Contents lists available at ScienceDirect

Neuroscience Letters

journal homepage: www.elsevier.com/locate/neulet

Review article

Historical pitfalls and new directions in the neuroscience of emotion

Lisa Feldman Barrett[a,b,*], Ajay B. Satpute[c]

[a] Department of Psychology, Northeastern University, Boston, MA, United States
[b] Athinoula A. Martinos Center for Biomedical Imaging and Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Charlestown, MA, United States
[c] Departments of Psychology and Neuroscience, Pomona College, Claremont, CA, United States

HIGHLIGHTS

- A dynamic, systems-based neuroscience view of emotion is proposed.
- An emotion word refers to a population of highly variable, situated instances.
- Emotions are not reactions to the world, but are predictions corrected by sensory inputs.
- Prediction signals are concepts that categorize and explain the causes of sensations.
- Affect is a property of consciousness and is not synonymous with emotion.

ARTICLE INFO

Article history:
Received 2 April 2017
Received in revised form 25 July 2017
Accepted 25 July 2017
Available online xxx

Keywords:
Emotion
Affect
fmri
Construction
Active inference
Predictive coding
Allostasis
Interoception
Theory of constructed emotion

ABSTRACT

In this article, we offer a brief history summarizing the last century of neuroscientific study of emotion, highlighting dominant themes that run through various schools of thought. We then summarize the current state of the field, followed by six key points for scientific progress that are inspired by a multi-level constructivist theory of emotion.

Contents

* Corresponding author at: Department of Psychology, Northeastern University, 360 Huntington Ave, Boston, MA, 02115 United States.
E-mail address: l.barrett@neu.edu (L.F. Barrett).

http://dx.doi.org/10.1016/j.neulet.2017.07.045

1. The brain basis of emotion: a brief history

The neuroscientific study of emotion began in the 19th century, when psychology was transformed from mental philosophy into a full-fledged scientific discipline. Scientists searched for the physical basis of fear, anger, sadness, among other mental categories, using the tools of physiology and neurology. During this era, the science of emotion was guided by a Linnaean-type taxonomy of emotion categories, a small set of which were each cast as independent mental organs with a unique neural circuit and/or pattern of physiological correlates (i.e., a physical fingerprint) waiting to be discovered.[1] This family of approaches [1,2], which we refer to as the classical view of emotions [3,4],[2] is an example of faculty psychology: – the belief that the human mind is structured as a set of mental abilities, each associated with a unique state caused by its own, unique biological system.[3] In the 19th century, faculties were the categories that described what the mind is (as a formal ontology), what the mind does (as a set of psychological processes), and how the mind is caused (i.e., a functional architecture of biological processes that create the psychological processes and corresponding mental phenomena). Scientists formed grand theories as they battled over whether emotions were constituted as biological changes in body or in brain (e.g., [5]). A compromise was struck—emotions were assigned to live in the ancient parts of the brain that control the body, dubbed the "limbic system", a.k.a., our inner beast, whereas cognitions were assigned to the cortex (mistakenly called the neocortex), like a crown designed for us by evolution (e.g., [6]. Eventually, these ideas became known as the triune brain concept [7].[4]

In the late 19th and early 20th centuries, studies accumulated in search of the biobehavioral basis of each emotion category, and scientists discovered that they were unable to identify any specific neural circuit or ensemble of autonomic nervous system changes (or facial movements, for that matter) that consistently distinguished the instances of one emotion category from another (for a review, see [156]). Some studies reported no biological or behavioral differences between categories, but more common was variation in distinctiveness. For example, the physical pattern for anger that distinguished it from sadness, fear and so on in one study was different from the pattern observed in other studies. Variation was the norm [3]. To deal with this variability, scientists recast emotions as functional states to be studied solely by manipulating their observable causes and measuring their observable effects.[5] Anger, sadness, fear, and so on were ill-defined (or defiantly not defined at all, e.g.,[6]). That is the beauty of functionalism; – a phenomenon can be studied as a set of causes and effects without ever operationalizing the phenomenon itself in scientific terms as a set of physical changes.

