The Brain and Emotion

Adrienne S. Bonar[1] and Kristen A. Lindquist[1]

1. University of North Carolina at Chapel Hill

To appear in *Oxford Research Encyclopedia of Psychology*

Summary

Since antiquity, scholars have sought to understand emotions by studying their biological basis. Assumptions within affective neuroscience, the study of how the creates emotion, have changed over time as neuroimaging methods and computation improved. Models of the brain basis of emotion have also been influenced by the interplay of psychological theories that differ fundamentally in their philosophical approaches. One family of theories, typological views of emotion, map specific emotions to discrete brain structures and functions; in a different approach, constructionist views of emotion examine how the brain's structural and functional principles constrain how it can create emotions.

Initial investigations in affective neuroscience suggested that subcortical regions (e.g., amygdala, insula) processed a select few emotions (e.g., fear, disgust). However, with advancements in computational methods, researchers were able to demonstrate that emotions engaged large-scale and domain-general brain systems. More recently, these advancements led to the discovery that neural patterns associated with a particular emotion, like fear, shift due to contextual factors. This work challenges the assumption that fear, disgust, sadness, anger and so on can be distinguished from one another through observation of a specific pattern of neural activity; the patterns of activity observed for a specific emotion are highly variable. In the wake of these technological and theoretical advancements, affective neuroscience might be in the midst of a paradigm shift—where older ideas about the brain basis of emotions are giving way to newer models of how the brain creates emotions, and mental experience, more generally.

Keywords: brain, emotion, affect, neuroscience, affective neuroscience, fMRI, theories of emotion

Introduction

How does the brain create emotion? Scholars have sought the biological basis of emotions such as "anger" and "fear" since as early as Hippocrates and his four humors. Today, our understanding of how the brain creates the mind have certainly expanded far beyond the notion that certain bodily substances—such as black bile—would be related to emotional dispositions such as sadness. Indeed, after decades of progress and accumulating data, the field of **affective neuroscience**—a branch of neuroscience dedicated to understanding the neurobiological basis of emotions, reward, value, and other related psychological constructs—has learned more about how the brain creates these quotidian experiences. Such knowledge has important bearing on our understanding of mental health and wellness and may even shed light on how the brain creates the mind, more generally.

Key Terms and Definitions

In this chapter, we review advancements in affective neuroscience over the past few decades to weigh in on current understanding about the brain basis of emotions. We draw primarily from findings using functional magnetic resonance imaging (fMRI), a neuroscience method that tracks changes in blood oxygenation to estimate blood flow to regions of the brain in response to stimuli and during different mental states. fMRI relies on the Blood-Oxygenated Level Dependent (BOLD) signal, which is a robust albeit indirect physiological marker of brain activation (Logothetis & Wandell, 2004). Of course, fMRI is not the only method for studying the brain basis of emotions and has crucial caveats and limitations. However, unlike methods that study the neural basis of "survival behaviors" in non-human animals (see (LeDoux & Daw, 2018) or methods that rely on naturally occurring or medically necessary lesions in humans with brain pathology (see Damasio, Adolphs, & Damasio, 2003), fMRI gives scientists the

unparalleled ability to non-invasively observe changes in neuronal activity in conscious, healthy humans (Raichle, 2001). Granted, the temporal resolution of fMRI is slower than that of EEG/MEG, but fMRI studies can also reveal both spatial and temporal relationships across the whole brain, which is useful for understanding how interactions within and between brain regions can give rise to psychological functions. fMRI studies of emotion have been available since the 1990's when they became the primary method of non-invasively studying human emotions, which means that the field of affective neuroscience has roughly 30 years' worth of data at its disposal for evaluating the brain basis of emotion. As we discuss, the knowledge generated by fMRI studies and their interpretation has changed over time as data has accumulated and as neuroimaging methods and computation have correspondingly improved.

Despite centuries' worth of debate about the nature of emotion (see Gendron & Barrett, 2009), scientists still tend to disagree about what an "emotion" is. Throughout this chapter, we use the term "emotion" to refer to discrete mental states that people experience as feelings within their own bodies, that impact their own perceptions and behaviors, and that they may label with terms such as (in English), "anger," "fear," or "happiness." Part of the confusion about what emotions are arises because the word "emotion" can refer to both an emotion *category* (e.g., "anger") and an *instance* of that category (e.g., feeling angry at a disrespectful stranger) in common parlance (Barrett & Lida, in press). A **category** is a group of objects, events, or instances that share similar features (Hoemann, Wu, et al., 2020; Murphy, 2002). The features that define a category can be both physical (e.g., chemical changes, visceromotor movement) and abstract (e.g., motivations, appraisals, value). An **instance** of a category refers to an object or event that is an exemplar of the category, but these instances are situation-bound, meaning that the features of that instance are embedded within and thus specific to a particular moment or

context (Barrett & Lida, in press). For example, the word “red” names a category that encompasses a wide variety of hues, brightnesses, and saturations, while an instance of seeing “red” could be the specific wavelengths of light you see when you look at an apple hanging from a tree on a bright, sunny day. This instance differs from the instance of “red” you see at a stoplight in the dark of night, or even the instance of red you see on that same apple on a cloudy day.

In affective neuroscience, a category such as “anger” refers to the abstraction across instances that form a prototypical instance of "anger." Of course, what is considered prototypical of a category might be culture- and person-specific. In many Western cultures, we think of anger as prototypically involving feelings of unpleasantness and high arousal, a furrowing of the brow and widening of the mouth to speak or yell, an increased heart rate and breathing rate, increased blood flow that reddens the skin of the face, clenched fists and outstretched hands, and the urge to punch or aggress against someone. Yet instances of “anger” are situation-specific and incredibly heterogeneous (e.g., anger at a stranger vs. at your child vs. at the opponent sports team all involve different facial expressions, physiology, behaviors and internal qualia).

We also differentiate between "emotion” and the term “affect”, which refers to a more global feeling that is typically characterized by two psychological dimensions: valence (feeling pleasure vs. displeasure) and arousal (feeling activated vs. still; Barrett & Bliss-Moreau, 2009; Russell, 2003). While the specific meanings of emotion categories vary widely across cultures and languages, emotion categories widely share features of affect across cultures (Jackson et al., 2019; Yik et al., 2023). Affect is also not unique to emotion, but is thought to be the brain’s abstract representation of the state of the body (Feldman et al., 2024; Shaffer et al., 2022) and contributes to perception, attitudes, decisions, and other so-called “cognitive” states. Some also

argue that affect is a core feature of consciousness (Barrett & Bar, 2009; Barrett & Bliss-Moreau, 2009; Damasio, 1999; Russell, 2003). Thus, an understanding of how the brain represents affect may shine light on how the brain contributes to an understanding of consciousness, more generally.

The Brain Basis of Emotion: Historical Influences and Emerging Consensus

Empiricism is always implicitly or explicitly shaped by theory, and the collection and interpretation of data produced by neuroimaging studies of emotion has been influenced across history by the interplay of psychological theories that differ fundamentally in their philosophical approaches. On the one hand are theories that use the hypothetico-deductive method to try to map brain structures and functions to the psychological categories that are named by human experiencers (e.g., "emotions," "fear," "anger"). That is, these theories take the psychological categories used commonly in daily discourse as scientific categories to be explained by brain data. We call these **typological views of emotion** (see Barrett & Theriault, in press)**;** examples of it in the literature are "basic emotion theory" (e.g., Ekman, 1992; Ekman & Cordaro, 2011; Levenson, 1999), "discrete emotion theory" (e.g., Izard, 1977; Tomkins, 1962, 1963), and "primary emotional affects" (Panksepp, 2004, 2016) (see Shiota, 2024 for discussion of these typological views).

On the other hand, are theories that try to inductively discern from the data how human psychological experiences come to be; these theories take psychological categories that are named by human experiencers to be folk categories, but not scientific categories that "carve nature at its joints" (cf., Plato, *Phaedrus 265e*) or categories which can be directly observable in the structure of biology. In the case of affective neuroscience, for instance, these approaches examine how the brain's structural and functional principles constrain how it can create the

experiences that some (but not all) humans refer to as “anger,” “fear,” “sadness,” etc. Such theories assume that categories such as “emotions,” “fear,” “anger” and so on are human constructions. Here, we refer to these as **constructionist views of emotion** (see Gendron & Barrett, 2009 for discussions of types of constructionist theories and more background on different philosophical assumptions between basic emotion and constructionist approaches). Examples of such views include “the theory of constructed emotion” (Barrett, 2017; Lindquist et al., 2022), the “Ortony, Clore and Collins (OCC)” model (Clore & Ortony, 2013), “the psychological construction of emotion” model (Russell, 2003), and the “entangled brain perspective” (Pessoa, 2023). With these methodological and theoretical forces in mind, we traverse the history of affective neuroscience and what is currently known about how the brain may create human emotions (see Figure 1 for a schematic).

