239&dq=%22the+interaction+of+affect+and+cognition%22&ots=kX7gHVjDYO&sig=\_HdpvVR3ko-eAsS2cAT3uWDxNPo

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

Fokkema, D. S., Koolhaas, J. M., & van der Gugten, J. (1995). Individual characteristics of behavior, blood pressure, and adrenal hormones in colony rats. *Physiology & Behavior*, *57*(5), 857–862.

Frank, R. H. (1988). *Passions Within Reason: The Strategic Role of the Emotions.* Norton. https://doi.org/10.2307/2072516

Friesen, W. V. (1973). *Cultural differences in facial expressions in a social situation: An experimental test on the concept of display rules.* University of California, San Francisco.

Frijda, N. H. (1986). The Emotions. In *The Emotions* (Issue November 2003). Cambridge University Press. https://doi.org/10.1093/0199253048.001.0001

Gaulin, S. J. C., & McBurney, D. H. (2001). *Psychology: An Evolutionary Approach*. Prentice Hall.

Gendron, M., Roberson, D., & Barrett, L. F. (2014). *Perceptions of Emotion From Facial Expressions Are Not Culturally Universal : Evidence From a Remote Culture*. *14*(2), 251–262. https://doi.org/10.1037/a0036052

Goldstone, A. P., Prechtl De Hernandez, C. G., Beaver, J. D., Muhammed, K., Croese, C., Bell, G., Durighel, G., Hughes, E., Waldman, A. D., Frost, G., & Bell, J. D. (2009). Fasting biases brain reward systems towards high - calorie foods. *European Journal of Neuroscience*, *30*, 1625–1635. https://doi.org/10.1111/j.1460-9568.2009.06949.x

Griffiths, P. E. (1997). *What emotions really are: The problem of psychological categories* (Vol. 1997). University of Chicago Press. http://books.google.com/books?hl=en&lr=lang\_en&id=t3BgX4KRVksC&oi=fnd&pg=PR9&dq=What+Emotions+Really+are&ots=6dQKjXxUh0&sig=T3BXLFNRijEO5vkLZmCJaauKvlU

Griffiths, P. E. (2003). Basic Emotions, Complex Emotions, Machiavellian Emotions. In A. Hatzimoysis (Ed.), *Philosophy and the Emotions* (pp. 39–67). Cambridge University Press. https://doi.org/10.1017/S1358246100007888

Griffiths, P. E. (2004). Towards a “Machiavellian” Theory of Emotional Appraisal. In *Emotion, Evolution, and Rationality*. https://doi.org/10.1093/acprof:oso/9780198528975.003.0005

Haggard, E., & Isaacs, K. (1966). Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy. *Methods of Research in Psychotherapy*.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Henrich, N. S., Gwako, E., Lesorogol, C., Marlowe, F., Tracer, D., & Ziker, J. (2006). Costly Punishment across Human Societies. *Science*, *312*(5781), 1767–1770. https://doi.org/10.1126/science.1127333

Izard, C. E. (1992). *Basic emotions, relations among emotions, and emotion-cognition relations.*

Izard, C. E. (2007). Basic Emotions, Natural Kinds, Emotion Schemas, and a New Paradigm. *Perspectives on Psychological Science*, *2*(3), 260–280. https://doi.org/10.1111/j.1745-6916.2007.00044.x

Kelly, D. R. (2011). *Yuck!: The Nature and Moral Significance of Disgust*. MIT Press.

Khalidi, M. A. (n.d.). Nature and Nurture in Cognition. In *academic.oup.com*. Retrieved August 9, 2020, from https://academic.oup.com/bjps/article-abstract/53/2/251/1546365

KhAlidi, M. ali. (2007). *Innate cognitive capacities*. https://philarchive.org/rec/KHAICC

LALAND, K. N., & READER, S. M. (1999). Foraging innovation in the guppy. *Animal Behaviour*, *57*(2), 331–340. https://doi.org/10.1006/anbe.1998.0967

Language, S. S.-C. P. M. and, 1972, undefined, & 2011, undefined. (n.d.). The idea of innateness. *Researchwithrutgers.Com*. Retrieved August 9, 2020, from https://www.researchwithrutgers.com/en/publications/the-idea-of-innateness

Levenson, R. W., Ekman, P., & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology*, *27*(4), 363–384.