In the first half of the 20th century, scientific studies guided by functionalism continued to produce a plethora of highly variable, contextual findings. This quickly gave way to behaviorism-the mind disappeared as a topic of scientific study altogether, and neuroscience focused on the biology of behavior. Correspondingly, emotions were no longer considered mental phenomena but were instead recast as states of the nervous system that caused specific behaviors. The neural circuitry for freezing or fleeing became the circuitry for fear. The neural circuitry for fighting became the circuitry for anger. And so on. Emotion categories, like all psychological categories, were ontologically reduced to behaviors. Experience and other mental phenomena became epiphenomenal to emotion, if they were considered at all.

Psychology emerged from behaviorism in the 1950s and 60s and the mind was reinstated as a topic of scientific inquiry. However, it was faculty psychology all over again, this time with computers (rather than organs of the body) as the driving metaphor. Emotions were recast as computations (either metaphorically or literally). For example, an "affect program" [8,9] for anger was said to be an

[1] For example, some writers in the 19th century, such as Carl Lange [138], adhered to a classical view of emotion, proposed that instances of the same emotion category share a distinctive pattern of peripheral nervous system activity (a metaphorical "fingerprint") and that different emotion categories have distinct, diagnostic fingerprints. William James has been consistently misquoted as claiming the same (for a discussion, see [139]. Modern versions of the classical view of emotion (including basic emotion approaches and causal appraisal approaches) are united by a similar hypothesis regarding emotion fingerprints. Real fingerprints are used to identify a person because the pattern of ridges and valleys on each finger is unique. The observed pattern (i.e., the print of the ridges) varies somewhat from one instance to the next depending on the degree of pressure used, the surfaces touched, the amount of sweat present, and so on, but in principle, the unique fingerprint can still be identified as belonging to one individual and one individual only. In the same way, the peripheral nervous system pattern for "anger" need not be identical for every instance for the classical view of emotion to be correct, so some variation from instance to instance is permitted, but the pattern should be *sufficiently similar* to identify those instances as anger and to distinguish them from instances of other emotion categories (named by words such as "sadness," "fear," "disgust," or "happiness") that each have their own unique physical fingerprints. This means that the biological changes for a given emotion category should be *consistent* from instance to instance (i.e., the same fingerprint should be observed) and *specific* (a fingerprint should uniquely identify instances of one emotion category and only that emotion category [140] under review).

[2] The classical view of emotion is a category of models that share certain hypotheses and assumptions. Like all categories, the models in this grouping also vary from one another in certain ways. For example, some approaches focus more on the hypothesis of neural essences (e.g., [13] whereas others emphasize a distinct physiological pattern for each emotion category (e.g., [11]. Some emphasize the hypothesis that animal species share homologous emotional expressions whereas others (e.g., [141] explicitly reject that hypothesis and instead propose the emotion categories evolved in our Pleistocene ancestors.

[3] Faculties were thought to be innate and caused by an all-powerful force (by the gods in ancient Greece, by a single god in the Middle Ages, and eventually by natural selection after [3]

[4] The triune brain is a combination of Plato's tripartite psyche and Aristotle's and Darwin's phylogenetic scale, tattooed onto the human brain (see [3]. Plato wrote that the human psyche consists of three parts: rational thoughts, passions (which today we call emotions), and appetites like the drive for hunger and sex. Rational thought was in charge, controlling the passions and appetites, an arrangement that Plato described as a charioteer wrangling two winged horses

[5] There are two popular flavors of functionalism that can be found in the science of emotion. Input-output functionalism (also called black-box functionalism) defines an emotion by its causes and effects. Adaptational functionalism defines an emotion as a state that supposedly evolved to serve a particular utility or purpose (it is sometimes called the "intentional design stance", e.g., [142].The state is thought to be caused by an evolved program that is responsible for creating the evolved state. The evolutionary biologist Ernst Mayr made a cogent argument for avoiding functionalism when studying the features that contribute to the adaptedness of an organism (e.g., [143,p. 48]) because they encourage metaphorical language that cannot be verified in physical terms.