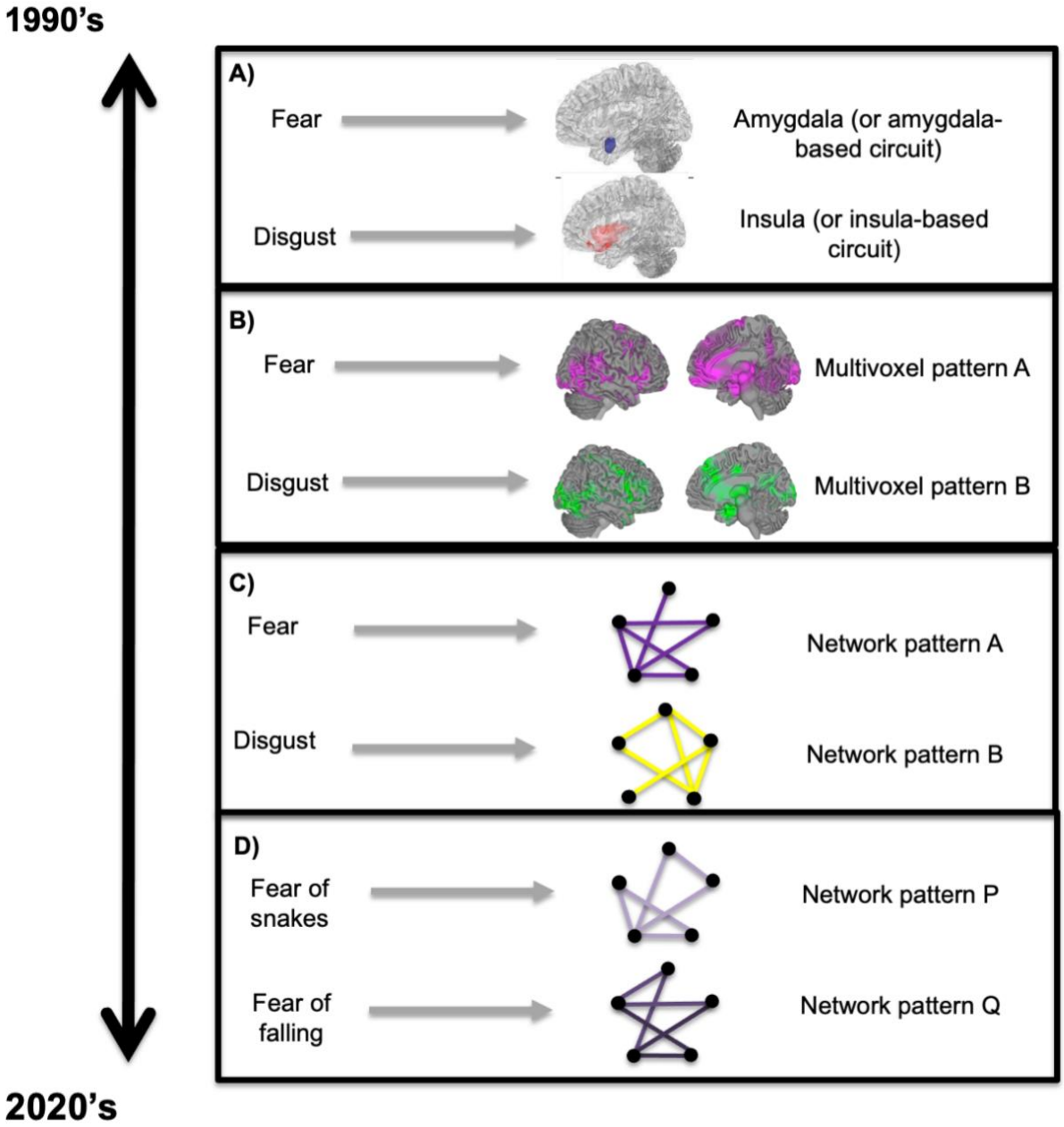

Figure 1. A schematic of shifts in understanding of the brain basis of emotions from the late 20th century to the beginning of the 21st century. Shifts from the localization of function to specific brain regions or circuits (A) to search for multivoxel patterns (B) to network-based understandings of emotion (C-D) to a situated understndings of those network patterns (D) were each the confluence of both theory and methodological advancements. In the final sections, we discuss the current research that is forming what comes after section D. Figure adapted from Lindquist & Barrett (2012).

Early Evidence for the Typology View and the Localization of Function

Early fMRI investigations into how sensations and perceptions emerged from the brain were conducted under the assumption of typological models of brain function. Due to notions from lesion studies that observed psychological deficits following relatively circumscribed brain damage, these earliest neuroimaging studies made assumptions of functional modularity: that a discrete brain region (e.g., fusiform face area) performs a single mental function (e.g., processing faces) (Bergeron, 2007; Fuster, 2000). The earliest affective neuroscience studies were no exception. Throughout the 20th century, emerging evidence from lesion studies (e.g., Adolphs et al., 1994; Hitchcock & Davis, 1986; K. LaBar et al., 1995; Raleigh et al., 1979) and electrical stimulation studies (e.g., Halgren et al., 1978; Mos et al., 1982; Panksepp, 1986; Sem-Jacobson, 1968) in non-human animals and humans alike were interpreted as evidence that specific emotion categories could be localized to specific anatomical brain structures, largely within the confines of the so-called "limbic system" (see Panksepp, 1998 for a review).[i] These lesion- and stimulation-based findings contributed support to typological theories that hypothesized that emotions such as "fear" and "anger" were each supported by specific anatomical structures. Based on these findings, affective scholars hypothesized that an instance of emotion such as "fear" emerged when subcortical regions of the brain were triggered by perceptions of an external stimulus, coordinating discrete response patterns in physiology, facial musculature, behavior, and subjective experience that are associated with a certain emotion category (Panksepp et al., 1998). In some of the models of emotion emerging during this time, emotion categories were localized to specific brain regions (e.g., Calder, 2003), while other models emphasized emotion categories as being rooted in anatomically-wired neural circuits that included multiple structurally connected brain areas (e.g., Adolphs, 2002; Izard, 1993; Panksepp

et al., 1998). Overall, these models largely situated emotions within the brainstem, subcortical structures of the brain (i.e., the amygdala, hypothalamus, insula, ventral striatum, periaqueductal gray) as well as a select few cortical areas like the orbitofrontal cortex (Dalgleish, 2004; Panksepp et al., 1998).

The hypothetico-deductive method did not alone influence interpretations of functional modularity in affective neuroscience. These interpretations were furthermore constrained by the computational methods used in early neuroimaging studies (e.g., Bergeron, 2007; Dunn & Kirsner, 2003), which tended to reveal activation in relatively circumscribed brain structures and thus reinforced typological theories and assumptions of functional modularity. This led to the interpretation that clusters of brain activity that were relatively localized to certain anatomical brain regions were *the* brain center for that emotion category (e.g., the amygdala and fear, Davis, 1992; Whalen, 1998; the insula and disgust, Phillips et al., 1997; Wicker et al., 2003). In other cases, anatomically defined regions of interest (e.g., the amygdala) were specifically targeted as regions of interest, due to the hypothesized links to specific emotions (e.g., LaBar et al., 1998; Morris et al., 1996). Collectively, these methods contributed to the typological theory hypothesis that certain emotion categories were consistently and specifically associated with certain relatively circumscribed brain anatomy (for a discussion see Barrett et al., 2007; Lindquist & Barrett, 2012).

With both the accumulation of increasing numbers of fMRI studies of emotion and advances in computational power, meta-analytic summaries could empirically test the typological hypothesis that an emotion category such as “fear” is consistently and specifically associated with activation within the amygdala. Very early meta-analytic summaries of the neuroimaging literature on emotion concluded that the body of evidence supported a consistent

and specific link between activation within certain brain structures and certain emotion categories (e.g., Phan et al., 2004). Yet other subsequent meta-analytic reviews containing more data failed to find evidence for consistent and specific associations between functional activation within specific neuroanatomically-defined structures and the experience of specific emotion categories (Kober et al., 2008; Lindquist & Barrett, 2012). For a brain region to be *the* neural circuit for a specific emotion category, it would have to evince both consistent and specific activation during all instances of that emotion category (see Lindquist et al., 2012 for a discussion). Instead, this work demonstrates that individual brain regions are associated with multiple emotion categories, and that one emotion category engages more than one brain region across studies (see Kober et al. 2008; Hamann, 2012; Lindquist et al. 2012).

As one illustrative example, when summarizing the entire neuroimaging literature on emotion published between 1992-2007, Lindquist and colleagues (2012) found that the amygdala was not consistently and specifically associated with experiences of fear across the literature. The right amygdala had activation greater than expected by chance in only about 30% of experimental contrasts assessing the experience of fear whereas the left amygdala showed greater activation than chance in 0% of those contrasts. Instead, the right amygdala showed the most consistent activation for experiences of disgust, reflecting a roughly 52% activation rate across studies of disgusting experiences and the left showed the most consistent activation for experiences of anger, reflecting a roughly 33% activation rate across studies of angering experiences. These sorts of findings began to draw into question typology accounts that had hypothesized and found a 1:1 association between emotion categories and specific neuroanatomy for many years.