Lindquist, K. A., Siegel, E. H., Quigley, K. S., & Barrett, L. F. (2013). *The Hundred-Year Emotion War : Are Emotions Natural Kinds or Psychological Constructions ? Comment on Lench ,*. *139*(1), 255–263. https://doi.org/10.1037/a0029038

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences*, *35*(03), 121–143. https://doi.org/10.1017/S0140525X11000446

McCullough, M. E., Kurzban, R., & Tabak, B. A. (2012). Cognitive systems for revenge and forgiveness. *The Behavioral and Brain Sciences*, *36*(1), 1–15. https://doi.org/10.1017/S0140525X11002160

Nonacs, P. (2001). State dependent behavior and the Marginal Value Theorem. *Behavioral Ecology*, *12*(1), 71–83. https://doi.org/10.1093/oxfordjournals.beheco.a000381

O’Connor, C. (2016). Guilt, Games, and Evolution. In A. Scarantino (Ed.), *Emotion Researcher*. ISRE.

O’Neill, E. (2015). Relativizing innateness: innateness as the insensitivity of the appearance of a trait with respect to specified environmental variation. *Biology & Philosophy*, *30*, 211–225.

Olton, D. (1979). Mazes, maps, and memory. *American Psychologist*.

Panksepp, J. (2004). *Affective Neuroscience: The Foundations of Human and Animal Emotions* (1st ed.). Oxford University Press, USA.

Potegal, M., & Davidson, R. J. (2003). Temper tantrums in young children: 1. Behavioral composition. *Journal of Developmental & Behavioral Pediatrics*, *24*(3), 140.

Reisenzein, R., Studtmann, M., & Horstmann, G. (2013). Coherence between Emotion and Facial Expression: Evidence from Laboratory Experiments. *Emotion Review*, *5*, 16–23. https://doi.org/10.1177/1754073912457228

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

Scarantino, A. (2015b). The motivational theory of emotions. In J. D’Arms & D. Jacobson (Eds.), *Moral Psychology and Human Agency*. Oxford University Press.

Scarantino, A. (2017). Do Emotions Cause Actions, and If So How? *Emotion Review*, 175407391667900. https://doi.org/10.1177/1754073916679005

Scarantino, A., & Nielsen, M. (2015). *Voodoo dolls and angry lions : how emotions explain arational actions*. 2975–2998. https://doi.org/10.1007/s11098-015-0452-y

Seeley, R. J., & Berridge, K. C. (2015). The hunger Games. *Cell*, *160*(5), 805–806. https://doi.org/10.1016/j.cell.2015.02.028

Sell, A. (2005). *Applying Adaptationism to Human Anger: The Recalibrational Theory*.

Shariff, A. F., & Tracy, J. L. (2011). What Are Emotion Expressions For? *Current Directions in Psychological Science*, *20*(6), 395–399. https://doi.org/10.1177/0963721411424739

Spunt, R. P., & Adolphs, R. (2017). A new look at domain specificity: insights from social neuroscience. *Nature Reviews Neuroscience*, *18*(9), 559–567. https://doi.org/10.1038/nrn.2017.76

Tebbich, S., Griffin, A. S., Peschl, M. F., & Sterelny, K. (2016). From mechanisms to function: an integrated framework of animal innovation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *371*(1690).

*The Distinction Between Innate and Acquired Characteristics (Stanford Encyclopedia of Philosophy)*. (n.d.). Retrieved August 9, 2020, from https://plato.stanford.edu/entries/innate-acquired/

Tolman, E. (1948). Cognitive maps in rats and men. *Psychological Review*.

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*(4), 397–405. https://doi.org/10.1177/1754073911410747

Wiegman, I. (2016). Angry Rats and Scaredy Cats: Lessons from Competing Cognitive Homologies. *Biological Theory*, *11*(4), 224–240. https://doi.org/10.1007/s13752-016-0249-z

Wiegman, I. (2019). Payback without bookkeeping: The origins of revenge and retaliation. *Philosophical Psychology*, *32*(7), 1100–1128. https://doi.org/10.1080/09515089.2019.1646896

Wiegman, I. (2020). Emotional Actions Without Goals. *Erkenntnis*. https://doi.org/10.1007/s10670-019-00200-8

Woodward, J. (2005). *Making things happen: A theory of causal explanation*. Oxford University Press.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*, 151–175.