[6] "I do not propose to attempt any description of the emotional qualities nor of the bodily expressions of "the emotions". If the reader does not know what it is to be afraid, or angry, or disgusted .... no amount of description, however, eloquent, will enlighten him." [144,p. 328-9])

inherited central organizing mechanism that produced a state of anger [10] and later [11]. A "primary process" system for fear was said to produce a state of fear [12, and later,13,14]. An evaluation of novelty (referred to as an "appraisal" of novelty) was said to produce a state of novelty [15] and later [16]. Once again, an emotion was equated with the process that caused it. Emotion categories remained ontologically reduced to behaviors, not because experience was irrelevant, but in the name of evolution; non-human animals clearly have emotions, the argument went, but they do not necessarily have emotional experiences. And scientists wanted a species-general explanation of emotion to generalize their findings from rats to humans.[7]

In the early 1990s, the cognitive revolution gave birth to the field of cognitive neuroscience, with affective neuroscience following soon after. Faculty psychology and its taxonomy of mental categories still reigned [17] and the goal (again) was to carve the brain into a set of independently functioning mental organs: one set of neurons dedicated to each mental category. Accordingly, distinct sets of neurons were thought to cause distinct processes that produced distinct states of the same name. Just as memories were thought to be produced by memory processes localized to memory circuits, attention was thought to be produced by attentional processes localized to attention circuits, and sensations produced by sensory processes localized to sensory circuits, so too emotions were thought to be caused by emotion processes localized to emotion circuits. The hypothesis that each emotion category is caused by its own dedicated neural circuit is a hallmark of the classical emotion view, according to a recent review, which states that the "agreed-upon gold standard is the presence of neurons dedicated to the emotion's activation" [141,p. 398]. In fact, only some versions of the classical view entertain this hypothesis (e.g., the most recent version of Ekman's basic emotion theory no longer explicitly hypothesizes that each emotion category has a dedicated neural circuit; see [11]).[8] Nevertheless, the triune brain concept was retained in the classical view, sometimes implicitly, for example, with the prefrontal cortex (as the putative home of cognition) downregulating the amygdala (as the putative home of emotion).

2. The current state of things

History has repeated itself. As neuroscientific studies of emotion accumulated over the last twenty years, it seemed at first as if the neurons responsible for each anger, sadness, fear, and so on could each be localized to specific brain regions (e.g., [19]). It quickly became clear, however, that scientists were facing the same challenges as a century earlier: studies that were purposely designed to isolate the specific neural basis of a mental category and distinguish it from other categories were consistently unable to do so; although an individual study, on its own, might appear to support the neural specificity of a given emotion category, meta-analytic evidence does not support this hypothesis (e.g., [20,21]).

Take, for example, links between fear and the amygdala. Much of what we know about the amygdala's role in fear in humans comes from studying Patient S.M., who lost both her amygdalae to Urbach-Wiethe disease (for a review of findings and relevant references, see an excellent review by [22]).[9] S.M. has difficulty experiencing fear in many normative circumstances (e.g. horror movies, haunted houses), but she experiences intense fear during experiments where she is asked to breathe air with higher concentrations of CO_2. She mounts a normal skin conductance response to an unexpectedly loud sound, but her brain seems unable to use arousal as a learning cue in milder laboratory situations (e.g. what used to be called 'fear learning' paradigms) and she therefore has difficulty learning from prior errors (e.g. she does not mount an anticipatory skin conductance response to aversive stimuli, during the Iowa Gambling Task, and she does not show loss aversion when gambling). Interestingly, however, there is evidence from S.M.'s everyday life that she is capable of aversive learning (i.e., 'fear learning') in certain circumstances. For example, S.M. is averse to breaking the law for fear of getting in trouble. She also spontaneously reports feeling worried. She 'learns fear' in the real world in certain circumstances, evidenced by the fact that she avoids seeking medical treatment and visiting the dentist because of pain she experienced on a previous occasion. Yet, S.M. also shows evidence of failure to "learn fear" in everyday life; for example, she returned to a park where she was attacked at knifepoint the previous day [23]. Therefore, an intact amygdala does not appear to be *necessary* to organize defensive responses to perceived threats in all contexts. This pattern of findings is consistent with the existence of degenerate circuitry for fear [3]: degeneracy exists when structurally distinct mechanisms perform the same function [24]. For example, animal species that have been engineered to knockout selected genes have some number of individuals (up to 30%) who continue to show the phenotype despite the absence of the selected genes (see [24]). Another example of degeneracy for fear can be found in the case of monozygotic twins, both of whom suffered amygdala damage from Urbach-Wiethe disease, but only one of whom shows problems with processing fear-inducing stimuli [25].