These inductive analyses were complemented by analyses of other sources of data from other modalities (e.g., behavior, facial expressions, peripheral physiology; Barrett, 2006) which failed to find evidence for discrete patterns associated with English language emotion categories. The brain-based meta-analyses also coincided with methodological advances in computation that moved neuroimaging from a focus on univariate analyses to multivariate analyses assessing patterns of brain activation between regions spread throughout the brain. Thus, the initial typology models of the 20th century began to give way to models which assumed that the neural representation of emotion was more distributed; few affective neuroscientists in the early 21st century now subscribe to the strong localization of function that predominated 20th century affective neuroscience, even if this work was foundational in early affective neuroimaging.

From Localization of Function to Distributed Representations

Evidence disconfirming the strong localization of function in emotion also could not have emerged without simultaneous methodological advances in **network neuroscience**, a field in which scientists examine how multivariate relations between various levels of neurobiological activity (e.g., neuronal firing, BOLD timeseries within brain regions) give rise to large-scale brain systems (see Barrett & Satpute, 2013; Bassett & Sporns, 2017). In neuroimaging, application of network methods has revealed evidence for sets of "intrinsic networks", large-scale communities of brain regions that show consistent low-frequency functional correlations when participants are "at rest" or not engaging in an experimental task. These findings, when paired with knowledge of structural connections between brain regions gleaned from non-human animal models and human white matter tract mappings, led to the conclusion that regions with correlated BOLD timeseries act as functionally wired units (e.g., Biswal et al., 2010; Bullmore & Sporns, 2009; Fox & Raichle, 2007; Sporns et al., 2004).

Findings spanning the human neurosciences revealed that these intrinsic networks were also activated during a variety of experimental tasks, including tasks involving autobiographical memory (e.g., (Buckner et al., 2008; Spreng et al., 2009), cognitive control (e.g., Cole et al., 2013; Zanto & Gazzaley, 2013), and visual attention (e.g., Corbetta & Shulman, 2002; Shulman et al., 2010), suggesting that they serve domain-general functions (Cole et al., 2014). This work challenged the idea that brain regions independently performed mental functions and instead demonstrates that mental processes depend on the functional interactions within and between distributed brain networks (Bressler & Menon, 2010; Sporns, 2014). Early constructionist theories deduced from the meta-analytic summaries of the emotion neuroimaging literature that these domain-general networks may similarly support emotions (Barrett & Satpute, 2013; Kober et al., 2008; Lindquist & Barrett, 2012; Pessoa, 2008).

Indeed, evidence establishing the brain's domain-general functional architecture ushered in a paradigm shift in the understanding of the brain basis of emotion. For one, the domain-generality of these intrinsic networks begins to finally puts to rest notions that emotion categories map onto one specific brain region or neural network. Networks that are seemingly linked to emotions are also linked to "cognitive" brain functions and vice versa. For instance, the so-called "salience network" has been associated with **interoception** and visceromotor regulation and representation of bodily sensations (Seeley, 2019; Uddin, 2016). It is thus not surprisingly associated with emotions (Barrett, 2017; Clore et al., 2021; Feldman et al., 2024; MacCormack & Lindquist, 2017; Satpute & Lindquist, 2019). Yet, its involvement is also observed across myriad "cognitive" tasks that have affective implications, such as those that involve the mobilization of internal resources for tracking goal-relevant stimuli (Lamichhane et al., 2016; Menon & Uddin, 2010). Indeed, its nodes within the anterior mid-cingulate cortex (associated

with somatovisceral engagement and motor movement), subgenual anterior cingulate cortex (associated with somatovisceral engagement) and anterior insula (associated with somatovisceral representations) make it well-suited for a role in generating and representing internal visceral states in the service of behavioral demands (see Lindquist & Barrett, 2012; Seeley et al., 2007).

In contrast, the default mode network is best known for its role in autobiographical memories, semantic representation, social perception and representation, prediction of the future, and context-specific visual perception (see Buckner et al., 2008; Spreng et al., 2009). It has been less traditionally associated with emotion per se, although data and theories from the early 21st century clearly demonstrate its role in emotion (see Amft et al., 2015; Satpute & Lindquist, 2019). Indeed, the default mode network's assortment of anterior and medial brain tissue has structural features that make it well-suited to represent abstract, heteromodal summaries of prior experiences (Satpute & Lindquist, 2019). In particular, the medial prefrontal cortex and anterior temporal pole possess fewer neurons that are less densely packed when compared to structures elsewhere in the brain (i.e., in many unimodal sensory cortices, e.g., primary visual cortex; (Finlay & Uchiyama, 2015). As a result, the default mode network is well-suited to represent low-dimensional summaries of information from the sensory modalities, a function that may be especially important to emotion. We return to this point later.

Finally, research points to the role of the frontoparietal control network and dorsal attention network in emotion. These networks are associated with cognitive control and visual attention, respectively (see Corbetta & Shulman, 2002; Dixon et al., 2018). Although executive control has long been associated with emotion regulation in the affective neuroscience literature (e.g., Ochsner et al., 2012), it has not traditionally been associated with emotion experience (also known as "emotion generation"), although those views are shifting in light of evidence for its

involvement in the generation of emotional experiences. Indeed, evidence suggests that the frontoparietal control network is relatively more engaged during the experience of emotion than during emotion regulation (J.-X. Zhang et al., 2023), a fact that defies the notion that its main purpose is to "regulate" so-called "irrational" emotional behaviors.

A separate, but distinct, multivariate analytical method called multivariate pattern analysis has also begun to shine light onto the distributed neural representation of emotions. Unlike network-based approaches that examine patterns of correlated BOLD activity across regions spanning the brain, multivariate pattern analyses tend to examine how BOLD signal dynamics within voxels of the brain within a region or spanning regions encode types of stimuli. This approach uses the pattern of increased or decreased activation within voxels that form a brain region to compute an overall multivariate summary associated with a stimulus class (e.g., "fear") (Haxby, 2012; Norman et al., 2006).

In one type of multivariate brain modelling, scholars use supervised machine learning methods to reveal whether the pattern of voxel activations and deactivations associated with one stimulus class is clear enough across trials to accurately predict the stimulus class associated with held out (un-analyzed) brain activity. In contrast to standard univariate fMRI analyses, which focus on identifying which brain regions on average show increased significant activity to an emotion category relative to some baseline, pattern classification takes a set of labelled multivarate patterns and learns to differentiate patterns associated with one emotion category from patterns associated with other emotion categories (Azari et al., 2020). When applied to emotion categories, pattern classification studies can "predict" with relatively high accuracy whether brain activity patterns spanning voxels belong to one category of emotion (e.g., "fear")

versus multiple others (e.g., anger, disgust, love, etc.) (e.g., Hamann, 2012; Kassam et al., 2013; Kragel & LaBar, 2014; Nummenmaa & Saarimäki, 2019).

Like the network-based findings, when applied to affective neuroscience, the multivoxel analyses reveal important new discoveries. Although some typology studies taking the hypothetico-deductive approach have interpreted evidence for multivoxel pattern findings as evidence for the typology hypothesis that emotion categories are biologically evolved and innate (e.g., Adolphs & Anderson, 2018; Saarimäki et al., 2022). One logical problem with this conclusion is that pattern classification studies can reveal evidence for the neural representation of categories that are clearly cultural artifacts (e.g., bicycles v. cars); this finding challenges the notion that finding a brain pattern for a category reveals anything about a biologically-evolved representation of that category.

Others have criticized multivoxel pattern analyses of emotion categories on other grounds because multivoxel patterns reflect statistical summaries of patterns of brain activation across instances of a category and across individuals; see Azari et al., 2020; Clark-Polner et al., 2017; Kragel et al., 2018 for further discussion). For instance, a recent paper showed that distinct multivoxel patterns exist for different situated instances of fear (fear of heights, fear of spiders, and fear of social threats) and that only about 2% of voxels explain situation-general experiences of fear (Wang et al. 2024). Rather, these studies collectively tell us how the brain uses tissue that is distributed across the cortex and subcortex to represent certain types of content, including, but not limited to, emotions (Satpute & Lindquist, 2019).

From Distributed Representations to the Interplay of Networks

Collectively, the multivariate findings—whether using a network-based approach or a multivariate pattern approach—have moved affective neuroscience further from its beginnings based in functional modularity. Instead, emerging evidence characterizes the pattern associated with instances of emotion as a dynamic array of brain regions within a specific whole-brain context (as in Ciric et al., 2017). Collectively, these network-based findings are quite inconsistent with the typology view of emotion; rather, they suggest that instances of emotion, even instances of the same emotion category, emerge through dynamic functional interactions between neural networks (see Lindquist & Barrett, 2012).

