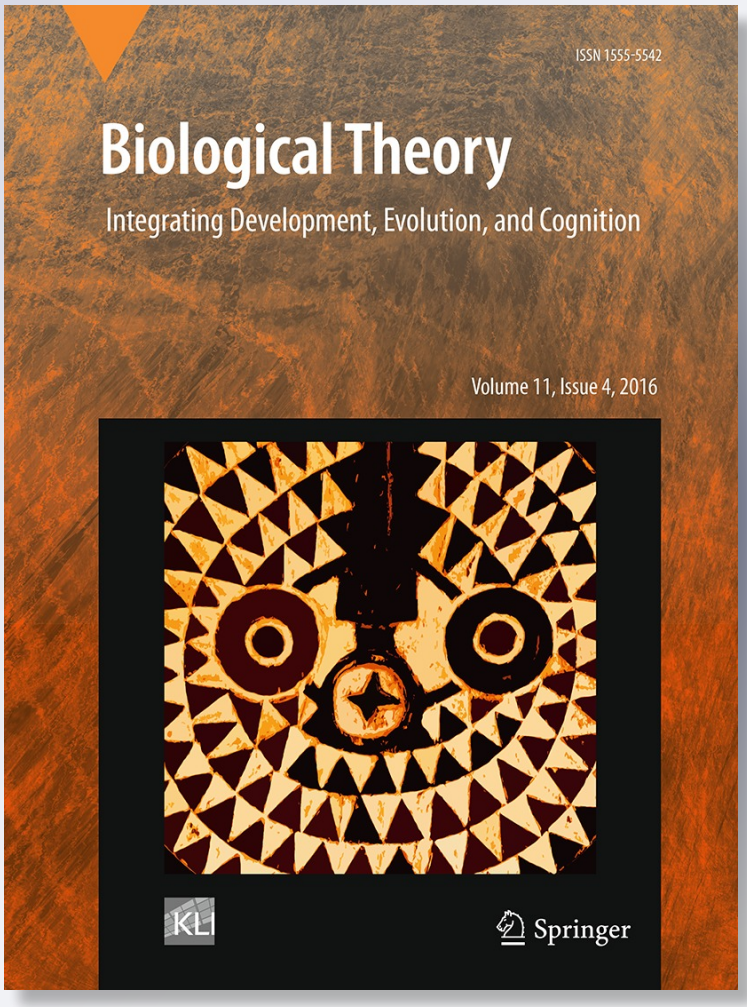
Angry Rats and Scaredy Cats: Lessons from Competing Cognitive Homologies

Isaac Wiegman

Biological Theory

ISSN 1555-5542
Volume 11
Number 4

Biol Theory (2016) 11:224-240
DOI 10.1007/s13752-016-0249-z



Springer

Biol Theory (2016) 11:224–240
DOI 10.1007/s13752-016-0249-z

ORIGINAL ARTICLE

Angry Rats and Scaredy Cats: Lessons from Competing Cognitive Homologies

Isaac Wiegman[1]

Received: 8 March 2016 / Accepted: 22 July 2016 / Published online: 6 September 2016

Abstract There have been several recent attempts to think about psychological kinds as homologies. Nevertheless, there are serious epistemic challenges for individuating homologous psychological kinds, or cognitive homologies. Some of these challenges are revealed when we look at competing claims of cognitive homology. This paper considers two competing homology claims that compare human anger with putative aggression systems of nonhuman animals. The competition between these hypotheses has been difficult to resolve in part because of what I call the boundary problem: boundaries between instances of psychological kinds (e.g., anger and fear) cannot be directly observed. Thus, there are distinctive difficulties for individuating psychological kinds across lineages. I draw four conclusions from this case study: First, recent evidence from the neuroscience of fear suggests that one of the proposed homologies involves a straightforward conflation of anger and fear. Second, this conflation arises because of the boundary problem. Third, there is an implicit constraint on the operational criteria that is easy to overlook in the psychological case. In this case, ignoring the constraint is part of the problem. Fourth, this is a clear case in which knowledge of homology cannot be accumulated piecemeal. Identifying homologs of human anger requires identifying homologs of fear.

Keywords Anger · Aggression · Cognitive homology · Fear · Homology

✉ Isaac Wiegman
Isaac.wiegman@txstate.edu

[1] Texas State University, San Marcos, TX, USA

Introduction

Homologous traits are traits that are derived from a single, ancestral trait. Hypotheses of homology are thus historical explanations for trait similarity (Ereshefsky 2012). Some have suggested that this kind of homology thinking is critical for individuating psychological kinds (e.g., Griffiths 1997, 2006; Matthen 2007; Ereshefsky 2012). For instance, Griffiths argues that homology classes (for now, classes of traits derived from common ancestral traits) share deeper similarities in causal structure than analogy classes (classes of traits that are similar due to convergence). So categories based on homology license extrapolative inferences with greater reliability (Griffiths 1994). It would follow that one research aim of the cognitive sciences should be to decompose the mind (when possible) into cognitive traits defined by homology, or *cognitive homologies*, to underwrite the kind of extrapolative inferences that mature scientific theories afford.

Some have expressed misgivings about this understanding of homology (e.g., Rosenberg and Neander 2009). However, even if the value of homology is taken for granted, there are many difficulties in reaping the fruits of this approach. One symptom of these difficulties is the existence of competing claims of cognitive homology. In this article, I consider the case of human anger, in which competing cognitive homologies have been proposed. I draw four conclusions from this case study: First, recent evidence from the neuroscience of fear suggests that one of the proposed homologies may involve a straightforward conflation of anger and fear. Second, I argue that this conflation arises because of the boundary problem: we cannot directly observe psychological kinds or the boundaries between instances of one kind and another. One of the most prominent strategies for assessing homology is to

identify similarities that indicate derivation from an ancestral trait (using the operational criteria of homology), but anger and fear have many similarities in both humans and other animals. Since we cannot directly observe boundaries between instances of anger and fear, their similarities are especially misleading. Third, there is an implicit constraint on the operational criteria that is easy to overlook in the psychological case: the *class-specificity constraint*. In this case, ignoring the constraint is part of the problem, suggesting that the implicit constraint needs to be an explicit guide for assessing cognitive homology. The class-specificity constraint motivates the search for boundaries between distinct psychological kinds in each organism being compared, and this can mitigate boundary problems. Fourth, this is a clear case in which knowledge of homology cannot be accumulated piecemeal. That is, identifying homologs of human anger requires also identifying homologs of human fear.

I begin in the second section ("Homology and Competing Hypotheses") by saying more about homology thinking. To understand what kind of evidence supports homology, I point out a range of hypotheses with which it competes and set out the kind of evidence that favors homology over and above them. I point out some of the difficulties in identifying cognitive homologies, and introduce the operational criteria of homology, with which some have attempted to address these problems. The operational criteria of homology (Remane 1971) can be understood as identifying similarities that provide evidence for homology over and above these competing hypotheses. Moreover, these criteria have been successfully applied to identify cases of cognitive homology.

However, in the third section ("Competing Hypotheses of Homology"), I consider a specific case in which two competing homology claims have been made concerning a single psychological trait, human anger. This emotion has been compared with two distinct aggression systems in nonhuman animals, what I call the confrontation and defense systems. Each of these proposed homologies seems to satisfy some of the operational criteria; nevertheless, recent evidence suggests that the defense system is homologous to fear rather than anger. Essentially, one of the two hypotheses involves the conflation of two distinct psychological kinds. It would be difficult to make this kind of mistake in the case of morphology.

In the fourth section ("Conflating Kinds: The Defense Hypothesis"), I attempt a diagnosis. I suggest that the problem arises because in the domain of psychology one cannot directly observe boundaries between instances of homologies, whereas these boundaries are observable in the paradigmatic case of skeletal morphology. In the domain of psychology, it is therefore easier to flout an implicit constraint on homology claims: what I call the *class-specificity constraint*. In the fifth section ("A Constraint on Homology Claims"), I argue that this implicit constraint on homology inferences restricts the kind of evidence admitted in favor of a hypothesis of homology. Using the criteria of homology in accordance with this constraint sometimes requires a search for boundaries or distinguishing characteristics between instances of two or more psychological kinds. I conclude by observing that knowledge of homology cannot always be acquired piecemeal, and in the cognitive domain this may be the rule rather than a mere exception.

Homology and Competing Hypotheses

To start, let us begin with the concept of homology. Though the concept of homology is crucial to evolutionary thinking, it was conceived in the service of biological taxonomy prior to Darwin's time. Owen (1846) thought of homology as the sameness of an organ or structure in different organisms under every *form* and *function*. In pre-Darwinian science, sameness was determined with reference to ideal animal archetypes that were divisible into parts. There is now broad agreement that common ancestry is a cause of many structural and functional similarities among organisms and that hypotheses of homology attempt to capture the similarities attributable to common ancestry. Despite this consensus, it remains controversial what exactly determines sameness of structure or function. Supposing that two traits derive from a trait of a common ancestor, we can ask, what is it that makes each of these traits identical to the trait of the common ancestor?

While contemporary accounts of homology give different answers to this question, these accounts are broadly consistent with the thought that homology is a causal-historical concept (for a clarification and defense of this claim, see Ereshefsky 2012).[1] Specifically, a homology refers to traits of various animals that are continuous across lineages. In this way, shared ancestry or continuity across a lineage is the common cause of each homolog (e.g., Assis and Brigandt 2009), and the effects of ancestry can be observed at different levels (phenotypic, developmental, genetic, etc.). Moreover, ancestry as a common cause provides a historical explanation for similarities between the homologous traits (whether they are traits of different organisms or traits of the same organism, as in the case of serial homology). In the words of Rieppel, homology is "…grounded in 'descent, with modification,' a process that belongs to the past" (Rieppel 2005).

[1] For a recent overview and an interesting proposal, see Ramsey and Peterson (2012). Their idea is that that sameness is determined by phylogenetic continuity and numerical identity across one or more hierarchical levels within organisms (e.g., levels at which phenotypes are observed or at which developmental mechanisms or gene networks operate).

As a causal-historical concept, we can sometimes refer to a homology without having detailed knowledge of the history of the developmental and hereditary mechanisms that give rise to each of its instances, just as we can refer to a disease entity, such as measles or chicken pox, without knowing about its underlying causes (Putnam 1969).[2] Nonetheless, we learn more about each homology as we learn more about its underlying causes, just as we learn more about chicken pox as we learn more about the virus that causes it.