The hypothesis that there may be degenerate circuitry supporting the emergence of fear is also supported by the study of non-human animals. For example, macaques monkey with amygdala lesions (in the central nucleus) freeze less when exposed to a potential threat (called the human intruder paradigm; [26].[10] Other studies indicate, however, that macaques with neonatal amygdala ablations show blunted responses to all evocative stimuli (not specific to threat; [27,28] and are able to learn about new threats via associative learning (i.e., 'fear learning'; see [29] under review) so that their socio-affective function appears to be relatively normal by adulthood [29]. In addition, there is some evidence that, in certain conditions, rodents without an intact amygdala can show normal responses in situations involving threat (e.g., [30].

Reviews of neuroscience research within a variety of measurement domains generally do not support the hypothesis that an individual emotion category can be consistently and specifically localized to a specific swath of brain tissue. The circuitry for a given emotion category has yet to be consistently and specifically localized to *individual neurons* (e.g., see [31] for a review of intracranial recording research on humans; see [32,33] for a review of electrical stimulation research on non-human animals), to *a brain region* (for a recent meta-analysis, see [21]), or to *a brain*

[7] There are some exceptions. For example, Panksepp [13] argued vigorously that non-human animals share human-like emotional experiences.

[8] Ekman's version of basic emotion theory, for example, is most clearly tested with observations about autonomic nervous system changes and facial movements (referred to as emotional expression). Specifically, Ekman and colleagues hypothesize that each family of emotion categories (i.e., the anger family) has distinctive patterns of autonomic nervous system physiology that distinguish them from the family of sadness, fear, and so on. Furthermore, they presume that facial movements express an internal emotional state, which translates into the hypothesis that certain configurations of facial movements (or facial actions) co-occur with other measurements of emotion, such as autonomic nervous system measurements, self-reports, or other behaviors. The data supporting these two hypotheses exist in a much larger pool of evidence which largely calls them into doubt (e.g., see [58,89,3,140,145–150]).

[9] S.M.'s brain shows abnormalities that extend beyond the amygdala, including the anterior entorhinal cortex and ventromedial prefrontal cortex, both of which show dense, reciprocal connections to the amygdala and may play a role in S.M.'s specific behavioral profile.

[10] Lesions in orbitofrontal cortex create similar behavioral changes [26].

network (for a review, see [34]; for an example of an individual experiment, see Touroutolgou et al., 2015). A growing number of brain imaging studies have successfully distinguished emotion categories from one another using multivariate patterns of activity distributed across the brain (i.e., pattern classification), but again, when it comes to the observed pattern for any single emotion category, variation is the norm. The diagnostic patterns observed for anger, sadness, fear, disgust and so on, are highly variable from study to study (e.g., compare patterns reported in [35–37] and the patterns derived from meta-analyses as reported in [38]).

Taken together, these findings call the classical view of emotion into doubt.[11] In response, functionalism has again been invoked to sustain the viability of the classical view (e.g., [39,40]). Functionalism introduces a host of scientific problems, however, the most serious being that it is rooted in psychological essentialism that renders the classical view non-falsifiable [41]. Psychological essentialism [42] allows people to define a psychological phenomenon by its causes and consequences while positing a hypothetical state and its hypothetical causal mechanism. People continue to believe in the existence of the state and its causal mechanisms, even when they remain unmeasurable and unspecified. Ekman's metaphorical affect program for fear, Panksepp's hypothetical FEAR system, and Adolph's proposed 'central fear state' are all examples of psychological essentialism.