Indeed, there is now growing evidence that instances of an emotion category are best represented as a functional assembly of within- and between-network patterns of functional connectivity in the brain. For instance, meta-analytic evidence finds on average greater functional co-activation amongst regions within the dorsal attention network and default mode network during anger when compared to a similarly valenced emotion such as fear (Wager et al., 2015). In contrast, fear, when compared to anger, is on average characterized by greater functional co-activation within both a basal ganglia network and a sensorimotor network (Wager et al., 2015). Other individual studies link self-reported state anxiety with increased connectivity within the salience and default mode networks (Saviola et al., 2020) and state anxiety associated with social anxiety disorder to connectivity between nuclei of the basal ganglia and regions of the frontoparietal control and salience networks (Anteraper et al., 2014).

In a recent study, Doyle and colleagues (2021) specifically revealed patterns of brain connectivity that characterized both between-emotion category differences and differences between instances of the same emotion category. Using a data-driven subgrouping procedure, they found that feelings of anger and anxiety that were induced via an autobiographical scenario

immersion technique were associated with different connectivity patterns among and within subnetworks of the salience, default mode, frontal parietal network, and dorsal attention network. Critically, such differences were not only revealed *between* the emotion categories (averaging over instances of that category), but also within each emotion category (differentiating amongst instances within that category). Specifically, within the anxiety induction, there were different subgroups of within and between network patterns; these patterns were identified in a data-driven fashion based exclusively on their network-based patterns of connectivity, and were not attributable to differences in the intensity or quality of emotions felt by participants nor to stable differences between individuals across emotions. As a case in point, whereas anxiety was in general associated with connectivity within aspects of the default mode network and between basal ganglia, somatomotor, salience, and right frontoparietal control network, two different subgroups were identifiable within this category-level average pattern. Specifically, subgroups differed in the extent to which they had connectivity amongst sub-networks comprising the default mode network. These types of network-based findings, in which researchers are examining complex within- and between-network patterns of connectivity during different emotions, increasingly represents the current state of the science in the early 21st century.

Clearly, we have traversed much terrain since the early days of fMRI studies in affective neuroscience. These findings collectively suggest that while the brain encodes emotions that humans name in common parlance (e.g., “anger,” “fear,” “joy”), its structure hardly seems to respect those categories. That is, emotion categories seem to emerge from the complex interplay between brain networks that themselves are linked to very general functions. What remains in question for scientists now is how these within- and between-network functional dynamics are

governed. Very new research based in predictive processing models of brain function are suggestive, although this work is still in its infancy.

Predictive Processing Models of Brain Function

To understand why and how distributed neural networks interact to create emotions, we first take a step back to examine what the field of neuroscience has revealed about the principles of brain function, more generally, in the past few decades. Here, the affective neuroscience work is squarely influenced by the constructionist theoretical perspective, which seeks to understand how the basic principles of brain function might constrain the brain's creation of human experiences of "anger," "fear," "joy," etc. We briefly introduce readers to this general approach and close by stating its implications for affective neuroscience.

Predictive Processing: A Precis

There is a growing consensus in cellular and computational neuroscience that the brain engages in **predictive processing**, a general signal processing strategy in which data is compressed to remove redundant information and conserve energy (Bastos et al., 2012; Clark, 2013; Friston, 2010; Sterling & Laughlin, 2015). Evidence from neuroscience, physiology, and ecology suggest that brains likely acquired predictive processing over the course of evolutionary history as a way to manage the metabolic costs of sensing and adapting to complex demands on processing (Bullmore & Sporns, 2012; Sterling & Laughlin, 2015). Specifically, as single cell and then multi-cellular organisms evolved increasingly complex and more metabolically demanding physiological systems (e.g., capacity for movement), selection pressures favored development of a central nervous system (i.e., brain) to model and regulate the energetic needs of the body (Shaffer et al., 2022; Sterling & Laughlin, 2015). A regulator of a complex system efficiently achieves regulation by generating an internal representational model of the system,

and updating the model when it is incorrect (i.e., learning; Bechtel & Bich, 2021; Clark, 2013; Sterling & Laughlin, 2015).

Predictive processing frameworks of emotion thus propose that the brain stores information based on learned patterns of sensory and motor input to generate predictions about future incoming sensory data and adaptive motor actions (Barrett, 2017; Theriault et al., 2020). The brain identifies patterns within internal visceromotor (i.e., termed "interoceptive") and external sensory input (vision, audition, olfaction, etc.), then recreates the patterns' structure and function to draw upon in the future, resulting in an internal "generative" model of the world. The brain then uses that model to predictively regulate the body in response to changing situations in the world around it (Barrett, 2017). For instance, a brain might learn that in a back alley, a sudden sound (e.g., snarling) and sight (e.g., the mouth of a dog) should engender quick mobilization of the autonomic nervous system, skeletomotor system, and retreat. However, in a different context (while playing with a dog in the backyard), no such action is necessary.

To achieve these functions, the brain thus needs structures that process incoming sensory signals from the body and world, and structures that generate, store, and apply an abstract model of the world based on prior experiences of those sensory signals. This is achieved via iterative connections between unimodal sensory regions - regions evolved to represent internal and external sensory signals in relatively high dimensions - and heteromodal association cortices such as the cortex contained in the default mode network. Evidence from comparative developmental neuroanatomy suggests that these brain regions are uniquely suited to represent low-dimensional summaries of unimodal sensory information given the structure, function, and development of these brain regions, both in utero and across primate evolution (for reviews see Finlay & Uchiyama, 2015; Katsumi et al., 2022; Satpute & Lindquist, 2019; Shaffer et al., 2022).

Indeed, regions such as medial prefrontal cortex, precuneus, and lateral temporal cortex that make up the default mode network have sparsely packed neurons and are late to develop both during neural tube development in utero and across mammalian brain evolution, suggesting that they may be uniquely suited for representing the sort of abstract, low-dimensional summary representations that have been associated with the default mode network throughout affective neuroscience studies of the 20th and 21st century.

Relevant to the connectivity amongst networks of the brain, these heteromodal association areas sit atop the primate brain's functional heterarchy.[ii] Neuroanatomical work mapping out the flow of information within primate cortex shows that the brain's internal model or "predictions" about the meaning of upcoming sensory and motor events are represented in heteromodal brain regions that can represent highly compressed, low-dimensional information that is abstracted from the modalities (Barbas, 2015; Barrett, 2017; Barrett & Simmons, 2015; Feldman et al., 2024). Within this heterarchy, efferent predictions are generated based on incoming sensory information from the body and exteroceptive modalities and projected to sensory regions to shape the next wave of incoming sense data. If afferent information does not confirm the prediction, the unanticipated input generates prediction error that is propagated back up the neuroaxis to be encoded into the brain's internal model (Barrett & Simmons, 2015). However, afferent sensory information that confirms predictions is effectively ignored, with the implication that a well-predicting internal model is an especially efficient one (Friston et al., 2015; Shaffer et al., 2022).

Implications for Affective Neuroscience

Although predictive processing models are typically tested at the cellular level in non-human animals (Barbas, 2015), advancements in computation and connectomics has led to a

surge of interest in studying how psychological functions emerge from functional **gradients** at various levels of neurobiological activity, including cytoarchitecture, gene expression, intrinsic functional and structural connectivity, and patterns of task-base activation (Bernhardt et al., 2022; Haak et al., 2018; Huntenburg et al., 2018; Oligschläger et al., 2019; Shafiei et al., 2020). These data reveal multiple domain-general gradients in the brain that describe low-dimensional and continuous representations of functional variation spanning the whole brain. It is these gradients that appear to have important implications for our understanding of neural network functions underlying emotion, and indeed, probably all mental states.

The general gradients of function observed in recent human neuroimaging studies, and replicating knowledge derived from comparative approaches, have important implications for modern network neuroscience models of emotion. Principle among these gradients is the "sensorimotor-association" gradient (also sometimes called "unimodal-transmodal" gradient; Margulies et al., 2016; C. Murphy et al., 2019; Vázquez-Rodríguez et al., 2019) . On the association end, brain regions that comprise the default mode network such as medial prefrontal, cingulate, precuneus, and middle temporal areas integrate multi-modal information to make predictions. Regions that comprise the frontoparietal network such as the lateral prefrontal and inferior parietal cortex fine-tune predictions by suppressing predictions when priors are very low. On the sensorimotor end, brain regions in primary sensory cortices process incoming sensations and send prediction errors back to association cortices (Katsumi et al., 2022; J. Zhang et al., 2019).These data suggest that during emotion, flow of information through networks should move from default mode sub-regions to unimodal sensory regions such as visual cortex, auditory cortex, interoceptive cortex, and motor cortex.