1. However, conscious recognition is obviously not the only route to achieve adaptive effects of ritualized signals. [↑](#footnote-ref-2)
2. For my purposes, innateness can be thought of roughly in terms of *phylogenetic information* (see, e.g., Browne, 2005), where this is understood as “fit” with the environment that is can only be explained by evolutionary adaptation. As Griffiths (*The Distinction Between Innate and Acquired Characteristics (Stanford Encyclopedia of Philosophy)*, n.d.) notes, this notion is closely connected with the evidential role of deprivation experiments in classic poverty of the stimulus arguments. (See, e.g., Khalidi, n.d.; KhAlidi, 2007; Language et al., n.d.; O’Neill, 2015) [↑](#footnote-ref-3)
3. Coltheart (1999) argues that modularity is best defined in terms of domain specificity. For a recent discussion of domain specificity in neuroscience, see also Spunt and Adolphs (2017). Kurzban and Barrett (2006) argue convincingly that modularity is primarily about functional specialization. Rather than committing myself to any particular way of defining modularity, I will merely use the word as a shorthand for the characteristics of emotion most relevant to behavioral variability, encapsulation and opacity, but those characteristics may only be imperfect indicators of modularity qua functional specialization (or domain specificity, or whatever else). [↑](#footnote-ref-4)
4. Of course, massive modularity theorists argue that central processes are also modular in some sense (e.g., Carruthers, 2006). However, their notion of modularity is not closely tied to encapsulation or opacity. Moreover, these accounts need not deny that there are important, observable differences between processes that are more automatic and mandatory and those that are more deliberate and controlled. Even on a massively modular view of cognition, it is still possible to articulate a relationship between emotion and cognition on which the one is encapsulated and opaque to the other. [↑](#footnote-ref-5)
5. It is beyond the scope of this paper to consider the neuroscience of emotions any further, given the vastness of that literature and given my focus on behavioral variability and modularity. As I understand it, much of the neuroscientific data on emotions is only tangentially relevant to these issues. This is especially true of the primary tool for investigating the neural basis of human emotions: fMRI. Celeghin et al (2017) capture some of the difficulties in using these neuroimaging methods to make any claims about basic emotions in humans, in particular claims about the speed or automaticity of affect programs:

   What inferences can be made on the basis of fMRI studies? What are the limitations of the methodology with respect to the debate on the existence of basic emotions?

   A first limit is epistemological. That is, imaging studies are good at revealing which neural structures are involved in the processing of basic emotions, but are silent with respect to what structures are necessary to recognize or express such emotions. In this sense, they offer a type of ‘weak’ or correlational evidence and should be interpreted in the light of other data, such as lesion studies, in which the correlational nature of fMRI data is elevated to a causal inference (Krakauer et al., 2017). Others and we tend to believe that the starting point to understand the neurobiology of emotions *is the analysis of behavior*, as we cannot rely solely on the correlational approach of neuroimaging data devoid of relation with behavioral outcomes. Indeed, the causal-mechanistic explanations are qualitatively different from understanding how component modules perform the computations that then combine to produce behavior (Krakauer et al., 2017)…

   To sum up, neuroimaging data and recent meta-analyses do not seem to us to provide sufficiently solid ground for rejecting the existence of basic emotions at the neurobiological level. In addition, other features considered typical of basic emotions, like automaticity or early onset during sensory processing (Ekman, 1999; see above), are not considered in these studies… (p. 4)