The problems of essentialism are also on display when scientists interpret pattern classifiers as neural "signatures" [35], "fingerprints" [43], and "biomarkers" [36] of emotion categories. Such words imply that the pattern for each emotion category is equivalent to the brain state for that category – something like a neural essence – such that successfully diagnosed instances of the category must contain its features that are both unique to that category and present in every instance.[12] Such interpretations represents a misunderstanding of pattern classification, however [44,45]. The derived patterns for each emotion category need not be observed in every (or even *any*) single instance of that category that is successfully classified (for an explanation and mathematical simulation, see [44,45]).[13]

Just as brain imaging studies find substantially different patterns of brain activity distinguish different emotion categories from one another across studies, studies that successfully distinguish between emotion categories using ANS measures report *different patterns from one study to the next,* even when those studies use the same stimuli, methods, and sample participants from the same population (e.g., compare findings from [36] with [157]). Empirical support for the classical view requires a *unique* pattern of physiological correlates (i.e., a physical fingerprint) that is consistent and specific for each emotion category. Instead, the body of published findings gives evidence of *variable* patterns for each emotion category across different contexts [e.g. 128,also see 130].[14] This is not evidence that some studies support the classical view of emotion while others fail to. Instead, it is evidence that the physical correlates of emotion, be they neural or autonomic, are highly variable within an emotion category from instance to instance (and the physiological changes associated with one emotion category in one study can be similar to the changes associated with a different emotion category in a different study). It is evidence that, when it comes to the biobehavioral correlates of any emotion category, variation is the norm. This degree of variability demands a mechanistic explanation that goes beyond explaining it away as epiphenomenal to emotion (i.e., it is not merely due to emotion regulation, display rules, insufficient laboratory methods, etc.).

Lesion and optogenetic studies of freezing, fighting, and fleeing in non-human animals are often interpreted as providing strong support for the unique neural basis of different emotion categories. A closer inspection reveals variable findings consistent with variation observed in humans. Again, turning to studies of behaviors that have been defined as "fear," various circuits acting in parallel or collaboratively support learning when to perform these behaviors. Studies have documented many (circuits) to one (behavior) mappings while others find one (circuit) to many (behavior) mappings [46,47]. Still other studies find that the amygdala, or specific parts (e.g. the basolateral nuclei) are not necessary for the expression of previously learned aversive behaviors (i.e. the way that learning is expressed depends on the context and the available options for behavior; [30]).

Furthermore, consider that animals perform many different behaviors in response to threat, each of which is supported by its own specific circuitry. For example, rats avoid the location of uncertain threat when they are free to move around, such as in a testing chamber with several arms (e.g., [158]); when they are not free to move around, they freeze. When threat is certain and they cannot escape, rats respond with defensive aggression – they kick bedding towards the threatening object (e.g., [159]) or jump on it and bite (e.g., [160]). And the circuitry that supports defensive aggression towards a cat (a predator) is distinct from that supporting defensive aggression towards another rat (a dominant intruder in the cage; [161]). Freezing, fleeing, and defensive aggression towards various predators are all responses to potential danger, so which corresponding circuit is the real fear circuit? Is this evidence of many fear circuits? If so, then what is the scientific value of the category "fear," other than its obvious utility for communicating research findings to people who do not study emotion? How do we know that defensive aggression is an instance of fear and not an instance of anger? Obviously, differential behavioral responses are not random – actions are contextually situated [48,49]. Therefore, context is important for understanding the variable biobehavioral patterns

[11] For more background on these issues and more nuanced scientific critiques, see Barrett [3], particularly chapter 8, and 2017b), Barrett et al. [164], Crivelli & Gendron [146], LeDoux [52], Lindquist et al. [21] and Quigley & Barrett [149]; for a defense of the classical view, see Adolphs [39], Izard [151], Keltner & Cordaro [152], Panksepp [153], and Tracy & Randles [18].