Given the causal-historical nature of homology, there is a vast range of evidence that could bear on whether or not one trait is homologous to another. Some of the best evidence pertaining to homology comes from cladistics. If one has an independently established phylogenetic tree, one can look at the distribution of a candidate homology, or character, on that tree. If, for instance, the existence of a homology is a more parsimonious explanation than convergent evolution on one or more occasion, then there is some reason to think that a trait is homologous.

Nevertheless, to bring this knowledge to bear on a hypothesis of homology, one must be able to decompose organisms into units; units that can be identified across species. Whereas the causal-historical approach explains how we can successfully refer to an entity or property in a given species, successful reference does not guarantee that one can identify the trait across the entire range of its manifestations. One might recognize paradigmatic instances of carbon combustion without successfully identifying other manifestations of the same oxidative process, such as rusting iron. Accordingly, a historical definition of homology actually presupposes a valid decomposition of organisms into units:

> Any phylogenetic investigation starts with a mental decomposition of the organisms into units of description or characters. Only then can the techniques to evaluate the historical relationships among character states be applied and genealogical continuity inferred. Character definition is expected to be non-arbitrary, such that the union of a hoof and the cerebellum is not acceptable as a character. (Wagner 1996, p. 36)

While we can assume that there are many ways to meaningfully decompose organisms (morphological, developmental, genetic, psychological, etc.), a central contention of this article is that meaningful decompositions can be more difficult to obtain in some domains (e.g., the cognitive) than others (e.g., the morphological).

This becomes clear when we look at how the homology concept has been extended to functional categories, including behavior and psychology (see, e.g., Wenzel 1992; Ereshefsky 2007; Love 2007; Clark 2009; García 2010; Murphy 2012). A symptom of the difficulty is this: the cladistic approach for identifying homologies cannot always be straightforwardly applied in the domain of psychology. Before we can even look at the distribution of a character on a phylogenetic tree, we need to know how to identify the character in each taxon. This is a difficult matter when dealing with behavioral and psychological characters. Psychological characters in particular are not directly observable, nor are the boundaries between each one.

Consider aggression research in nonhuman animals. Some neurophysiologists claim to have identified two separate aggression systems[3] that lead to different forms of aggression in cats (see, e.g., Siegel 2004; Siegel and Victoroff 2009). Both forms of aggression can be induced through electrical brain stimulation (EBS) in distinct regions of the hypothalamus. That is to say that they are neurally dissociated. The so-called *defense* system leads to unconstrained attacks when areas in the cat ventromedial hypothalamus (VMH) are stimulated, whereas the so-called *predation* system leads to "quiet biting attack" (directed at prey species) when distinct regions are stimulated. This is just a small piece of the larger body of evidence used to infer the boundaries between putative psychological or behavioral categories of defense and predation. Rather than being directly observable from behavior, these boundaries were inferred from this considerable evidence base (which includes neurophysiological work across several species).

Similarly, scientists in the ethological tradition claim to have identified two separate psychological (or perhaps behavioral) systems that lead to different forms of aggression in various rodent species (e.g., Blanchard and Blanchard 1984, 2003; Adams 2006; Blanchard et al. 2009). For instance, in one experimental paradigm, the *resident-intruder* paradigm, a male rat, the resident, encounters an unfamiliar male rat, the intruder, in the resident's territory. The result is that the resident will engage in repeated attempts to bite the back of the "intruder." This confrontational form of attack is distinct from the attacks mounted by the intruder, which tend to be aimed at the snout of the resident rat. This latter form of attack seems to be aimed at preventing back biting and creating a window for escape and avoidance, and it is coordinated with other behaviors (e.g., freezing) that also subserve defense from predators. A wealth of ethological observation and experimental data suggests that these

[2] We may not be able to *individuate* homologies without respect to processes of development, etc. Nevertheless, it is clear that we can successfully refer to an entity without knowing how to individuate it.

[3] I take it that "aggression" is a behavioral or psychological category rather than a neurophysiological one.

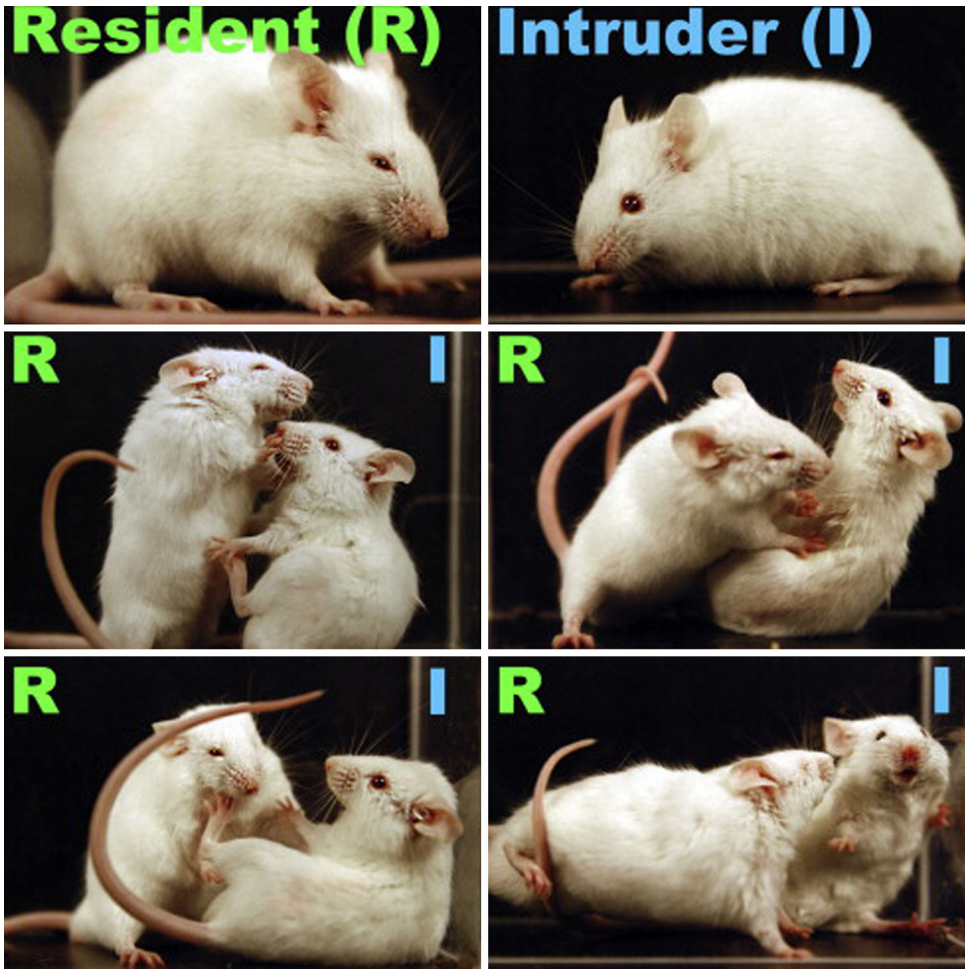


Fig. 1 Confrontation and avoidance behaviors (e.g., facial expressions, postures and maneuvers) of resident and intruder mice (respectively) From Defensor and Corley (2012), p. 683 © Elsevier. Originally published in *Physiology and Behavior*

forms of attack are subserved by distinct psychological systems, what I call the *confrontation* and *avoidance* systems (see Fig. 1).[4] But again, these putative psychological systems are not directly observed. Rather, scientists have used empirical methods to infer the boundaries between putative psychological categories of confrontation and avoidance.

However, the methods for inferring these boundaries in rats and cats are simply not appropriate to doing so in many other species. For instance, the ethological methods appropriate to rats are inappropriate to humans and many other species due to differences in social ecology. Human males, for instance, are not consistently required to physically defend territories or dominance position for breeding purposes. Similarly, an ethical constraint on invasive surgical intervention in humans (of the sort required for EBS in the hypothalamus) is that the interventions have a therapeutic aim for the individual being operated on. This limitation makes it difficult, though not impossible, to test the effects of EBS in the hypothalamus in humans. Due to relatively recent development of deep brain stimulation techniques, it has become possible on occasion to investigate the relation between psychological states (of panic, aggression, etc.) and neural activity in the hypothalamus (as I discuss below).

In any case, cladistic tests require methods for individuating a psychological trait in each taxon, and I suggest later that this sometimes requires distinguishing it *from* at least one other trait. Since there are going to be many cases in which we cannot avail ourselves of the same methods to individuate psychological traits in humans, and since we will often be in doubt as to which methods are sufficiently similar, this cladistic approach to homology sometimes cannot be applied to psychological traits of humans.

[4] Most researchers call these the offense and defense systems. While some researchers assume that what I call the avoidance system in rats is of a piece with defense in cats, it has also become the norm to assume that the avoidance system is closely related to, or even part of, the fear system (Blanchard and Blanchard 2008; though see also Gross and Canteras 2012). Nevertheless, other researchers argue that the defense system is closely related to anger (Panksepp 1998), as I also discuss below. Introducing my own labels is a way to maintain distinctions between different lines of research which happen to use similar terms and to avoid prejudging the matter one way or another.

One way of addressing this problem is to use the operational criteria of homology (e.g., Ereshefsky 2007; Love 2007; Clark 2009; García 2010; Murphy 2012). These criteria need not function as a definition of homology but instead, we can use them to establish a consistent set of methods for ascertaining homologies and by extension, identical traits. Importantly, the identification of homologies via the operational criteria does not depend on any method for individuating a given trait across clades.

How then do the criteria function? They identify particular kinds of similarity, the kinds that are best explained by positing continuity of traits across lineages over and above a range of competing hypotheses. What are the competing hypotheses? One is that the similarity is due to chance convergence (e.g., as a result of drift). Another more probable possibility is that convergent evolution (due to adaptation) explains the similarity. This kind of convergence is a clear case of analogy. Still another possibility in the behavioral domain is that similarity is explained by plastic developmental processes. In the clearest cases of plasticity, similarity can be explained entirely by convergent learning, perhaps shaped largely by task demands.[5] The main competition is thus between hypotheses of homology, analogy, and developmental plasticity. Insofar as they function as evidence, the criteria of homology should help pick out similarities between traits that are explained by continuity of traits across lineages and not convergent evolution or plastic developmental processes.