   More specific to the issues at hand, it is difficult to see how neural data could possibly help to elucidate whether an emotion can control flexible patterns of behavior or only rigid ones (as the strong invariability assumption states). At present it simply is not possible to use neural data to determine whether a behavior is caused by a basic emotion or not. To do so would involve a reverse inference from patterns of brain activation to the involvement of a basic emotion, and we would only have that if we could resolve one of the most heated debates in current neuroscience of emotion (see, e.g., Celeghin et al., 2017; Lindquist et al., 2012). Other lines of research in the neuroscience of emotion are no more relevant to the issues of behavioral variability and modularity: EEG studies tend to focus on perceptual processing of emotional facial expressions rather than emotional responses as such, and lesion studies provide only ambiguous and indirect evidence concerning changes in emotional behavior. So, while neuroscientific research is important for assessing certain claims about the existence of basic emotions, it does little to help assess the invariability assumptions that are my focus here. [↑](#footnote-ref-6)
6. For a critical perspective, see, Celeghin et al (2017). [↑](#footnote-ref-7)
7. Ekman apparently retains this view in more recent writings (2003). He writes the following in an endnote: “Frijda’s description of the actions that characterize each emotion includes what I have said and quite a bit more. I believe it is only these rudimentary initial postural moves [e.g. looking down on an object of contempt, fixed attention on the object of surprise, movement toward the source of sensory pleasure, and slumping posture and loss of muscle tone in sadness] that are inbuilt, automatic, and universal.” (p. 268) Moreover, a more recent piece by Ekman and Cordaro (Ekman & Cordaro, 2011) suggest that the view remains largely the same (see esp. pp. 367 and 368). [↑](#footnote-ref-8)
8. At least, similar outcomes with some reasonable exceptions, such as when the emotion is suppressed or regulated in some other way. Ekman does appear to appreciate this point, and makes some comments that suggest he is skeptical that anger has any “central direction.” Nevertheless, once we distinguish between the direct motivational effects of anger on the one hand, and the individually variable effects of emotion regulation on the other (a distinction that Ekman does not consistently draw), I believe it becomes possible to see the outlines of a central direction. This makes the assessment of an emotion’s central direction less empirically tractable. Nevertheless, the problems with the central direction of hunger are exactly analogous, but clearly not insurmountable. The fact that I am hungry may not necessarily manifest itself in my behavior in any straightforward way (e.g., if I am feverishly working on an essay that I hope to submit in short order). [↑](#footnote-ref-9)
9. Scarantino (2015b) proposes a motivational theory of emotions that makes motivational states essential to emotions, basic or otherwise. See also, Scarantino (2017), Scarantino & Nielsen (2015), and Wiegman (2020). [↑](#footnote-ref-10)
10. See, e.g., Daly and Wilson (1988, Chapter 10) and McCullough et al (2012) [↑](#footnote-ref-11)
11. Anger need not always lead to such actions any more than hunger always leads to pursuit of food, but that is just to say that like hunger, the behavioral dispositions of anger can be suppressed by countervailing desires or by systems for executive control. [↑](#footnote-ref-12)
12. Wiegman (2020) argues that some emotional actions do not have goals, in a sense roughly identical to Dickinson’s. These arguments undermine any simple distinction between “emitted, flexible behaviors and elicited, fixed ones,” where the category of “action” and “flexible behaviors” are subsumable to “goal-directed behaviors.” [↑](#footnote-ref-13)
13. Though he confesses that he and his associates never tested this hypothesis quantitatively “because it seemed so obvious” (Ekman, 2006, p. 208). [↑](#footnote-ref-14)
14. While there is some experimental evidence of this capacity (see e.g. LALAND & READER, 1999), this kind of evidence appears to be surprisingly scant, perhaps because it seems obvious enough to take for granted. [↑](#footnote-ref-15)
15. Though there may be distinct motivations for hunting and feeding. See Berg and Baenninger (1974). [↑](#footnote-ref-16)
16. For an integrated theoretical treatment of the many influences on foraging behavior, see Nonacs (2001). [↑](#footnote-ref-17)
17. The idea of integration/modularity of motivational states needs sharpening. Were there space, a more thorough way to do this is to start with a causal or computational characterization of encapsulation and opacity, perhaps along manipulationist lines (e.g. Woodward, 2005). In both cases, the limitation is in the ability of other systems to manipulate (or cause changes in) the target system at intermediate levels of processing in a way that influences its performance of the relevant function. [↑](#footnote-ref-18)
18. Another thing worth noting here is that there may actually be some variability in certain signaling functions, which could also make evolved facial signals highly sensitive to context. See Griffiths (Griffiths, 2003, 2004) for detailed discussions. [↑](#footnote-ref-19)