Using a constructionist inductive approach to understand these dimensions suggests that brain regions involved in representing memory and concepts make meaning out of sensory information in a given context, shaping what is seen, heard, felt, and acted upon (Barrett, 2017; Satpute & Lindquist, 2019). Moreover, this process can be influenced by attention and goals relevant to higher-order contextual demands. These neural findings offer profound new understandings of how prior experiences, learning, cultural knowledge, and the power of the situational demands could be infusing the brain's construction of emotional experiences, even before they are triggered by external sensory events.

The Brain Basis of Emotion: Looking Towards the Future

In the past two decades, research on the brain basis of emotion has moved beyond the assumption that discrete emotion categories like "fear" or "disgust" reside in subcortical structures alone. As outlined in this chapter, advances in technology, including technology that allowed researchers to study humans experiencing emotions in real time, such as functional magnetic resonance imaging (fMRI) and advancements in data computation, revealed that emotions emerge from dynamic, context-dependent interactions across distributed neural networks. These data suggest that neural representations of emotion engage regions involved in generating predictions of the future based on the past, regions involved in visceromotor engagement, and regions that help select amongst those predictions. This work has forwarded new hypotheses on the neural basis of emotions, leading to new distributed neural network models of emotion in the first decades of the 21st century, and even more recently, to new predictive processing models of emotion.

Predictive processing models of emotion, like the theory of constructed emotion (Barrett, 2017) and constructed mind approach (Shaffer et al., 2022), present exciting new methodological

challenges for the future of affective neuroscience. First, emerging constructionist models of the brain argue that all psychological phenomena, including emotion, are supported by large-scale functional interactions in the brain (Barrett & Satpute, 2019; Hutchinson & Barrett, 2019). In this chapter, we focus on functional interactions along a sensorimotor-to-association gradient, where brain regions perform various functions along a continuum from lower-level sensory processing to higher-level sensory integration and abstraction (e.g., Huntenburg et al., 2018). Tract-tracing and diffusion mapping studies confirm that intrinsic activity across measurement occasions, modalities, and subjects can be mapped onto this common sensorimotor-association gradient (e.g., (Haak et al., 2018; Katsumi et al., 2022). If affective neuroscience is to keep pace with advancements in cortical mapping, an important next step for future research will be understanding the spatiotemporal dynamics of this gradient during instances of emotion. To date, most connectivity gradient analyses examine functional processing at rest; few studies investigate whether the dimensions of the sensorimotor-association gradient are stable across tasks and no work, to our knowledge, examines whether it best represents brain activation during emotion. More targeted analysis of the sensorimotor-association gradient, leveraging multivariate approaches across different emotion task paradigms, will be required to test whether this gradient reflect neural activity during instances of emotion.

In addition to the emerging consensus that psychological phenomena are supported by large-scale functional interactions that span the cerebral cortex, predictive processing models suggest that neural networks should not just be modeled in terms of static activity, but as "shifting populations of neurons" that are informed by previous brain states (p. 6, Barrett and Satpute, 2019). The brain generates expectations and predictions about incoming sensory information based on prior knowledge and experience, suggesting that past neural activity may

play a crucial role in shaping ongoing neural processes (Barrett, 2017; Friston, 2010; Lee et al., 2021). As Lee and colleagues outline in their 2021 review, fMRI tasks assessing emotion have traditionally tended to focus on measuring brain responses to randomized stimuli in isolation, but they may not actively engage processes where the brain learns from these stimuli and adjusts its predictions accordingly (Lee et al., 2021). Moreover, emerging models of the intrinsic functional architecture of the brain suggest that connectivity within and between networks shifts over short periods of time (e.g., (Allen et al., 2014; Ciric et al., 2017). An important direction for future work will thus be to better understand the spatial and temporal variability contributing to the neural representations of emotion.

Concluding Remarks

Nearly two hundred years after the first works in affective neuroscience, the field finds itself at the crossroads of what might be a paradigm shift—where older ideas about the biological basis of emotions are giving way to newer models of how the brain creates emotions, and mental experience, more generally. While this paradigm shift toward predictive and context-sensitive neurobiological models of emotion requires new computational tools and experimental paradigms, the theory of constructed emotion provides a useful conceptual framework for answering the age-old question of how the brain creates emotion. We will no doubt continue to learn more as methods advance, but a cautionary tale from this literature is that methods and scientific philosophy together shape the generation of knowledge and its interpretation.

Further Reading:

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), Article 1. https://doi.org/10.1093/scan/nsw154

Barrett, L. F. & Theriault, J. (in press). What's real? A philosophy of science for social psychology. Chapter to appear in D. Gilbert, S. T. Fiske, E. Finkel, & W. B. Mendes (Eds.), *The Handbook of Social Psychology*, 6th Edition.

Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, *20*(3), 353–364. https://doi.org/10.1038/nn.4502

Bressler, S. L., & Menon, V. (2010). Large-scale brain networks in cognition: Emerging methods and principles. *Trends in Cognitive Sciences*, *14*(6), Article 6. https://doi.org/10.1016/j.tics.2010.04.004

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Clark-Polner, E., Johnson, T. D., & Barrett, L. F. (2017). Multivoxel pattern analysis does not provide evidence to support the existence of basic emotions. *Cerebral Cortex*, *27*(3), Article 3. https://doi.org/10.1093/cercor/bhw028

Cole, M. W., Bassett, D. S., Power, J. D., Braver, T. S., & Petersen, S. E. (2014). Intrinsic and task-evoked network architectures of the human brain. *Neuron*, *83*(1), 238–251. https://doi.org/10.1016/j.neuron.2014.05.014

Finlay, B. L., & Uchiyama, R. (2015). Developmental mechanisms channeling cortical evolution. *Trends in Neurosciences*, *38*(2), 69–76. https://doi.org/10.1016/j.tins.2014.11.004

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience*, *6*(4), Article 4. https://doi.org/10.1080/17588928.2015.1020053

Hutchinson, J. B., & Barrett, L. F. (2019). The power of predictions: An emerging paradigm for psychological research. *Current Directions in Psychological Science*, *28*(3), 280–291. https://doi.org/10.1177/0963721419831992

Katsumi, Y., Kamona, N., Zhang, J., Bunce, J. G., Hutchinson, J. B., Yarossi, M., Tunik, E., Quigley, K. S., Dickerson, B. C., & Barrett, L. F. (2021). *Functional connectivity gradients as a common neural architecture for predictive processing in the human brain* [Preprint]. Neuroscience. https://doi.org/10.1101/2021.09.01.456844

Kragel, P. A., Koban, L., Barrett, L. F., & Wager, T. D. (2018). Representation, pattern information, and brain signatures: From neurons to neuroimaging. *Neuron*, *99*(2), 257–273. https://doi.org/10.1016/j.neuron.2018.06.009

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,* 121-143. https://doi.org/10.1017/S0140525X11000446.

Satpute, A. B., & Lindquist, K. A. (2019). The default mode network's role in discrete emotion. *Trends in Cognitive Sciences*, *23*(10), Article 10. https://doi.org/10.1016/j.tics.2019.07.003

Wager, T. D., Kang, J., Johnson, T. D., Nichols, T. E., Satpute, A. B., & Barrett, L. F. (2015). A Bayesian model of category-specific emotional brain responses. *PLoS Computational Biology*, *11*(4), 1004066. https://doi.org/10.1371/journal.pcbi.1004066

References

Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, *12*(2), 169–177. https://doi.org/10.1016/S0959-4388(02)00301-X

Adolphs, R., & Anderson, D. J. (2018). *The Neuroscience of Emotion: A New Synthesis*. Princeton University Press. https://doi.org/10.23943/9781400889914

Adolphs, R., Tranel, D., & Damasio, H. (1994). *Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala*. *372*.

Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., & Calhoun, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cerebral Cortex (New York, NY)*, *24*(3), 663–676. https://doi.org/10.1093/cercor/bhs352

Amft, M., Bzdok, D., Laird, A. R., Fox, P. T., Schilbach, L., & Eickhoff, S. B. (2015). Definition and characterization of an extended social-affective default network. *Brain Structure & Function*, *220*(2), 1031–1049. https://doi.org/10.1007/s00429-013-0698-0

Anteraper, S. A., Triantafyllou, C., Sawyer, A. T., Hofmann, S. G., Gabrieli, J. D., & Whitfield-Gabrieli, S. (2014). Hyper-connectivity of subcortical resting-state networks in social anxiety disorder. *Brain Connectivity*, *4*(2), 81–90. https://doi.org/10.1089/brain.2013.0180

Azari, B., Westlin, C., Satpute, A. B., Hutchinson, J. B., Kragel, P. A., Hoemann, K., Khan, Z., Wormwood, J. B., Quigley, K. S., Erdogmus, D., Dy, J., Brooks, D. H., & Barrett, L. F. (2020). Comparing supervised and unsupervised approaches to emotion categorization in the human brain, body, and subjective experience. *Scientific Reports*, *10*(1), Article 1. https://doi.org/10.1038/s41598-020-77117-8

Barbas, H. (2015). General cortical and special prefrontal connections: Principles from structure to function. *Annual Review of Neuroscience*, *38*(1), 269–289. https://doi.org/10.1146/annurev-neuro-071714-033936

Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, *10*(1), Article 1. https://doi.org/10.1207/s15327957pspr1001_2

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), Article 1. https://doi.org/10.1093/scan/nsw154

Barrett, L. F., & Bar, M. (2009). *See it with feeling: Affective predictions during object perception*. 10.