The most prominent criteria for homology were developed by Adolf Remane (1971) and can be deployed for this purpose. While these criteria were developed prior to cladistic approaches to homology,[6] they are fully consistent with such approaches. Specifically, there is no contradiction in supposing *both* that animals are best classified into monophyletic groupings that reflect historical divergences *and* that the criteria of homology can be used as evidence that animals share various traits because those traits existed prior to a historical divergence. Moreover, the criteria may be a useful compliment to cladistic approaches when difficulties arise, as in the case of cognitive homologies. Remane's criteria include *position*, *special quality*, and *continuity of intermediates*, each of which I describe briefly. The criterion of *position* applies to the radius and ulna bones (of tetrapods), because even with different forms and functions across different organisms, they retain their position relative to other bones of tetrapod forelimbs (humerus and the bones of the wrist). It would be highly unlikely for these characters to have evolved de novo in each of the different animals that possesses it and yet to have the same relative position to other structures. Moreover, there are few shared functions across different tetrapod species that would explain the distinctive correspondence. While corresponding position sounds like a spatial property, it is actually topological, and can include corresponding positions in temporal sequence or corresponding positions across cognitive architectures (e.g., "boxologies"). For instance, a typical boxological diagram in psychology may decompose a capacity or process (e.g., attention, memory) into distinct processes (short-term store, selective filter, etc.) and depict the flow of information from one process to another. The abstract structure of such a model could be shared in one or more organisms, or even duplicated within a single organism (as in the case of serial cognitive homologies), and this would apparently be evidence for homology, for the same reason that spatial position can be evidence for morphological homology.[7]

The criterion of *special quality* concerns "…shared features [that] cannot be explained by the role of a part in the life of the organism. The fact that in the vertebrate eye the blood supply to the retina lies between the retina and the source of light is a famous example of a 'special quality'" (Griffiths 2007, p. 648). If two characters are highly complex, then it is less likely that they would have independently evolved to have similar qualities. The location of blood supply to the vertebrate retina is both complex and nonessential (and even slightly counterproductive) given the functional role of the retina (what it is used for in the organism), so it identifies a correspondence that provides strong evidence that the various instances of this character derive from continuity across lineages.

Finally, the *continuity of intermediates* allows identification of homologous forms, A and C, because of the existence of one or more transitional states, B_1, …, B_n, between the two forms. In many cases, the homology between transitional forms, say between A and B_1 or between B_1 and B_2, is determined by applying the other two criteria. For instance, there are transitional states between the bones of the mammalian inner ear and the bones of the reptilian jaw. We know this because the bones of the reptilian jaw share the same *position* (relative to other bones of the jaw) as the bones of several intermediate forms, some of which share the same position as the bones of the mammalian inner ear.

In recent discussions, these criteria have been used to support relations of homology between cognitive systems in various species. For instance, García (2013) argues that processes for face recognition in humans and chimpanzees

[5] See Brown (2013) for a detailed discussion of the difficulties (e.g., due to the plasticity and transformability of behavior) in applying the criteria of homology to behavior.

[6] Thanks to an anonymous reviewer for raising the point about the historical precedence of the operational criteria.

[7] For a more detailed discussion of how the position criterion should be understood and assessed, see Love (2007) and Murphy (2012).

are cognitive homologs. Likewise, Clark (2009) argues that the human emotion of shame exists in two distinct forms, one of which is shared with other primates, the other of which may require distinctively human forms of self-awareness. Clark argues that these forms of shame are serial homologs, meaning that these distinct emotions derive from a single trait of an ancestral organism and are thus repeated structures within a single organism. Similarly, Murphy (2012) argues that various forms of imagery share common representational codes with non-imaginary representations (e.g., of visual or tactile objects) and that processes of imagination are thus serial homologs of first-order perceptual processes. Thus, it is clear that these criteria have already been fruitfully employed to begin decomposing human, primate, and even mammalian minds more broadly into units defined by homology.

Nevertheless, the fruitfulness of this approach is limited. Specifically, no extant account of cognitive homology addresses what I call the *boundary problem*, which requires an explication of the boundaries between distinct processes. This problem arises when, for instance, we ask of Clark: why should we think that there are serial homologs of shame as opposed to a single emotion that interacts with different mechanisms (some of which may be uniquely human) to produce distinct shame phenomena? Questions like this do not always pose problems for scientific research. However, such problems do arise when we take a closer look at competing hypotheses of homology.

Competing Hypotheses of Homology

I suspect that there are many instances of competing homology claims, but I focus here on competing claims concerning the human emotion of anger. Many emotion researchers and theorists have suggested that anger is an innate adaptation (Ekman 1999; Sell et al. 2009). Indeed, a wealth of research suggests that facial expressions of human anger appear early in development (even in children born deaf and blind) and that these expressions also appear across cultures (Ekman et al. 1969; Eibl-Eibesfeldt 1973; Izard 1994; Matsumoto and Willingham 2009). Moreover, some have suggested that similar facial expressions appear in other primates (Chevalier-Skolnikoff 1973; Parr and Waller 2006). This raises the question of which behaviors might be manifestations of anger in nonhuman animals. Given the tight link between anger and aggression in humans, some aggression researchers propose that innate patterns of aggression in nonhuman animals (mentioned in the previous section) are manifestations of anger. In other words, they propose that the systems responsible for these phenomena are homologous with human anger, meaning that these complex traits (human anger and one or more aggression systems of nonhuman animals) are derived from a common ancestral trait (and are thus continuous across the relevant lineages).

As plausible as this may sound, there have been two incommensurate proposals along these lines, and there has been little progress in adjudicating between them:

1. *Confrontation hypothesis*: human anger is homologous to the behavior program responsible for confrontational behaviors of rats.
2. *Defense hypothesis*: human anger is homologous with the system responsible for defensive aggression in cats (which arises from a neural system that includes the VMH, the amygdala, and parts of the brain stem).

The *confrontation hypothesis* holds that confrontational behaviors (mentioned in the previous section) observed in resident rats reflect "an underlying emotional state" that is a primitive version of anger (Blanchard and Blanchard 1984, 1988, 2003). This behavioral repertoire is set in opposition to the avoidance behaviors observed in intruder rats, which are thought to reflect fear. Moreover, the hypothesis holds that these two distinct emotional systems provide the best way of understanding angry aggression and fearful aggression (respectively) in humans.

By contrast, the *defense hypothesis* is that human experiences of anger "emerge" from a pan-mammalian brain system (including the VMH and amygdala among others) that produces defensive behaviors when electrically stimulated (Panksepp 1998; Panksepp and Zellner 2004; Blair 2012; Panksepp and Biven 2012). As mentioned in section two, these behaviors are set in opposition to predatory behaviors, which are neurally dissociable from the defensive behaviors. In other words, this hypothesis holds that there are two neural systems for aggression, and that one of them, the defense system, provides the primary neural substrate for human anger and is the proximate cause of "the feeling states and behavioral acts" (Panksepp 1998, p. 14) distinctive of human anger.[8] In addition, Panksepp assumes throughout his work that this system is distinct from the neural system for fear. This is an important component of Panksepp's overall research program, the core of which are the claims that "The available evidence now overwhelmingly supports the conclusion that basic emotional processes emerge from homologous brain mechanisms in all mammals" (1998, p. 51) and that there are at least "…four primal emotional circuits [that] mature

[8] The focus on "feeling states" can be a distraction when considering the psychological traits of nonhuman animals. I think it is less tendentious to think of emotions as psychological entities that explain a cluster of visible symptoms (e.g., physiological arousal, facial expressions, motivation, etc.). In any case, this is what I intend when I talk about emotions like anger and fear.

soon after birth, as indexed by the ability of localized brain stimulation to evoke coherent emotional displays in experimental animals, and these systems appear to be remarkably similarly organized in humans" (1998, p. 52). While he does point out that "there is considerable overlap and hence neural interaction among systems" (1998, p. 53), his apparent assumption is that each of these neural systems are distinct (see especially Fig. 3.5 on p. 53) at some important level of analysis. While some of the proponents of the defense hypothesis do not share these aims, there is broad agreement among them that we can best understand certain types of human aggression, namely impulsive and instrumental forms of aggression, in terms of dissociable neural systems for defense and predation, respectively. Whereas the *confrontation hypothesis* identifies distinctions between psychological categories through patterns of behavior, the *defense hypothesis* identifies these distinctions through neural dissociations.

Importantly, these hypotheses are incompatible. Within the neurophysiological tradition, the neural dissociation between predatory and defensive aggression is the main reason to consider them fundamental, distinct categories of aggression. However, confrontation and avoidance behaviors in rats do not exhibit this kind of clean neural dissociation (Siegel 2004, Chap. 1), or what evidence there is suggests a neural substrate for confrontational aggression that is distinct (and perhaps also dissociable) from the substrate for defense and predation both (see, e.g., Adams and Boudreau 1993; Canteras 2002). While defensive and predatory aggression have been elicited in rats by EBS in the hypothalamus (in roughly homologous brain regions), neither form of aggression is identical with the confrontational aggression observed in ethological research (Kruk 1991). Confrontational aggression is distinct from predatory aggression in the following respects: Even though predatory and confrontational attacks both involve back biting, predatory aggression in rats (elicited by EBS) is only directed at rat pups and mice, and usually involves "killing bites" to the neck (Woodworth 1971). By contrast, patterns of confrontational aggression are mostly directed at uncastrated adult males (as opposed to females or rat pups), and are nonlethal. Thus, it appears to be aligned with phenomena of intermale competition, which appear to be distinct from phenomena of predation (in part because of these behavioral differences).

Moreover, confrontational aggression is distinct from defensive aggression in these respects: Many of the defensive attacks induced by EBS in the VMH do not specifically target the back and many are accompanied by alarm calls or escape behaviors (Lammers et al. 1988). By contrast, confrontational aggression targets the dorsal surfaces of the intruder's neck and back and is accompanied by approach behaviors, and threat signals. In other words, the aggression phenomena identified by these different research programs are behaviorally distinct, and distinct neural mechanisms seem to be responsible for them.