19. I think this justification ultimately fails on independent grounds: Griffiths is running together two things that should be kept separate. One is a question about existence: are there any psychological states in the ballpark of basic level emotions categories (e.g. anger, fear, disgust, interest) that are evolutionary adaptations? Another is a question about essence: are there any “psychological and neurological principles” that capture the category defining features of the class of basic emotions? While both questions are important, the existence of special purpose adaptations in the ballpark of anger and fear (etc.) simply does not depend on whether there is some important similarity among the states in those (possibly very separate) ballparks. Much less does the question of existence depend on whether the relevant states are contextually invariable. [↑](#footnote-ref-20)
20. Caroline and Robert Blanchard call them *offense* and *defense*, respectively (Blanchard et al., 2009; Blanchard & Blanchard, 1984, 2003). [↑](#footnote-ref-21)
21. See, e.g., Wiegman (2019) [↑](#footnote-ref-22)
22. Recall from section 2 that there is no reason to suppose that basic emotions are unique to humans or that the set of basic emotions in humans is the same as in other species. [↑](#footnote-ref-23)
23. Perhaps a key difference between this research and research on basic emotions in humans is that no evidence has been presented that the facial expressions and physiological responses are specific to the relevant emotional state. There is even some evidence against this claim regarding rats and mice: The facial expressions appear to have a protective function rather than a communicative function. See Defensor et al (2012). Nevertheless, lack of specificity does nothing to negate the evidence that these distinct behavior suites exist and are adapted to solve basic life problems of sociality. [↑](#footnote-ref-24)
24. In connection with this point, Adams and Schoel (1982) have pointed out that captive macaques also attempt to bite the backs of “intruders.” [↑](#footnote-ref-25)
25. For instance, in some resource competition scenarios, signaling aggressive intent is not an adaptive strategy (Caryl, 1979). As mentioned above, the facial expression associated with the confrontation in rats is not actually specific to confrontation. For instance, the rat will make a similar expression if its vibrissae are repeatedly stroked, suggesting that it may instead be a protective response to potential contact with the snout (Defensor et al., 2012). So if these expressions have any signaling function it appears to be secondary to their protective function, and perhaps also dissociable from it. [↑](#footnote-ref-26)
26. See Gaulin and McBurney (2001, p. 266) for a similar argument. [↑](#footnote-ref-27)
27. Lisa Barrett’s (2017) alternative to BET also predicts contextual variability. So contextual variability will not tell us which of these competing theories is most likely to be true. While this discussion has focused on behavioral invariability and modularity, other kinds of invariability assumptions have also come under scrutiny recently, especially invariability of neural responses, or the assumption that there is a one-to-one mapping between brain regions and emotions (cf. Lindquist et al., 2012). Similar to the case I have made concerning behavioral variability, it is unlikely that this commitment is part of the theoretical core of BET, because no one has shown how it follows from the central commitments of the theory. [↑](#footnote-ref-28)
28. Frank’s (Frank, 1988) notion of a commitment device comes close to providing such a rationale, but many emotions are clearly not commitment devices (e.g., disgust, happiness) or at least that is not their first and foremost function. For a discussion of disgust’s signaling function that considers and rejects its role as a commitment device, see Kelly (2011). [↑](#footnote-ref-29)
29. See Scarantino (2015a) for an argument along similar lines. One reply to this line of argument is just that BET should include the commitments that its proponents consider essential to the theory. In other words, BET is falsified by the contextual variability of emotions because Ekman and others articulate these assumptions as central to the theory. I am happy to grant this in exchange for a concession: the falsity of BET does not prove there are no special purpose adaptations in the ballpark of basic level emotion categories like anger, fear, etc. (I have yet to see Barrett and others make this kind of concession. On the contrary, they appear to take the falsity of BET as a reason to reject domain-specific theories of emotion quite broadly.) If the arguments above are correct, then special purpose adaptations may very well be cognitively integrated in a way that supports contextual variability. If they exist, some would lack so-called “behavioral signatures,” but they might resemble basic level emotion categories. For example, even if confrontation and avoidance systems in rats should not be called “basic emotions” because of their contextual variability, they are still special purpose adaptations that resembles human emotions in various respects (e.g., as irruptive, impulsive, or reactive motivational states). And if there are motivational adaptations like these in rats, they may very well be conserved from our common ancestors. Thus, there may very well be echoes of these special purpose adaptations in the human lineage as well (see e.g., Wiegman, 2016). [↑](#footnote-ref-30)