Barrett, L. F., & Bliss-Moreau, E. (2009). Chapter 4 Affect as a psychological primitive. In *Advances in Experimental Social Psychology* (Vol. 41, pp. 167–218). Elsevier. https://doi.org/10.1016/S0065-2601(08)00404-8

Barrett, L. F., Lindquist, K. A., Bliss-Moreau, E., Duncan, S., Gendron, M., Mize, J., & Brennan, L. (2007). Of Mice and Men: natural kinds of emotions in the mammalian brain? A response to Panksepp and Izard. *Perspectives on Psychological Science*, *2*(3), 297–312. https://doi.org/10.1111/j.1745-6916.2007.00046.x

Barrett, L. F., & Satpute, A. B. (2013). Large-scale brain networks in affective and social neuroscience: Towards an integrative functional architecture of the brain. *Current Opinion in Neurobiology*, *23*(3), Article 3. https://doi.org/10.1016/j.conb.2012.12.012

Barrett, L. F., & Satpute, A. B. (2019). Historical pitfalls and new directions in the neuroscience of emotion. *Neuroscience Letters*, *693*, 9–18. https://doi.org/10.1016/j.neulet.2017.07.045

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, *16*(7), Article 7. https://doi.org/10.1038/nrn3950

Barrett, L. F. & Theriault, J. (in press). What's real? A philosophy of science for social psychology. Chapter to appear in D. Gilbert, S. T. Fiske, E. Finkel, & W. B. Mendes (Eds.), *The Handbook of Social Psychology*, 6th Edition.

Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, *20*(3), 353–364. https://doi.org/10.1038/nn.4502

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*(4), 695–711. https://doi.org/10.1016/j.neuron.2012.10.038

Bechtel, W., & Bich, L. (2021). Grounding cognition: Heterarchical control mechanisms in biology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1820), 20190751. https://doi.org/10.1098/rstb.2019.0751

Bergeron, V. (2007). Anatomical and functional modularity in cognitive science: Shifting the focus. *Philosophical Psychology*, *20*(2), 175–195. https://doi.org/10.1080/09515080701197155

Bernhardt, B. C., Smallwood, J., Keilholz, S., & Margulies, D. S. (2022). Gradients in brain organization. *NeuroImage*, *251*, 118987. https://doi.org/10.1016/j.neuroimage.2022.118987

Biswal, B. B., Mennes, M., Zuo, X.-N., Gohel, S., Kelly, C., Smith, S. M., Beckmann, C. F., Adelstein, J. S., Buckner, R. L., Colcombe, S., Dogonowski, A.-M., Ernst, M., Fair, D.,

Hampson, M., Hoptman, M. J., Hyde, J. S., Kiviniemi, V. J., Kötter, R., Li, S.-J., … Milham, M. P. (2010). Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences*, *107*(10), 4734–4739. https://doi.org/10.1073/pnas.0911855107

Bressler, S. L., & Menon, V. (2010). Large-scale brain networks in cognition: Emerging methods and principles. *Trends in Cognitive Sciences*, *14*(6), Article 6. https://doi.org/10.1016/j.tics.2010.04.004

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network. *Annals of the New York Academy of Sciences*, *1124*(1), 1–38. https://doi.org/10.1196/annals.1440.011

Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*(3), 186–198. https://doi.org/10.1038/nrn2575

Bullmore, E., & Sporns, O. (2012). The economy of brain network organization. *Nature Reviews Neuroscience*, *13*(5), 336–349. https://doi.org/10.1038/nrn3214

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

Cesario, J., Johnson, D. J., & Eisthen, H. L. (2020). Your brain Is not an onion with a tiny reptile inside. *Current Directions in Psychological Science*. https://doi.org/10.1177/0963721420917687

Chanes, L., & Barrett, L. F. (2016). Redefining the role of limbic areas in cortical processing. *Trends in Cognitive Sciences*, *20*(2), Article 2. https://doi.org/10.1016/j.tics.2015.11.005

Ciric, R., Nomi, J. S., Uddin, L. Q., & Satpute, A. B. (2017). Contextual connectivity: A framework for understanding the intrinsic dynamic architecture of large-scale functional brain networks. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-06866-w

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Clark-Polner, E., Johnson, T. D., & Barrett, L. F. (2017). Multivoxel pattern analysis does not provide evidence to support the existence of basic emotions. *Cerebral Cortex*, *27*(3), Article 3. https://doi.org/10.1093/cercor/bhw028

Clore, G. L., & Ortony, A. (2013). Psychological construction in the OCC model of emotion. *Emotion Review*, *5*(4), 335–343. https://doi.org/10.1177/1754073913489751

Clore, G. L., Proffitt, D. R., & Zadra, J. R. (2021). Feeling, seeing, and liking: How bodily tesources inform perception and emotion. In M. D. Robinson & L. E. Thomas (Eds.), *Handbook of Embodied Psychology: Thinking, Feeling, and Acting* (pp. 43–64). Springer International Publishing. https://doi.org/10.1007/978-3-030-78471-3_3

Cole, M. W., Bassett, D. S., Power, J. D., Braver, T. S., & Petersen, S. E. (2014). Intrinsic and task-evoked network architectures of the human brain. *Neuron*, *83*(1), 238–251. https://doi.org/10.1016/j.neuron.2014.05.014

Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., & Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature Neuroscience*, *16*(9), 1348–1355. https://doi.org/10.1038/nn.3470

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215. https://doi.org/10.1038/nrn755

Dalgleish, T. (2004). The emotional brain. *Nature Reviews Neuroscience*, *5*(7), 583–589. https://doi.org/10.1038/nrn1432

Damasio, A. R. (1999). *The Feeling of what Happens: Body and Emotion in the Making of Consciousness*. Houghton Mifflin Harcourt.

Davis, M. (1992). The role of the amygdala in fear and anxiety. *Annual Review of Neuroscience, 15,* 353-375.

Dixon, M. L., De La Vega, A., Mills, C., Andrews-Hanna, J., Spreng, R. N., Cole, M. W., & Christoff, K. (2018). Heterogeneity within the frontoparietal control network and its relationship to the default and dorsal attention networks. *Proceedings of the National Academy of Sciences*, *115*(7), E1598–E1607. https://doi.org/10.1073/pnas.1715766115

Doyle, C. M., Lane, S. T., Brooks, J. A., Wilkins, R. W., Gates, K. M., & Lindquist, K. A. (2022). Unsupervised classification reveals consistency and degeneracy in neural network patterns of emotion. *Social Cognitive and Affective Neuroscience*, nsac028. https://doi.org/10.1093/scan/nsac028

Dunn, J. C., & Kirsner, K. (2003). What Can we Infer from Double Dissociations? *Cortex*, *39*(1), 1–7. https://doi.org/10.1016/S0010-9452(08)70070-4

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review*, *3*(4), 364–370. https://doi.org/10.1177/1754073911410740

Feldman, M. J., Bliss-Moreau, E., & Lindquist, K. A. (2024). The neurobiology of interoception and affect. *Trends in Cognitive Sciences*, S1364661324000093. https://doi.org/10.1016/j.tics.2024.01.009

Finlay, B. L., & Uchiyama, R. (2015). Developmental mechanisms channeling cortical evolution. *Trends in Neurosciences*, *38*(2), 69–76. https://doi.org/10.1016/j.tins.2014.11.004

Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience*, *8*(9), 700–711. https://doi.org/10.1038/nrn2201

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), Article 2. https://doi.org/10.1038/nrn2787

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience*, *6*(4), Article 4. https://doi.org/10.1080/17588928.2015.1020053

Fuster, J. M. (2000). The module: Crisis of a paradigm. *Neuron*, *26*(1), 51–53. https://doi.org/10.1016/S0896-6273(00)81137-X

Girn, M., Setton, R., Turner, G. R., & Spreng, R. N. (2024). The "limbic network", comprising orbitofrontal and anterior temporal cortex, is part of an extended default network: Evidence from multi-echo fMRI. *Network Neuroscience*, 1–40. https://doi.org/10.1162/netn_a_00385

Haak, K. V., Marquand, A. F., & Beckmann, C. F. (2018). Connectopic mapping with resting-state fMRI. *NeuroImage*, *170*, 83–94. https://doi.org/10.1016/j.neuroimage.2017.06.075

Halgren, E., Walter, R. D., Cherlow, D. G., & Crandall, P. H. (1978). Mental phenomena evoked by electrical stimulation of the human hippocampal formation and amygdala. *Brain*, *101*(1), 83-115. https://doi.org/10.1093/brain/101.1.83