As a result, these proposals make incompatible inferences about what anger is and also about which aggression phenomena are its manifestations. According to the confrontation hypothesis, anger in humans is responsible for aggression in response to *conspecific challenge* (specifically from an "intruder" or a subordinate), which we should expect to be distinct from fearful aggression in humans (e.g., aggression as a response to a life-threatening situation or a challenge from a formidable opponent). According to the defense hypothesis, anger in humans is responsible for impulsive aggression more broadly, which includes aggression in response to serious *threats* in addition to challenges (see, e.g., Blair 2012). In other words, these proposals make incommensurate inferences about the nature of angry behavior in humans and other animals. Importantly, part of the background of this disagreement is broad agreement that anger is a separate emotion from fear. Moreover, the main proponents of each hypothesis also defend putative homologies for fear (Blanchard and Blanchard 1984, 2008; Panksepp 1998, Chap. 11). It follows from this that if either the defense or confrontation system is homologous with anger, then it cannot be homologous with fear.

Conflating Kinds: The Defense Hypothesis

While proponents of these hypotheses aim to identify homologies, there has been little progress in forging a consensus. I think this is a symptom of a deep epistemic problem for assessing cognitive homologies. This is not because of any expectation that consensus should be reached quickly. Rather it is because the case for the defense hypothesis has been strikingly tenuous from the beginning. We can see this by evaluating the evidence in favor of the defense hypothesis. Since Panksepp proposed the defense hypothesis, there has been increasing evidence that the defense system (what he calls the RAGE system) is not distinct from fear and may very well be a proper part of the fear system. If so, then the defense hypothesis amounts to a conflation of two putatively distinct psychological kinds in cats, anger and fear. I suggest that a mistake of this kind would be almost impossible in the domain of morphology. It arises primarily because of the *boundary problem*: we cannot directly observe boundaries between instances of psychological kinds.

Before I make this case, I will first say something more about what I take to be the target of comparison for both hypotheses of homology concerning anger. I take it that the appropriate target of the confrontation and defense

hypotheses is *basic human anger*, the cluster of properties (e.g., physiological, endocrine, and postural changes) associated with involuntary facial expressions of human anger (Ekman 2003, Chap. 6; Izard 2007; Ekman and Cordaro 2011). To briefly defend this choice, this is the most closely studied set of "anger" phenomena the structure of which is likely explained by biological inheritance (as opposed to cultural inheritance or similar selection regimes, etc.), therefore it is the most plausible target for homology claims. This is because homology claims identify traits across taxa that are continuous across lineages in large part due to processes of biological inheritance. As suggested above, biological inheritance appears to be one of the causal homeostatic mechanisms that preserve the structure of homologous traits across lineages (cf. Assis and Brigandt 2009; Brigandt 2009). Thus, if there is something like anger in nonhuman animals, then it is most likely to correspond with phenomena in humans that are explainable by inheritance, namely the basic emotion of anger.

The defense system bears some similarities with human anger that seem to satisfy the criteria of homology. First, there may be some evidence for *continuity across intermediates*: stimulation of the hypothalamus of fish, lizards, chickens, opossums, cats, dogs, rats, and marmosets leads to defensive forms of attack (Roberts et al. 1967; Bergquist 1970; Panksepp 1971; Woodworth 1971; Lipp and Hunsperger 1978).[9] In macaques, ventromedial hypothalamic stimulation also results in attack under certain conditions (Alexander and Perachio 1973), some of which depend on whether the EBS occurs in the presence of a higher or lower ranking conspecific (attack being more likely in the latter case).

Importantly, the criteria of homology are focused on internal properties of an entity rather than its relationships to external entities, and the effects of brain stimulation are not obviously internal to the entity in question,[10] the neural substrate for the defense system. So the facts about EBS to the VMH by themselves are not complete evidence for continuity of intermediates. Rather, this evidence should be integrated with information about how other components of the defense system interact with the VMH. For instance, stimulation of the medial amygdala (a putative component of the defense system) in cats can potentiate defensive attacks elicited by EBS in the hypothalamus (e.g., Shaikh et al. 1993). If there is evidence for a similar relationship (potentiation of EBS-induced attack by the amygdalae) in other species, then this would further strengthen claims about continuity of the defense system as a whole.

There are some hints that continuity obtains with humans. For instance, there is a handful of case studies concerning hypothalamic stimulation in humans (with Parkinson's disease or obsessive compulsive disorder) where aggression has been elicited by EBS (see below). Moreover, in some neuroimaging studies, anger induction (e.g., through remembering an angering event) has been correlated with hypothalamus activation (e.g., Damasio and Grabowski 2000). There is also evidence that amygdala stimulation can produce feelings of anger in humans (e.g., Hitchcock and Cairns 1973). One might take this as preliminary evidence that human anger is subserved by some of the same pathways that implement defensive behaviors.

Second, consider the criterion of *position*. As with the offensive attack observed in ethological work, physiological arousal and threat signals do occur prior to defensive attacks elicited by EBS (e.g., Stoddard-Apter et al. 1983). The cat's defensive posture and facial expressions also bear an apparent similarity to anger, though this similarity has not been analyzed further. While few call on this as evidence for the hypothesis, I suspect that this similarity is part of what led scientists to call this behavioral syndrome "defensive rage," which may have subsequently colored the way the phenomenon was perceived.

However compelling these similarities may seem, they are actually quite flimsy as evidence for homology. This becomes apparent when we look at these results in the context of a larger body of work concerning the hypothalamus and aggression elicited by brain stimulation. First consider humans. The region that is associated with aggression in humans, the triangle of Sano, is not specific to the VMH, but instead overlaps with the ventromedial and dorsolateral hypothalamus (the area associated with predatory aggression) and includes areas between the posterior hypothalamus and subthalamic nucleus (Sano et al. 1970; Bejjani et al. 2002; Rosa et al. 2012; Franzini et al. 2013).[11] Surgical lesions (as well as continuous EBS) within the triangle of Sano region have been shown to abolish abnormal aggressive behaviors, but most of these areas are centered around the posterior hypothalamus

[9] Delgado (1968) produced aggressive behaviors with electrical stimulation of the thalamus and cerebellum of chimpanzees and macaques. However, these brain structures are notably absent from the defense hypothesis and its descriptions of brain structures involved in aggression. While Delgado and colleagues did observe facial expressions during attacks, these facial expressions were not analyzed.

[10] Thanks to an anonymous reviewer for pointing this out.

[11] Panksepp and Biven (2012, p. 150) seem to contradict this by saying, "If these kinds of brain-stimulation procedures are carried out in human beings, people tend to clench their jaws and to report feelings of intense anger (King 1961; Hitchcock and Cairns 1973; Mark et al. 1972)." However, one can easily see that each of these studies involves stimulation of the amygdala rather than the hypothalamus. Moreover, stimulation of the amygdala produces a multitude of emotional experiences and behaviors, as I discuss in this section.

(Rosa et al. 2012; Franzini et al. 2013). Other regions in the triangle of Sano that produce aggression via EBS are areas outside the hypothalamus, between the subthalamic nucleus and the posterior hypothalamus (Bejjani et al. 2002). In sum, while this research suggests a connection with aggression in humans and activity in and around the hypothalamus, it does not show localization to the VMH (or even the hypothalamus more generally).

Nor does stimulation of the VMH in humans provide any clear evidence that areas within this region are part of a neural system specific to anger, as some proponents of the defense hypothesis might predict.[12] Stimulation of the VMH in humans sometimes leads to panic and the feeling of being chased (Wilent et al. 2010). Similarly, stimulation of the amygdala during brain surgery induces several other emotional experiences besides anger, including anxiety, guilt, embarrassment, jealousy, and a "desire for flight or escape" (Hitchcock and Cairns 1973). The latter feelings are usually associated with fear (see, e.g., Frijda et al. 1989). Thus, neuroscientific research does not unambiguously support the defense hypothesis. For all this research, there is as yet no reason to think that anger is uniquely associated with regions in the VMH or the amygdala in humans, because it remains possible that the aggression produced by stimulation of these areas is associated with fear or emotion more broadly (including, e.g., shame, disgust, guilt, depression, etc.).

The last 15 years of neurophysiological research in rats is also telling. Some of this research demonstrates that in rats, defensive behaviors are associated with some of the very neural circuits that are specialized for anti-predator responses (i.e., tonic immobility, freezing, flight, and fight). In an influential study, Canteras and Chiavegatto (1997) exposed rats to a natural predator (a cat), and observed subsequent Fos immunoreactivity (a well-known indicator of preceding neural activity) in the hypothalamus and surrounding area. This revealed activation in a constellation of sites, including among others the VMH (*dorsomedial* aspect), the anterior hypothalamic nucleus, the dorsal premammillary nucleus, and importantly, the perifornical region. Similar research on reproduction and agonistic behavior (including confrontation behaviors) reveals Fos immunoreactivity in distinct regions of male rat brains—the VMH (*ventrolateral* aspect, see also Fig. 2c), the medial preoptic nucleus, and the ventral premammillary nucleus (Kollack-Walker and Newman 1995)—after either mating or intermale competition (a resident-intruder confrontation with another male rat). Subsequent work supports the claim that there are two distinct systems here, one that underpins innate defensive responses to predators and another that underpins innate reproductive and agonistic responses to conspecifics