[12] A "biomarker" is defined as "a broad subcategory of medical signs ;– that is, objective indications of medical state observed from outside the patient, – which can be measured accurately and reproducibly"; they are "objective, quantifiable characteristics of biological processes" [165], p. 463). Importantly, a "sign" has causal relation to its process. This means that a biomarker should reliably diagnose a state, condition, or phenomenon, regardless of context. For example, cancerous tumors often shed genetic material into the bloodstream that is then used as a biomarker to detect the underlying (less easily measurable) condition Liotta et al., 2003. To describe a voxel-based classifier as a biomarker, then, is to imply that the specified configuration of voxels is present and unchanging in all the instances of the same emotion category (in the same way that DNA from a tumor can be used to uniquely identify the presence of a cancer). According to Wikipedia, a biomarker is "a measurable substance in an organism whose presence is indicative of some phenomenon." https://en.wikipedia.org/wiki/Biomarker (Nov 27, 2015). A successful pattern classifier yields an abstract statistical summary that is produced from a population of highly variable instances. This means that the specific nature of all instances of a given emotion category cannot be inferred from the locations and configurations of the voxels that most effectively distinguish instances of one emotion category from another in a particular study (i.e., the pattern for an emotion category in a particular study does not necessarily reveal the mechanisms that create all instances of that emotion category; see [44].

[13] A classic behavioral study by Posner & Keele [154] demonstrated a similar phenomenon almost half a century ago (using abstract categories of random dot configurations)

[14] A recent meta-analysis from our lab, for example, summarized findings from over 200 experiments measuring autonomic nervous system reactivity during instances of emotion categories and failed to find distinct autonomic fingerprints for any emotion category [140] under review). Individual studies often find differences in ANS measures during anger, sadness, fear, disgust and happiness, but the differences do not replicate from one study to the next. This translates into tremendous ANS variation both within and across categories.

within any emotion category, as well as the potential similarities across categories.

A more serious issue is whether neuroscience studies of behavior in non-human animals can ever be used to test the hypothesis of specific neural circuits for emotion categories in the first place. The neural circuitry for freezing only reveals something about "fear" when you equate emotion with behavior and stipulate, at the outset, that freezing occurs during a state of fear (and only during fear). Equating freezing with fear is an example of mentalizing on the part of the scientist – it is inferring a mental state that is not directly visible. Thus, interpreting neural evidence about circuitry for behavior as evidence for the brain basis of emotion rests on the scientific justification and validity of mental inference [3]. The role of mental inference in interpreting neuroscience research on non-human animals is rarely discussed explicitly, and as a consequence, many scientists who work in that domain routinely fail to recognize the role that their own (often implicit) mental inferences automatically play in interpreting their data. They refer to circuits as controlling fear when in fact they are studying context-dependent behaviors that may not have a one-to-one correspondence with a category of fear. Animal models that allow for probing the brain basis of behavior may be extremely useful for providing hypotheses about the neurobiology of human emotions (even if there is a many-to-many mapping of actions to emotion categories; e.g., [50]).

One way forward is to do away with emotion words and just study the variable instances themselves. One lesson of behaviorism, however, is that a science of instances is not viable. Induction and generalizability depend on the ability to form categories. The problem is not in forming a category – it is in essentializing the category. A category is a group of objects or instances that are treated as similar for some purpose. To essentialize is to believe that these similarities are fixed and real in nature in a perceiver-independent (or ontologically objective) way rather than created by a human mind. Essentialism is the belief that category members (like instances of fear) all share a deep, unchanging feature (or set of features) – an essence – giving the category its identity. Any attempt to localize a mental category to a single set of dedicated of neurons, in any form, rests on essentialism. From the standpoint of the classical view, the way to deal with the tremendous variation observed thus far in the neuroscientific study of emotion is to create more fine-grained taxonomies, attempting to bring nature under control and make it easier to identify the unique neural essence of each emotion category. Yet it is unclear whether this approach has thus far produced a generalizable and generative science of emotion.