Hamann, S. (2012). Mapping discrete and dimensional emotions onto the brain: Controversies and consensus. *Trends in Cognitive Sciences*, *16*(9), 458–466. https://doi.org/10.1016/j.tics.2012.07.006

Haxby, James. V. (2012). Multivariate pattern analysis of fMRI: The early beginnings. *Neuroimage*, *62*(2), 852–855. https://doi.org/10.1016/j.neuroimage.2012.03.016

Hitchcock, J., & Davis, M. (1986). Lesions of the amygdala, but not of the cerebellum or red nucleus, block conditioned fear as measured with the potentiated startle paradigm. *Behavioral Neuroscience*, *100*(1), 11–22. https://doi.org/10.1037/0735-7044.100.1.11

Hoemann, K., Wu, R., LoBue, V., Oakes, L. M., Xu, F., & Barrett, L. F. (2020). Developing an understanding of emotion categories: Lessons from objects. *Trends in Cognitive Sciences*, *24*(1), Article 1. https://doi.org/10.1016/j.tics.2019.10.010

Huntenburg, J. M., Bazin, P.-L., & Margulies, D. S. (2018). Large-scale gradients in human cortical organization. *Trends in Cognitive Sciences*, *22*(1), 21–31. https://doi.org/10.1016/j.tics.2017.11.002

Hutchinson, J. B., & Barrett, L. F. (2019). The power of predictions: An emerging paradigm for psychological research. *Current Directions in Psychological Science*, *28*(3), 280–291. https://doi.org/10.1177/0963721419831992

Izard, C. E. (1993). Four systems for emotion activation: Cognitive and noncognitive processes. *Psychological Review*, *100*(1), 68–90. https://doi.org/10.1037/0033-295X.100.1.68

Jackson, J. C., Watts, J., Henry, T. R., List, J.-M., Forkel, R., Mucha, P. J., Greenhill, S. J., Gray, R. D., & Lindquist, K. A. (2019). Emotion semantics show both cultural variation and universal structure. *Science*, *366*(6472), Article 6472. https://doi.org/10.1126/science.aaw8160

Kassam, K. S., Markey, A. R., Cherkassky, V. L., Loewenstein, G., & Just, M. A. (2013). dentifying emotions on the basis of neural activation. *PLoS ONE*, *8*(6), e66032. https://doi.org/10.1371/journal.pone.0066032

Katsumi, Y., Kamona, N., Zhang, J., Bunce, J. G., Hutchinson, J. B., Yarossi, M., Tunik, E., Quigley, K. S., Dickerson, B. C., & Barrett, L. F. (2021). *Functional connectivity gradients as a common neural architecture for predictive processing in the human brain* [Preprint]. Neuroscience. https://doi.org/10.1101/2021.09.01.456844

Katsumi, Y., Theriault, J. E., Quigley, K. S., & Barrett, L. F. (2022). Allostasis as a core feature of hierarchical gradients in the human brain. *Network Neuroscience*, *6*(4), 1010–1031. https://doi.org/10.1162/netn_a_00240

Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K., & Wager, T. D. (2008). Functional grouping and cortical-subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *NeuroImage*, *42*(2), Article 2. https://doi.org/10.1016/j.neuroimage.2008.03.059

Kragel, P. A., Koban, L., Barrett, L. F., & Wager, T. D. (2018). Representation, pattern information, and brain signatures: From neurons to neuroimaging. *Neuron*, *99*(2), 257–273. https://doi.org/10.1016/j.neuron.2018.06.009

Kragel, P. A., & LaBar, K. S. (2014). Advancing emotion theory with multivariate pattern classification. *Emotion Review*, *6*(2), 160–174. https://doi.org/10.1177/1754073913512519

Kuppens, P., Van Mechelen, I., Smits, D. J. M., & De Boeck, P. (2003). The appraisal basis of anger: Specificity, necessity and sufficiency of components. *Emotion*, *3*(3), 254–269. https://doi.org/10.1037/1528-3542.3.3.254

LaBar, K., LeDoux, J., Spencer, D., & Phelps, E. (1995). Impaired fear conditioning following unilateral temporal lobectomy in humans. *The Journal of Neuroscience*, *15*(10), 6846–6855. https://doi.org/10.1523/JNEUROSCI.15-10-06846.1995

LaBar, K. S., Gatenby, J. C., Gore, J. C., LeDoux, J. E., & Phelps, E. A. (1998). Human amygdala activation during conditioned fear acquisition and extinction: A mixed-trial fMRI study. *Neuron*, *20*(5), 937–945. https://doi.org/10.1016/S0896-6273(00)80475-4

Lamichhane, B., Adhikari, B. M., & Dhamala, M. (2016). Salience Network Activity in Perceptual Decisions. *Brain Connectivity*, *6*(7), 558–571. https://doi.org/10.1089/brain.2015.0392

LeDoux, J., & Daw, N. D. (2018). Surviving threats: Neural circuit and computational implications of a new taxonomy of defensive behaviour. *Nature Reviews Neuroscience*, *19*(5), 269–282. https://doi.org/10.1038/nrn.2018.22

Levenson, R. W. (1999). The intrapersonal functions of emotion. *Cognition & Emotion*. https://doi.org/10.1080/026999399379159

Lindquist, K. A., & Barrett, L. F. (2012). A functional architecture of the human brain: Emerging insights from the science of emotion. *Trends in Cognitive Sciences*, *16*(11), Article 11. https://doi.org/10.1016/j.tics.2012.09.005

Lindquist, K. A., Jackson, J. C., Leshin, J., Satpute, A. B., & Gendron, M. (2022). The cultural evolution of emotion. *Nature Reviews Psychology*, *1*(11), 669–681. https://doi.org/10.1038/s44159-022-00105-4

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,* 121-143. https://doi.org/10.1017/S0140525X11000446.

Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the BOLD Signal. *Annual Review of Physiology*, *66*(1), 735–769. https://doi.org/10.1146/annurev.physiol.66.082602.092845

MacCormack, J. K., & Lindquist, K. A. (2017). Bodily contributions to emotion: Schachter's legacy for a psychological constructionist view on emotion. *Emotion Review*, *9*(1), Article 1. https://doi.org/10.1177/1754073916639664

Margulies, D. S., Ghosh, S. S., Goulas, A., Falkiewicz, M., Huntenburg, J. M., Langs, G., Bezgin, G., Eickhoff, S. B., Castellanos, F. X., Petrides, M., Jefferies, E., & Smallwood, J. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences*, *113*(44), 12574–12579. https://doi.org/10.1073/pnas.1608282113

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: A network model of insula function. *Brain Structure and Function*, *214*(5), 655–667. https://doi.org/10.1007/s00429-010-0262-0

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., & Dolan, R. J. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature*, *383*(6603), 812–815. https://doi.org/10.1038/383812a0

Mos, J., Kruk, M. R., Van Poel, A. M. D., & Meelis, W. (1982). Aggressive behavior induced by electrical stimulation in the midbrain central gray of male rats. *Aggressive Behavior*, *8*(3), 261–284. https://doi.org/10.1002/1098-2337(1982)8:3<261::AID-AB2480080304>3.0.CO;2-N

Murphy, C., Wang, H.-T., Konu, D., Lowndes, R., Margulies, D. S., Jefferies, E., & Smallwood, J. (2019). Modes of operation: A topographic neural gradient supporting stimulus dependent and independent cognition. *NeuroImage*, *186*, 487–496. https://doi.org/10.1016/j.neuroimage.2018.11.009

Murphy, G. (2002). *The Big Book of Concepts*. The MIT Press. https://doi.org/10.7551/mitpress/1602.001.0001

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–430. https://doi.org/10.1016/j.tics.2006.07.005

Nummenmaa, L., & Saarimäki, H. (2019). Emotions as discrete patterns of systemic activity. *Neuroscience Letters*, *693*, 3–8. https://doi.org/10.1016/j.neulet.2017.07.012

Ochsner, K. N., Silvers, J. A., & Buhle, J. T. (2012). Functional imaging studies of emotion regulation: A synthetic review and evolving model of the cognitive control of emotion. *Annals of the New York Academy of Sciences*, *1251*, E1-24. https://doi.org/10.1111/j.1749-6632.2012.06751.x

Oligschläger, S., Xu, T., Baczkowski, B. M., Falkiewicz, M., Falchier, A., Linn, G., & Margulies, D. S. (2019). Gradients of connectivity distance in the cerebral cortex of the macaque monkey. *Brain Structure and Function*, *224*(2), 925–935. https://doi.org/10.1007/s00429-018-1811-1

Panksepp, J. (1986). Chapter 4—THE ANATOMY OF EMOTIONS. In R. Plutchik & H. Kellerman (Eds.), *Biological Foundations of Emotion* (pp. 91–124). Academic Press. https://doi.org/10.1016/B978-0-12-558703-7.50010-3

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

Panksepp, J. (2016). The cross-mammalian neurophenomenology of primal emotional affects: From animal feelings to human therapeutics. *Journal of Comparative Neurology*, *524*(8), 1624–1635. https://doi.org/10.1002/cne.23969