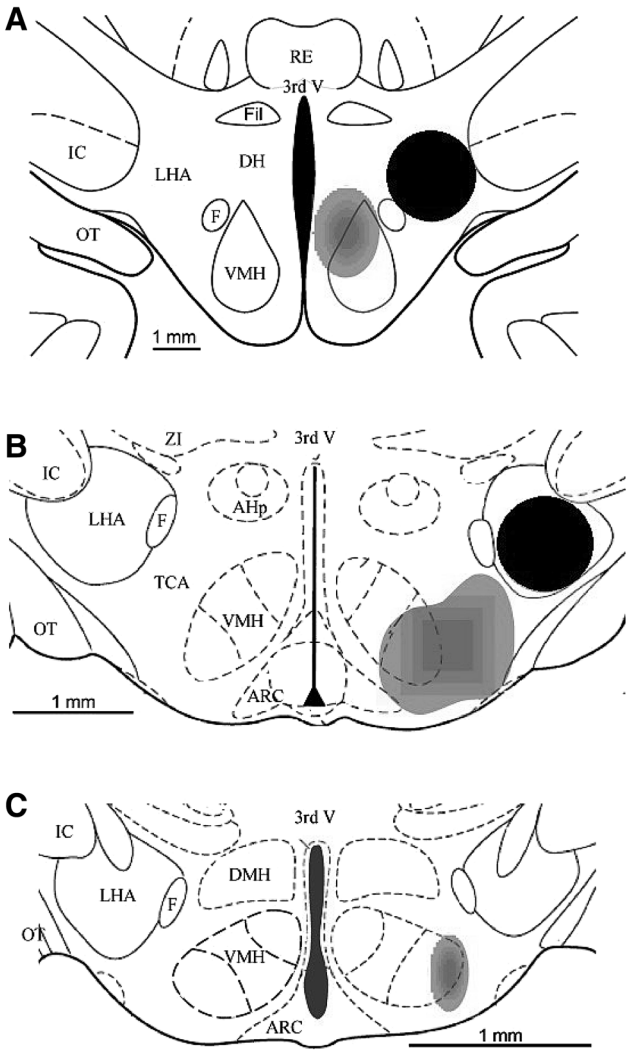


Fig. 2 Hypothalamic areas associated with aggression. **a** Sites at which EBS elicits defensive rage (*gray shading*) and predatory attack (*round black shading*) in cats (based on Siegel et al. 1999). Note that the *gray* is located on the dorsomedial aspect of the VMH, and thus corresponds more closely with the regions associated with anti-predator behavior in rats (cf. Canteras and Chiavegatto 1997). **b** Sites from which EBS elicits conspecific biting attacks (*gray shading* directed at the back, head and neck regions) and mouse killing attacks (*round black shading*) in rats (based on Lammers et al. 1988; Woodworth 1971). **c** Sites from which optogenetic stimulation elicits biting attacks (usually directed at the back) on both male and female intruders in mice (based on Lin et al. 2011). Note that this is the ventrolateral aspect of the VMH. Abbreviations (excerpted from Haller 2013): *3rd V* 3rd ventricle, *AHp* anterior hypothalamic nucleus, posterior part, *ARC* arcuate nucleus, *DH* dorsal hypothalamic area, *DMH* dorsomedial hypothalamic nucleus, *F* fornix, *Fil* nucleus filiformis, *IC* internal capsule, *LHA* lateral hypothalamic area, *OT* optic tract, *PVN* paraventricular hypothalamic nucleus, *RE* nucleus reuniens, *VMH* ventromedial hypothalamic nucleus, *TCA* tuber cinereum area, *ZI* zona incerta From Haller (2013), p. 99 © Elsevier. Originally published in *Brain Research Bulletin*

[12] This is predicted under Panksepp's assumption that distinct neural systems underpin distinction emotions (e.g., RAGE and FEAR), as discussed above. Thanks to two anonymous reviewers for pressing me to clarify this point.

including confrontation and avoidance behaviors (for a review, see Canteras 2002).

These results call into question whether research on defensive rage in cats is really focused on a psychological state like anger in humans, which is supposed to be distinct from fear and anti-predator behaviors. In cats, it is primarily the dorsomedial aspect of the VMH that causes defensive rage, the very same region in rats that becomes active after exposure to a predator (see Fig. 2a). If this area were specific to anger across a range of species, one would not expect it to be associated with anti-predator responses in rodents, as Canteras and others have observed. Moreover, EBS-induced defensive attacks in rats (which have been compared with defensive rage in cats) are clearly part of the repertoire of anti-predator behaviors, e.g., flight and freezing behaviors that are commonly associated with fear. For example, it is well known that rats will mount leaping attacks toward cats once they are cornered or in close proximity (usually after freezing and flight responses have been ineffective, Blanchard and Blanchard 2008, p. 66). Moreover, these attacks appear to serve the purpose of creating a window for escape and avoidance. Accordingly, subsequent research has shown that these defensive attacks can also be elicited by stimulation of the perifornical region (see references in Roeling et al. 1994), a region in which stimulation (whether electrical or chemical) can elicit either escape or antipredator attack behaviors (in rats). At first glance, the defense hypothesis would not necessarily predict that multiple areas associated with defensive attacks would also be associated with antipredator behavior.[13] These observations are better predicted by the hypothesis that defensive behaviors are the products of fear or of systems for predator avoidance, both of which are supposed to be distinct from anger. Given the bulk of this research, defensive aggression is more likely to be a context-dependent component of the fear response in rats as a wealth of other work suggests (e.g., Adams 2006; Blanchard and Blanchard 2008). At the very least, the current evidence clearly does not support a unique localization within the VMH for defensive aggression *as distinct from* components of fear (or antipredator) responses in rats, nor is there any evidence (to my knowledge) that defensive rage is distinct from antipredator responses in cats. If so, then defensive aggression (in both rats and cats) may very well be a product of fear rather than anger.

As it turns out, much of the work that Panksepp (Panksepp 1998; Panksepp and Zellner 2004; Panksepp and Biven 2012) cites in favor of the defense hypothesis is not linked to anger in any distinctive way. For instance, Panksepp references the work of Allan Siegel to support the defense hypothesis, but Siegel does not even advocate the defense hypothesis, and in many cases makes claims that contradict it. In several places (including Siegel 2004) Siegel compares defensive behaviors with a disorder known as episodic discontrol, which is marked by "…decreased impulse control—a characteristic common to defensive behavior—and altered perceptual states following stimuli evoking *anger, fear or rage*" (Siegel and Victoroff 2009, p. 213; emphasis mine). Indeed, many of the similarities that are noted between defensive behaviors and these forms of human aggression are characteristics of affectively driven behavior in general. Impulsivity is a characteristic of many kinds of emotion expression (see, e.g., Frijda 1986, 2010), including fear, anger, disgust, sadness, and joy.

So it looks as if the similarities between the defense system and human anger may be only apparent. In reality, the evidence is weak that this system is distinct from a neural system (or systems[14]) for fear, and in fact, there never was any such evidence. At this juncture, it seems much more likely that the defense system is simply part of the fear system.

Importantly, this discussion has been operating under the assumptions that (a) homologous emotional states remain tied to homologous brain structures across lineages, and that (b) the relevant neural regions are emotion-specific and dissociable. Both are claims to which Panksepp is clearly committed (cf. the discussion above). Indeed, the main body of his work in affective neuroscience appears to be an extended defense of these claims (or perhaps an outgrowth of these assumptions). Nevertheless, these assumptions stand in tension with what I said previously about cognitive homologies. Such homologies are homologies of function (Love 2007) that need not be linked to homologous structures (cf. Ereshefsky 2007) and that could potentially be realized by some of the same neural components. Accordingly, one *can* distinguish the defense

[13] Panksepp does try to explain why neural systems for fear and anger should be "interdigitated": "It makes good evolutionary sense for FEAR and RAGE circuits to be intimately related for one of the functions of anger is to provoke fear in competitors, and one of the functions of fear is to reduce the impact of angry behaviors from threatening opponents" (1998, p. 208). While this may explain the functional relationships between anger and fear, it remains puzzling how this would explain their neuroanatomical relationship, which is what "interdigitation" apparently refers to. That is, why and how would *inter*-organismal interactions predict close *intra*-organismal neuroanatomical organization? Regardless of whether there is a sensible answer to this question, interdigitation does not cancel functional, and perhaps also physical separateness of these neural systems for Panksepp. The next sentence reads as follows: "Although it has not been empirically demonstrated, it is reasonable to suppose that at low levels of arousal, the two systems are mutually inhibitory…."

[14] There is now some evidence in rats to distinguish fear of predators and fear of conspecifics (Gross and Canteras 2012).

hypothesis as Panksepp articulates it from other versions of the hypothesis, which relax Panksepp's two assumptions.

Nevertheless, the case for the defense hypothesis is not substantially strengthened by relaxing these assumptions. First, if one relaxes the assumption (a) that emotions are tied to homologous brain areas, the hypothesis still fails to predict extensive overlap between regions that elicit defensive attack as well as other antipredator responses in rats. On the contrary, relaxing this assumption actually diminishes the testability of the defense hypothesis by making it less committed to specific predictions about the effects of EBS. To see this, consider that the modified hypothesis would look something like this: human anger is homologous with the system responsible for defensive aggression in cats, which arises from a neural system that includes the VMH, the amygdala, and parts of the brain stem in cats *but may or may not arise from these regions in humans*. If we cannot assume that homologous emotions track homologous brain regions, we cannot make clear predictions about what emotional outcomes we should expect EBS to have in homologous brain regions in humans. If so, then the results of EBS cannot provide substantial support for this theory over a range of other theories. At least, this remains true unless the hypothesis is supplemented by plausible assumptions about which kinds of shifts in function brain regions will undergo over evolutionary time and phylogenetic space.

Second, one could relax the apparent assumption (b) that a brain area such as the VMH is associated with a single cognitive function. Relaxing this assumption seems plausible because, as Anderson (2007, 2010) argues, there is considerable evidence that different neural systems get deployed for multiple different functions depending on task demands. Moreover, one might expect that more ancient brain regions (VMH being quite ancient) are likely to be deployed for a greater number of different functions. As applied to the defense hypothesis, we might imagine that certain constituents of the defense system can also be deployed to serve separate functions, such as the production of other emotional states. If either of these assumptions is relaxed, then the defense hypothesis no longer predicts that neural substrates for anger will be distinct from the substrates for other emotions or that areas like the central amygdala or VMH will be specific to anger. Nevertheless, in that case the defense hypothesis also loses a good deal of its testability. For one, there would be few interesting predictions to make concerning the effects of EBS on its own. Rather, we might expect effects of EBS to be radically context specific (especially in humans), since activity in a given region could interact with a number of different functions that may or may not be engaged in a single context. The overall point is that hypothesis testing requires that theories make competing predictions. But relaxing either of these two assumptions undercuts the ability of the defense hypothesis to make clear predictions that distinguish it from the confrontation hypothesis, and hence undercuts the evidential value of EBS data in humans and rats.

So why did the defense hypothesis seem compelling in the first place? Part of the problem is that homologies are more difficult to assess in the cognitive domain than in the morphological domain. Recall that the boundaries between psychological units within a given organism are typically inferred, rather than being directly observed. This results in the boundary problem: that it is difficult to determine the boundaries between instances of psychological states. The same cannot be said of skeletal homologies or the developmental units (e.g., limb buds, gill arches) from which they arise. While instances of morphological units (and the boundaries that may separate them) may not all be directly observable, they are a good deal closer to the observable end of the spectrum than instances of psychological kinds. The latter are perhaps one of the clearest cases of unobservable entities. To be clear, the point is not that the existence of psychological *kinds* are inferred (as is the existence of abstract morphological units), rather it is that the existence of *instances* of psychological kinds is inferred in a wide range of cases (though I do not entirely rule out the possibility that instances can be directly perceived). By contrast, in the domain of morphology, even variables that are less directly observable, such as distinct patterns of gene expression in development, appear to be tightly linked to distinct developmental/morphological units, such as distinct limb buds or distinct gill arches. This makes competing claims of morphological homology much easier to adjudicate in the domain of morphology.

If we consider cases in which there was uncertainty, perhaps even competition, between hypotheses of morphological homology, we can see that these issues are often resolved by the careful study of developmental structures and the processes by which they differentiate into distinct skeletal structures. For instance, resolution of the cases of mammalian ossicle homology and avian digit homology (e.g., Takechi and Kuratani 2010; Wagner 2005, respectively) depended on a kind of continuity of *developmental* intermediates assessed by a combination of methods like embryo dissection, evaluating the effects of selective gene knockouts, and detecting patterns of gene expression.[15] While some of these methods are theory-laden, they could not proceed without access to clearly observable boundaries between instances of morphological units, such as the boundaries between bone structures or the boundaries between limb bud and body cavity. These are exactly the

[15] Thanks to Alan Love for pointing out the relevance of these examples.

kind of observable boundaries that are lacking in the psychological domain.

Now, if the neural system for defensive aggression is really homologous with fear in humans, as I have suggested, then the confusion here is actually on par with confusing the anterior and posterior limb buds on an embryo or the radius and the ulna on a tetrapod skeleton. This kind of confusion is hardly even possible in the morphological domain, precisely because we are able to directly observe the boundaries between instances of morphological units (e.g., posterior and anterior limb buds, radius and ulna). Were it possible to directly observe the cause of defensive aggression in cats and observe its relation (or lack thereof) to other causes of feline aggression, then questions would immediately arise about which (if any) cause of aggression bears *distinctive* similarities with human anger. In the morphological case, one is able to notice these boundaries without making any prior inferences about them, and this knowledge inevitably guides the identification of homologs across various species.

A Constraint on Homology Claims

As a result, it is worth making explicit the role that this knowledge plays in homology inferences. I suggest that this knowledge constrains the kind of similarities admitted as evidence for homology. The relevant constraint derives from the fact that there are different homology relations at different levels of generality. To see this, notice first that each general homology captures all the traits (of various organisms) that are continuous and correspondent at some level of a hierarchy (cf. Ramsey and Peterson 2012) . For instance, all tetrapods with forelimbs have forelimbs with a similar structure, and this is because each tetrapod forelimb is continuous with the forelimb of the common ancestral tetrapod (the first population of vertebrates to live on land) and numerically identical across the range of ancestors that link each lineage to the ancestral tetrapod. As a result, tetrapod forelimbs form a homology class. Moreover, the class of tetrapod forelimbs is nested within a larger, more general homology class, the class of paired appendages. This is because the forelimb of the ancestral tetrapod itself was derived from (and hence continuous with) the paired appendages of its common ancestor with gnathostomes more broadly.[16] Accordingly, the forelimbs of amphibians, reptiles, avians, and mammals are members of the homology class of tetrapod forelimbs, but they are also members of the more general, inclusive homology class of paired appendages, which also includes the pectoral fins of sharks and bony fish, among others.[17] Thus, a given homology class can be nested within a broader homology class.

Importantly, while pectoral fins *are* homologous with instances of tetrapod forelimbs *as paired appendages*, the similarities between pectoral fins and tetrapod forelimbs *do not* provide good evidence for homology in the less inclusive class of *tetrapod forelimbs*. Inclusion in this more specific class is indicated by bone structures (e.g., radius and ulna) that are absent in pectoral fins. These structures are due to modifications that occurred subsequent to the divergence of tetrapods from teleosts, and that is why teleost pectoral fins are not included in this homology class.

Consequently, we can see that some similarities only indicate inclusion in a broader homology class (e.g., paired appendages), whereas other similarities indicate inclusion in narrower homology classes (e.g., tetrapod forelimbs). It follows that, when evaluating similarities between traits, it is sometimes necessary to consider which homology class a similarity indicates. Moreover, when one identifies a specific homology class, similarities can only count as evidence for inclusion in that class if the following constraint is met.

> *Class-specificity constraint*: To provide evidence for relations of homology relative to homology class G as opposed to the more inclusive class, H, requires that some similarities between relata are not shared by traits in the more inclusive class, H.

This constraint limits the evidence for homology in relation to a specific homology class at a specific level of generality. The constraint captures why the similarities between human forelimbs and the dolphin pectoral fin indicate membership in the homology class of tetrapod forelimbs while similarities with the shark pectoral fin do not.

This constraint might easily be confused with a rule to avoid conflating symplesiomorphies and synapomorphies. Put simply, synapomorphies are characters that are shared by members of monophyletic group, a group that includes all and only the descendants of a common ancestor. By contrast, symplesiomorphies are similarities shared only between paraphyletic groups, such as the similarities between shark fins and teleost fish fins (which are not shared by all the

[16] This clade includes animals descended from cartilaginous (chondrichthyes) and bony fish (osteichthyes), the latter of which were the ancestors of the tetrapods.

[17] Another way of putting this point is to distinguish between characters and character-states, where characters are determinables and character-states are determinates (cf. Brigandt and Griffiths 2007, p. 635). For instance, animals can possess or lack paired appendages (character), and this character can appear in different states within a lineage (ray fin, tetrapod forelimb, etc.). This characterization too is subject to levels of generality because something that is a character-state at one level of generality (e.g., tetrapod forelimb) is a character with distinct character-states (e.g., amphibian forelimb, avian forelimb, mammalian forelimb, etc.) at a lower level of generality.

common ancestors of chondrichthys and osteichthys). The class-specificity constraint does not function to distinguish characters in this way. If we note the similarity in pectoral fins, the constraint does *not* tell us whether the similarity is a symplesiomorphy or a synapomorphy. Rather, the constraint says that *if* this character indicates homology (which it may), it indicates homology only at the broadest level at which relevant similarities obtain. Let us consider the relevant similarities. The main similarity in pectoral fins is that the left and right fins are symmetrical, but there are also similarities in internal structure (e.g., divisions between axial and radial portions) and perhaps also developmental mechanisms. The question posed by the constraint is whether these similarities only obtain between ray fin fish and cartilaginous fish. If the similarities obtain more broadly (we should predict that most of them will), then the indicated homology class is more inclusive than ray fin fish and cartilaginous fish. After all, there *is* a relationship of homology between pectoral fins of these species, it is just not exclusive to these two clades. Specifically, paired appendages are shared by the entire monophyletic group of gnathostomes (with the exception of species in which paired appendages were lost, as in serpentes, cetacea, and caecilia). As a result, the class-specificity constraint allows finer-grained distinctions between traits than the distinction between synapomorphies and symplesiomorphies, since it operates within monophyletic groups and can distinguish between two or more synapomorphies. For instance, it is easy to see how it could distinguish between tetrapod forelimbs (shared by dolphins and humans, for instance) and paired appendages (shared by dolphins, humans, and sharks) given the similarities that they share.

The examples so far deal straightforwardly with morphology or body structure. Moreover, one might think that as it applies to these cases, the class-specificity constraint is so obvious as to make its articulation unnecessary. On the contrary, I suspect that what is obvious concerning morphology is easily confused concerning behavior or psychology. That is, one can find similarities that seem to indicate cognitive homology and more easily misidentify the homology class that this evidence concerns. In doing so, one violates the class-specificity constraint. I think this is one of the main reasons why the defense hypothesis has persisted without refutation over such a long period of time. After all, there are similarities between anger and fear that may indicate some kind of continuity. For instance, it is possible that they are serial homologs, which are structures or functions that are duplicated within a single organism (like the vertebrae or retinotopic maps in visual areas of the brain, respectively; see especially Murphy (2012)) and many of which may derive from a single trait in an organism's ancestors. Nevertheless, the current evidence suggests that the similarities between the defense system and human anger do not indicate homology at a level that includes anger and excludes fear (as both the defense hypothesis and the confrontation hypothesis propose).

What the class-specificity constraint requires then, is information regarding distinctness in each species being compared. In the example concerning shark and ray fin fish, the appropriate relation of homology is clarified by information about the internal structure of the fins, information that distinguishes the more from the less general class. Similarly, the appropriate relation of homology (or lack thereof) between the defense system in cats and anger in humans could be clarified by information that distinguishes a more general class (e.g., that includes all emotions) from a less general one (e.g., that includes only anger). While proponents of the defense hypothesis have attended to evidence for a distinction between defense and predation, they have not attended to evidence for a distinction between defense and fear.[18]

By contrast, consider the confrontation hypothesis. There is clear evidence for distinctness between characters that may correspond to anger and fear in humans. Manifestations of the confrontation and avoidance systems in rats can be distinguished by quantifiable differences in the facial expressions of residents and intruders (Defensor et al. 2012), just as manifestations of anger and fear in humans can be distinguished by their distinctive facial expressions (e.g., Ekman and Friesen 1971). Moreover, resident and intruder rats have distinct forms of attack with distinct target sites. Thus, it is possible to distinguish *within rats* at least two different patterns of affective behavior accompanied by distinct facial expressions.

On its own, this is not a particularly strong form of evidence in favor of this hypothesis. However, there are other similarities between confrontation behaviors in rats and angry behaviors in humans that are dissimilar to fearful behaviors in various respects. One interesting piece of evidence for homology is a special quality that is shared by rats and stumptail macaques. Adams and Shoel (1981) note that dominant macaques and resident rats both implement strategies aimed at accessing and biting the backs of subordinates.[19]

[18] However, ethologists have pointed to some evidence for a distinction between two patterns of intraspecific aggression: aggression in subordinates (which may correspond with defensive aggression) and "offensive" aggression in dominants (see Leyhausen 1979). This distinction may parallel that between confrontational and avoidant aggression in rats and a range of other mammals (see, e.g., Blanchard and Blanchard 1984). If so, then defensive aggression in cats may correspond with avoidant aggression.

[19] This may be an instance of a broader set of behaviors in mammals involving ritualized aggression that involves *target attack sites*. Target sites are usually accompanied by specialized defenses or weapons such as the lion's mane and the ram's horns. See Blanchard and Blanchard (1984) as well as Leyhausen (1979) for an extensive overview of target sites and ritualized aggression in a wide range of mammals.