Panksepp, J., Knutson, B., & Pruitt, D. L. (1998). Toward a neuroscience of emotion. In M. F. Mascolo & S. Griffin (Eds.), *What Develops in Emotional Development?* (pp. 53–84). Springer US. https://doi.org/10.1007/978-1-4899-1939-7_3

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature Reviews Neuroscience*, *9*(2), Article 2. https://doi.org/10.1038/nrn2317

Pessoa, L. (2023). The Entangled Brain. *Journal of Cognitive Neuroscience*, *35*(3), 349–360. https://doi.org/10.1162/jocn_a_01908

Phan, K. L., Wager, T. D., Taylor, S. F., & Liberzon, I. (2004). Functional neuroimaging studies of human emotions. *CNS Spectrums*, *9*(4), 258–266. https://doi.org/10.1017/S1092852900009196

Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., Bullmore, E. T., Perrett, D. I., Rowland, D., Williams, S. C. R., Gray, J. A., & David, A. S. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature*, *389*(6650), 495–498. https://doi.org/10.1038/39051

Raichle, M. E. (2001). Bold insights. *Nature*, *412*(6843), 128–130. https://doi.org/10.1038/35084300

Raleigh, M. J., Steklis, H. D., Ervin, F. R., Kling, A. S., & McGuire, M. T. (1979). The effects of orbitofrontal lesions on the aggressive behavior of vervet monkeys (Cercopithecus aethiops sabaeus). *Experimental Neurology*, *66*(1), 158–168. https://doi.org/10.1016/0014-4886(79)90071-2

Roy, M., Shohamy, D., & Wager, T. D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends in Cognitive Sciences*, *16*(3), Article 3. https://doi.org/10.1016/j.tics.2012.01.005

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. https://doi.org/10.1037/0033-295X.110.1.145

Satpute, A. B., & Lindquist, K. A. (2019). The default mode network's role in discrete emotion. *Trends in Cognitive Sciences*, *23*(10), Article 10. https://doi.org/10.1016/j.tics.2019.07.003

Saviola, F., Pappaianni, E., Monti, A., Grecucci, A., Jovicich, J., & De Pisapia, N. (2020). Trait and state anxiety are mapped differently in the human brain. *Scientific Reports*, *10*(1), Article 1. https://doi.org/10.1038/s41598-020-68008-z

Seeley, W. W. (2019). The salience network: A neural system for perceiving and responding to homeostatic demands. *Journal of Neuroscience*, *39*(50), Article 50. https://doi.org/10.1523/JNEUROSCI.1138-17.2019

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., Reiss, A. L., & Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience

processing and executive control. *Journal of Neuroscience*, *27*(9), 2349–2356. https://doi.org/10.1523/JNEUROSCI.5587-06.2007

Sem-Jacobson CW. *Depth-electroencephalographic stimulation of the human brain and behavior.* Springfield, IL: Charles C. Thomas Publishing; 1968

Shaffer, C., Westlin, C., Quigley, K. S., Whitfield-Gabrieli, S., & Barrett, L. F. (2022). Allostasis, action, and affect in depression: Insights from the theory of constructed emotion. *Annual Review of Clinical Psychology*, *18*(1), 553–580. https://doi.org/10.1146/annurev-clinpsy-081219-115627

Shafiei, G., Markello, R. D., Vos de Wael, R., Bernhardt, B. C., Fulcher, B. D., & Misic, B. (2020). Topographic gradients of intrinsic dynamics across neocortex. *eLife*, *9*, e62116. https://doi.org/10.7554/eLife.62116

Shiota, M. N. (2024). Basic and Discrete Emotion Theories. In A. Scarantino (Ed.), *Emotion Theory: The Routledge Comprehensive Guide* (1st ed.). (pp. 310-330). Routledge.

Shulman, G. L., Pope, D. L. W., Astafiev, S. V., McAvoy, M. P., Snyder, A. Z., & Corbetta, M. (2010). Right hemisphere dominance during spatial selective attention and target detection occurs outside the dorsal frontoparietal network. *Journal of Neuroscience*, *30*(10), 3640–3651. https://doi.org/10.1523/JNEUROSCI.4085-09.2010

Sporns, O. (2014). Contributions and challenges for network models in cognitive neuroscience. *Nature Neuroscience*, *17*(5), 652–660. https://doi.org/10.1038/nn.3690

Sporns, O., Chialvo, D. R., Kaiser, M., & Hilgetag, C. C. (2004). Organization, development and function of complex brain networks. *Trends in Cognitive Sciences*, *8*(9), 418–425. https://doi.org/10.1016/j.tics.2004.07.008

Spreng, R. N., Mar, R. A., & Kim, A. S. N. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode:

A quantitative meta-analysis. *Journal of Cognitive Neuroscience*, *21*(3), 489–510. https://doi.org/10.1162/jocn.2008.21029

Sterling, P., & Laughlin, S. (2015). *Principles of Neural Design*. The MIT Press. http://www.jstor.org/stable/j.ctt17kk982

Theriault, J. E., Young, L., & Barrett, L. F. (2020). The sense of should: A biologically-based framework for modeling social pressure. *Physics of Life Reviews*. https://doi.org/10.1016/j.plrev.2020.01.004

Tomkins, S.S. (1962). *Affect, Imagery, Consciousness: Vol. 1. The Positive Affects.* New York: Springer

Tomkins, S.S. (1963). *Affect, Imagery, Consciousness: Vol. 2. The Negative Affects*. New York: Springer

Uddin, L. Q. (2016). *Salience Network of the Human Brain*. Academic Press.

van den Heuvel, M. P., & Sporns, O. (2011). Rich-club organization of the human connectome. *Journal of Neuroscience*, *31*(44), Article 44. https://doi.org/10.1523/JNEUROSCI.3539-11.2011

Vázquez-Rodríguez, B., Suárez, L. E., Markello, R. D., Shafiei, G., Paquola, C., Hagmann, P., Van Den Heuvel, M. P., Bernhardt, B. C., Spreng, R. N., & Misic, B. (2019). Gradients of structure–function tethering across neocortex. *Proceedings of the National Academy of Sciences*, *116*(42), 21219–21227. https://doi.org/10.1073/pnas.1903403116

Wang, Y., Kragel, P.A, & Satpute, A.B. (2024). Neural predictors of fear depend on the situation. *Journal of Neuroscience,* 44, e0142232024; **DOI:**10.1523/JNEUROSCI.0142-23.2024

Wager, T. D., Kang, J., Johnson, T. D., Nichols, T. E., Satpute, A. B., & Barrett, L. F. (2015). A Bayesian model of category-specific emotional brain responses. *PLoS Computational Biology*, *11*(4), 1004066. https://doi.org/10.1371/journal.pcbi.1004066

Whalen, P. J. (1998). Fear, vigilance, and ambiguity: Initial neuroimaging studies of the human amygdala. *Current Directions in Psychological Science*, *7*(6), 177–188. https://doi.org/10.1111/1467-8721.ep10836912

Wicker, B., Keysers, C., Plailly, J., Royet, J.-P., Gallese, V., & Rizzolatti, G. (2003). Both of is disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron*, *40*(3), 655–664. https://doi.org/10.1016/S0896-6273(03)00679-2

Yik, M., Mues, C., Sze, I. N. L., Kuppens, P., Tuerlinckx, F., De Roover, K., Kwok, F. H. C., Schwartz, S. H., Abu-Hilal, M., Adebayo, D. F., Aguilar, P., Al-Bahrani, M., Anderson, M. H., Andrade, L., Bratko, D., Bushina, E., Choi, J. W., Cieciuch, J., Dru, V., … Russell, J. A. (2023). On the relationship between valence and arousal in samples across the globe. *Emotion*, *23*(2), 332–344. https://doi.org/10.1037/emo0001095

Zanto, T. P., & Gazzaley, A. (2013). Fronto-parietal network: Flexible hub of cognitive control. *Trends in Cognitive Sciences*, *17*(12), 602–603. https://doi.org/10.1016/j.tics.2013.10.001

Zhang, J., Abiose, O., Katsumi, Y., Touroutoglou, A., Dickerson, B. C., & Barrett, L. F. (2019). Intrinsic functional connectivity is organized as three interdependent gradients. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-51793-7

Zhang, J.-X., Dixon, M. L., Goldin, P. R., Spiegel, D., & Gross, J. J. (2023). The neural separability of emotion reactivity and regulation. *Affective Science*, *4*(4), 617–629. https://doi.org/10.1007/s42761-023-00227-9

NOTES

[i] Note that current models of brain function see the concept of the limbic system, in particular, and the triune brain concept, more generally, as outdated (Cesario et al., 2020; Chanes & Barrett, 2016).

[ii] Note that the term heterarchical is preferred to hierarchical because it does not assume that projections exist in a single direction. See Lee et al. 2021 for a discussion.