In macaques, this behavior seems arbitrary with respect to the (probable) function of inflicting nonlethal damage on the subordinate. Macaques have a much larger repertoire of bodily movements than rats, many of which could serve the function of inflicting nonlethal harm (pushing, kicking, scratching, slapping, holding, etc.). Thus, back biting is a *special quality*, and the best explanation of this behavior may appeal to products of common ancestry (and thus phylogenetic continuity). In other words, the reason that the attacks of both rats and macaques are aimed at biting the neck and back may be that they share a common ancestor with a corresponding aggressive strategy and perhaps similar motivational mechanisms for negotiating conflict with conspecifics.[20]

At this point, no solid connection has been made with human anger. That is, we have no independent reason to think that dominant macaques attack subordinates out of anger. However, a tenuous case can be made on the basis of facial expressions. In other studies of macaque behavior, macaques with higher dominance status do display facial expressions (i.e., "stare," "round-mouthed stare", and "open-mouthed stare") toward lower-ranking macaques in aggressive encounters, expressions that resemble anger expressions in humans and are distinct from fear expressions (Chevalier-Skolnikoff 1974). Chevalier-Skolnikoff (1973) argues that two of these expressions are continuous (recruiting homologous facial muscles) across macaques, chimps, and humans. Some confirmation of these comparisons has been attained by using a facial action coding system to quantify chimpanzee facial expressions (Parr et al. 2007). Thus, there is tenuous evidence for *continuity across intermediates* between human anger and a putative confrontation system across the common ancestors of these species. Importantly, the similarities between human anger and the confrontation system in rats do not violate the class-specificity constraint on homology claims (relativized to a homology class that only includes the emotion of anger). Specifically, there is tentative evidence for aggression syndromes in rats, macaques, and humans that are distinct from fear (or rather traits that seem to correspond with fear) in each of these species. Some of the observed similarities between these "anger" syndromes (e.g., back biting in rats and macaques, facial expressions of anger in primates) are not shared with fear (or any other emotions for that matter). Thus, these similarities satisfy the operational criteria of homology in ways that are not also satisfied by other emotions like fear.

[20] Adams and Schoel actually argue for homology by considering similarity in the dynamics of attack and submission across both species. I do not find their argument very compelling.

Conclusion

While the case here is far from conclusive, we can derive some lessons from the contrast between the two hypotheses. First and foremost, to satisfy the class-specificity constraint, we need evidence of distinctness between characters and continuity of that distinction across intermediates. This kind of evidence is not as easy to come by in the domains of psychology and behavior by comparison with morphology. In the latter domain, the method of individuation can be as straightforward as boiling an animal's corpse. With this and other methods, we can directly observe the boundaries between radius and ulna, forelimb and hindlimb, in each species we want to compare. By contrast, in the psychological case we need a way of *inferring* distinct boundaries in each species we compare and ensuring that the *same* boundaries exist in each animal we compare. This problem may be even more pressing in the domain of nosology, where distinctions between behaviors or cognitive systems are sometimes essential for diagnosing distinct forms of dysfunction and where animal models of dysfunction are important tools for testing therapeutic interventions.

A second moral of this story is that homologies are rarely discovered piecemeal. This case of cognitive homology helps us to see that homologies sometimes must be assessed in pairs or larger n-tuples. If skeletal homology led us to believe otherwise, it was only because the boundaries between units are often so clearly observable. The inferences afforded by our visual systems work so well, their operations are almost invisible to us.

Third and less obvious, while homology may be the key to decomposing the mind into natural units, it is no silver bullet. Stereotyped behaviors (e.g., involuntary facial expressions), rigid behavioral goals (e.g., biting the back of a subordinate), and neural localization can provide some easy-to-infer anchors for assessing cognitive homology. Nevertheless, when goals become more diffuse, means more variegated, and soft assembled neural systems recruited (e.g., Anderson 2007), more tenuous inferences become necessary.

Consider an example: we clearly share certain appetites with nonhuman animals. We might easily conjecture that thirst, hunger, and lust are driven by highly conserved mechanisms across mammals and perhaps tetrapods more broadly. Nevertheless, if all we know is that a human male is cooking a gourmet meal, we need a good deal more information and inference to tease apart which aspects of this performance are means to satisfaction of hunger, lust, or even a secondary appetite for money. Insofar as evolved emotions include flexible motivational states, the causes of human aggression may be equally difficult to pull apart. If

all we know is that someone was verbally excoriated, we need a good deal more information and inference to assess whether fear, anger, jealousy, or even money were the motives behind the verbal abuse. The more flexible the motivational state, the more diffuse its connection with observable behavior, and the more difficult it is to homologize.

To sum up, I have suggested that competing claims of cognitive homology are symptomatic of a unique and unaddressed problem for homology thinking in the cognitive domain: the boundary problem. The problem is that the instances of many psychological kinds are not directly observable. One way to mitigate this problem is to supplement the operational criteria of homology with explicit use of the class-specificity constraint. In combination with the operational criteria of homology this constraint helps to specify what kind of evidence supports homology claims, namely, identification of *unique* correspondences that indicate common membership in a *specific homology class*; correspondences that provide evidence for common ancestry (or phylogenetic continuity more broadly) as opposed to common selection pressures (whether, cultural, developmental, or ancestral). Moreover, it motivates the search for boundaries between two or more candidate homologies within each organism being compared. With this constraint in hand, homology thinking can provide independent criteria for evaluating substantive disagreements concerning the nature of psychological kinds. Because this constraint was neglected, an ultimately flimsy hypothesis of homology, the defense hypothesis, has remained a serious competitor for far too long. Thinking carefully about homology in the cognitive domain and paying due respect to the class-specificity constraint can help to guide the future search for homologies and thus aid the project of decomposing the mind into natural units.

Acknowledgments Many thanks to those who have read through previous versions of this paper and provided helpful comments and feedback. This includes Caroline Blanchard, Carl Craver, John Doris, Marc Ereshefsky, Alan Love, Ron Mallon, Taylor Murphy, Anya Plutynski, and Michael Potegal.

References

Adams DB (1981) Motor patterns and motivational systems of social behavior in male rats and stumptail macaques—are they homologous? Aggress Behav. doi:10.1002/1098-2337(1981)7:1<5:AID-AB2480070104>3.0.CO;2-7

Adams DB (2006) Brain mechanisms of aggressive behavior: an updated review. Neurosci Biobehav Rev 30(3) 304–18. doi:10.1016/j.neubiorev.2005.09.004

Adams D, Boudreau W (1993) Offense produced by chemical stimulation of the anterior hypothalamus of the rat. Physiol Behav 53:1127–1132

Alexander M, Perachio AA (1973) The influence of target sex and dominance on evoked attack in rhesus monkeys. Am J Phys Anthropol 38:543–548

Anderson ML (2007) Massive redeployment, exaptation, and the functional integration of cognitive operations. Synthese 159(3):329–345. doi:10.1007/s11229-007-9233-2

Anderson ML (2010) Neural reuse: a fundamental organizational principle of the brain. Behav Brain Sci 33:245–313

Assis LCS, Brigandt I (2009) Homology: homeostatic property cluster kinds in systematics and evolution. Evolut Biol 36(2):248–255. doi:10.1007/s11692-009-9054-y

Bejjani BP, Houeto JL, Hariz M et al (2002) Aggressive behavior induced by intraoperative stimulation in the triangle of Sano. Neurology 59(9):1425–1427. doi:10.1212/01.WNL.0000031428.31861.23

Bergquist EH (1970) Output pathways of hypothalamic mechanisms for sexual, aggressive and other motivated behaviors in opposum. J Comp Physiol Psychol 70:389–398

Blair RJR (2012) Considering anger from a cognitive neuroscience. Cognit Sci 3:65–74. doi:10.1002/wcs.154

Blanchard DC, Blanchard RJ (1984) Affect and aggression: an animal model applied to human behavior. In: Blanchard RJ, Blanchard DC (eds) Advances in the study of aggression, vol 1, pp 1–62

Blanchard DC, Blanchard RJ (1988) Ethoexperimental approaches to the biology of emotion. Annu Rev Psychol 39:43–68

Blanchard DC, Blanchard RJ (2003) What can animal aggression research tell us about human aggression? Horm Behav 44(3):171–177. doi:10.1016/S0018-506X(03)00133-8

Blanchard DC, Blanchard RJ (2008) Defensive behaviors, fear, and anxiety. Handb Behav Neurosci 7339(07):5–7. doi:10.1016/S1569-7339(07)00005-7

Blanchard DC, Litvin Y, Pentkowski NS, Blanchard RJ (2009) Defense and aggression. In: Berntson GG, Cacioppo JT (eds) Handbook of neuroscience for the behavioral sciences. Wiley, Sussex, pp 958–974

Brigandt I (2009) Natural kinds in evolution and systematics: metaphysical and epistemological considerations. Acta Biotheor 57(1–2):77–97. doi:10.1007/s10441-008-9056-7

Brigandt I, Griffiths PE (2007) The importance of homology for biology and philosophy. Biol Philos 22(5):633–641. doi:10.1007/s10539-007-9094-6

Brown RL (2013) Identifying behavioral novelty. Biol Theory. doi:10.1007/s13752-013-0150-y

Canteras N, Chiavegatto S (1997) Severe reduction of rat defensive behavior to a predator by discrete hypothalamic chemical lesions. Brain Res Bull 44:297–305

Canteras NS (2002) The medial hypothalamic defensive system: hodological organization and functional implications. Pharmacol Biochem Behav 71(3):481–491

Chevalier-Skolnikoff S (1973) Facial expression of emotion in nonhuman primates. In: Cole J (ed) Darwin and facial expression: a century of research in review. Academic Press, New York, pp 11–89

Chevalier-Skolnikoff S (1974) The ontogeny of communication in the stumptail macaque (macaca arctoides). Karger, Basel

Clark J (2009) Relations of homology between higher cognitive emotions and basic emotions. Biol Philos 25:75–94. doi:10.1007/s10539-009-9170-1

Damasio A, Grabowski T (2000) Subcortical and cortical brain activity during the feeling of self-generated emotions. Nat Neurosci 3:1049–1056

Defensor EB, Corley MJ, Blanchard RJ et al (2012) Facial expressions of mice in aggressive and fearful contexts. Physiol Behav 107(5):680–685. doi:10.1016/j.physbeh.2012.03.024

Delgado JMR (1968) Offensive-defensive behaviour in free monkeys and chimpanzees induced by radio stimulation of the brain. In:

Sigg EB, Garattini S (eds) Aggressive behavior. Wiley, New York, pp 109–119
Eibl-Eibesfeldt I (1973) The expressive behavior of the deaf-and-blind-born. Soc Commun Mov 163–194
Ekman P (1999) Basic emotions. In: Dalgleish T, Power M (eds) The handbook of cognition and emotion. Wiley, Sussex, pp 45–60
Ekman P (2003) Emotion revealed: understanding faces and feelings. Phoenix Press, New Haven
Ekman P, Cordaro D (2011) What is meant by calling emotions basic. Emot Rev 3(4):364–370. doi:10.1177/1754073911410740
Ekman P, Friesen WV (1971) Constants across cultures in the face and emotion. J Pers Soc Psychol 17:124–129
Ekman P, Sorenson ER, Friesen WV (1969) Pan-cultural elements in facial displays of emotion. Science 164:86–88
Ereshefsky M (2007) Psychological categories as homologies: lessons from ethology. Biol Philos 22(5):659–674. doi:10.1007/s10539-007-9091-9
Ereshefsky M (2012) Homology thinking. Biol Philos 27(3):381–400. doi:10.1007/s10539-012-9313-7
Franzini A, Broggi G, Cordella R et al (2013) Deep-brain stimulation for aggressive and disruptive behavior. World Neurosurg 80(3–4):S29.e11–S29.e14. doi:10.1016/j.wneu.2012.06.038
Frijda NH (1986) The emotions. In: The emotions, vol. 1. Cambridge University Press, Cambridge. doi:10.1093/0199253048.001.0001
Frijda NH (2010) Impulsive action and motivation. Biol Psychol 84(3):570–579. doi:10.1016/j.biopsycho.2010.01.005
Frijda NH, Kuipers P, ter Schure E (1989) Relations among emotion, appraisal, and emotional action readiness. J Person Soc Psychol 57(2):212–228. doi:10.1037//0022-3514.57.2.212
García CL (2010) Functional homology and functional variation in evolutionary cognitive science. Biol Theory 5(2):124–135. doi:10.1162/BIOT_a_00036
García CL (2013) La homología de sistemas cognitivos: el caso del reconocimiento de caras en primates humanos y no humanos. In: Gonzalez J, Davalos PK (eds) Ciencias cognitivas: entre la cooperacion y la integracion. UAQ/Editorial Porrua, Mexico City
Griffiths PE (1994) Cladistic classification and functional explanation. Philos Sci 61(2):206–227
Griffiths PE (1997) What emotions really are: the problem of psychological categories, vol 1997. University of Chicago Press, Chicago
Griffiths PE (2006) Function, homology, and character individuation. Philos Sci 73(1):1–25
Griffiths PE (2007) The phenomena of homology. Biol Philos 22(5):643–658. doi:10.1007/s10539-007-9090-x
Gross CT, Canteras NS (2012) The many paths to fear. Nat Rev Neurosci 13(9):651–658. doi:10.1038/nrn3301
Haller J (2013) The neurobiology of abnormal manifestations of aggression—a review of hypothalamic mechanisms in cats, rodents, and humans. Brain Res Bull 93:97–109. doi:10.1016/j.brainresbull.2012.10.003
Hitchcock E, Cairns V (1973) Amygdalotomy. Postgrad Med J 49(578):894–904
Izard CE (1994) Innate and universal facial expressions: evidence from developmental and cross-cultural research. Psychol Bull 115(2):288–299
Izard CE (2007) Basic emotions, natural kinds, emotion schemas, and a new paradigm. Perspect Psychol Sci 2(3):260–280. doi:10.1111/j.1745-6916.2007.00044.x
King H (1961) Psychological effects of excitation in the limbic system. In: Sheer DE (ed) Electrical stimulation of the brain. University of Texas Press, Austin, pp 477–486
Kollack-Walker S, Newman SW (1995) Mating and agonistic behavior produce different patterns of Fos immunolabeling in the male Syrian hamster brain. Neuroscience 66(3):721–736. doi:10.1016/0306-4522(94)00563-K
Kruk MR (1991) Ethology and pharmacology of hypothalamic aggression in the rat. Neurosci Biobehav Rev 15(4):527–538
Lammers JH, Kruk MR, Meelis M et al (1988) Hypothalamic substrates for brain stimulation-induced attack, teeth-chattering and social grooming in the rat. Brain Res 449(1–2):311–327
Leyhausen P (1979) Cat behavior: the predatory and social behavior of domestic and wild cats, 1st edn (Tonkin BA, trans). Taylor & Francis/Garland STPM Press, New York
Lin D, Boyle MP, Dollar P et al (2011) Functional identification of an aggression locus in the mouse hypothalamus. Nature 470(7333):221–226. doi:10.1038/nature09736
Lipp HP, Hunsperger RW (1978) Threat, attack and flight elicited by electrical stimulation of the ventromedial hypothalamus of the marmoset monkey. Brain Behav Evol 15:260–293
Love AC (2007) Functional homology and homology of function: biological concepts and philosophical consequences. Biol Philos 22(5):691–708. doi:10.1007/s10539-007-9093-7
Mark V, Ervin F, Sweet W (1972) Deep temporal lobe stimulation in man. In: Eleftheriou BE (ed) The neurobiology of the amygdala. Plenum Press, New York, pp 485–507
Matsumoto D, Willingham B (2009) Spontaneous facial expressions of emotion of congenitally and noncongenitally blind individuals. J Person Soc Psychol 96(1):1–10. doi:10.1037/a0014037
Matthen M (2007) Defining vision: what homology thinking contributes. Biol Philos 22(5):675–689. doi:10.1007/s10539-007-9088-4
Murphy TS (2012) Cognitive homology: psychological kinds as biological kinds in an evolutionary developmental cognitive science. University of Alberta, Edmonton
Owen R (1846) Lectures on the comparative anatomy and physiology of the vertebrate animals, vol 2. Printed for Longman, Brown, Green, and Longmans, London
Panksepp J (1971) Aggression elicited by electrical stimulation of the hypothalamus in albino rats. Physiol Behav 6(4):321–329. doi:10.1016/0031-9384(71)90163-6
Panksepp J (1998) Affective neuroscience: the foundations of human and animal emotions, 1st edn. Oxford University Press, New York
Panksepp J, Biven L (2012) The archaeology of mind: neuroevolutionary origins of human emotions. Norton, New York
Panksepp J, Zellner MMR (2004) Towards a neurobiologically based unified theory of aggression. Revue Int Psychol Soc 17(2):37–61
Parr LLA, Waller BBM (2006) Understanding chimpanzee facial expression: insights into the evolution of communication. Soc Cognit Affect Neurosci 1:221–228. doi:10.1093/scan/nsl031
Parr L, Waller B, Vick S et al (2007) Classifying chimpanzee facial expressions using muscle action. Emotion 7(1):172–181. doi:10.1037/1528-3542.7.1.172.Classifying
Putnam H (1969) Brains and behaviour. In: Block N (ed) Analysis, vol 30. Blackwell, Oxford, pp 1–19
Ramsey G, Siebels Peterson A (2012) Sameness in biology. Philos Sci 79(2):255–275. doi:10.1086/664744
Remane A (1971) Die grundlagen des natürlichen systems: der vergleichenden anatomie und der phylogenetik. Verlag, Koeltz Koenigstein
Rieppel O (2005) Modules, kinds, and homology. J Exp Zool Part B Mol Dev Evol 304(1):18–27. doi:10.1002/jez.b.21025
Roberts WW, Steinberg ML, Means LW (1967) Hypothalamic mechanisms for sexual, aggressive and other motivational behaviors in the oppossum didelphis virginiana. J Comp Physiol Psychol 64:1–15
Roeling TAP, Veening JG, Kruk MR et al (1994) Efferent connections of the hypothalamic "aggression area" in the rat. Neuroscience 59(4):1001–1024. doi:10.1016/0306-4522(94)90302-6

Rosa M, Franzini A, Giannicola G et al (2012) Hypothalamic oscillations in human pathological aggressiveness. Biol Psychiatry 72(12):e33–e35. doi:10.1016/j.biopsych.2012.06.007

Rosenberg A, Neander K (2009) Are homologies (selected effect or causal role) function free ? Philos Sci 76:307–334

Sano K, Mayanagi Y, Sekino H et al (1970) Results of stimulation and destruction of the posterior hypothalamus in man. J Neurosurg 33(6):689–707. doi:10.3171/jns.1970.33.6.0689

Sell A, Tooby J, Cosmides L (2009) Formidability and the logic of human anger. Proc Natl Acad Sci USA 106(35):15073–15078. doi:10.1073/pnas.0904312106

Shaikh MB, Steinberg A, Siegel A (1993) Evidence that substance P is utilized in medial amygdaloid facilitation of defensive rage behavior in the cat. Brain Res 625(2):283–294

Siegel A (2004) Neurobiology of aggression and rage, 1st edn. Informa Healthcare, London

Siegel A, Ta Roeling, Gregg TR, Kruk MR (1999) Neuropharmacology of brain-stimulation-evoked aggression. Neurosci Biobehav Rev 23(3):359–389

Siegel A, Victoroff J (2009) Understanding human aggression: new insights from neuroscience. Int J Law Psychiatry 32(4):209–215. doi:10.1016/j.ijlp.2009.06.001

Stoddard-Apter S, Levin B, Siegel A (1983) A sympathoadrenal and cardiovascular correlate of aggressive behavior in the awake cat. J Auton Nerv Syst 8:343–360

Takechi M, Kuratani S (2010) History of studies on mammalian middle ear evolution: a comparative morphological and developmental biology perspective. J Exp Zool Part B Mol Dev Evol 314 B(6):417–433. doi:10.1002/jez.b.21347

Wagner G (1996) Homologues, natural kinds and the evolution of modularity. Am Zool 36:36–43

Wagner GP (2005) The developmental evolution of avian digit homology: an update. Theory Biosci 124(2):165–183. doi:10.1016/j.thbio.2005.07.002

Wenzel J (1992) Behavioral homology and phylogeny. Annu Rev Ecol Syst 23(1992):361–381

Wilent WB, Oh MY, Buetefisch CM et al (2010) Induction of panic attack by stimulation of the ventromedial hypothalamus. J Neurosurg 112(6):1295–1298. doi:10.3171/2009.9.JNS09577

Woodworth CH (1971) Attack elicited in rats by electrical stimulation of the lateral hypothalamus. Physiol Behav 6(4):345–353