3. A multi-level constructionist approach to the study of emotion

Another approach is to follow the data rather than subjugate it. The weight of evidence across different neuroscientific domains suggests the need for a contextually sensitive, constructionist approach to understanding the neurobiological basis of emotion. In the science of emotion, past constructionist approaches largely ignored biology: social construction focused on how emotions are influenced by social roles and values and psychological construction focused on how emotions emerge from more basic psychological processes related to making meaning of affective feelings. But a new breed of constructionist approaches is working to integrate social and psychological construction with neural construction (research that focuses on how experience-driven brain development wires a brain that can create emotions;for examples, see [51–55,56] [162]). The *theory of constructed emotion* is an example of this approach (see [118,120,121,3,4] and we use it here as a guiding example for the future of neuroscientific investigations of emotion to give a wide-ranging biologically plausible account of how emotions, as mental events, are constructed.[15]

3.1. A recognition that definitions of emotion are stipulated, not discovered

In philosophy, a stipulation is definition by fiat. The history of emotion research reveals that emotions are treated as discoveries, but in fact they are prescriptions. Scientists began with a mental framework for emotion (using categories from mental philosophy) which then dictates the sorts of questions asked, experiments designed, and interpretations offered (and even what counted as data in the first place). It is only this stipulation – a mental inference linking observed actions (e.g., freezing) to certain function or goals (e.g., for fear) – that allows scientists to claim that the circuitry for the actions is evidence for emotion circuits. What we learn (or fail to learn) about emotion in any experiment is determined by how we define emotions in the first place. Rather than starting with emotion categories and searching for their physical basis in the brain and the rest of the nervous system, the theory of constructed emotion begins with what is known about the structural and functional organization of nervous systems more generally and asks how they create emotions (i.e. a reverse engineering approach). Ultimately, this approach is revealing that emotional events are constructed by domain general processes, meaning that the category of emotion is not a natural kind [57], just as emotion categories themselves are not natural kinds [58].

3.2. Allostasis, interoception and affect are at the core of the brain

A brain is one big structure composed of neurons that pass information back and forth to one another. This structure is bathed in a chemical system (supported by "helper" glial cells) that continually modifies the ease and speed with which neurons share information. Information flow is also managed, in part, by inhibitory neurons whose activity fluctuates dynamically in time (e.g., [59,60]) rather than by a static modularity that is anatomically built into its architecture. Ongoing fluctuations in information flow allow this single physical structure to realize an astoundingly large number of spatiotemporal patterns as neurons dynamically coordinate their activity into different coherent states (i.e., the brain is a complex system). Part of the brain's complexity comes from its ability to generate novel representations by combining ("remembering") bits and pieces of past experience; in short, the brain is a complex, information gaining system [61,62]. This more dynamic systems-based view of the brain makes a compelling a case for moving away from faculty psychology's search for independent mental organs in the brain and thereby relinquishing essentialism (because an instance of perception, emotion, or any mental event, is better thought of as a brain state than as one brain region passing perceptual information to another, which computes emotion, after which it passes information to yet another brain regions, which launches behavior (see [57]).

Furthermore, brains did not evolve for rationality, happiness, or accurate perception. Brains evolved, in part, to efficiently ensure resources for physiological systems within an animal's body (i.e., its internal milieu) so that an animal can grow, survive, and reproduce [63]. This balancing act is called *allostasis* [64]. The brain networks that are most important for implementing allostasis (called the "default mode" and "salience" networks) also represent the sensory consequences of allostasis (i.e., interoception; [65,66]; for related discussions, see [67–69].

[15] The specific details of this theory are beyond the scope of this paper but interested readers are referred to [65,104,66].

The two interconnected intrinsic brain networks that regulate allostasis and represent interoception are also at the core of many other psychological phenomena, including memory, decision making, theory of mind, attention, and a host of others [66], and form the backbone for neural communication in the brain [70]. Such findings prompt the intriguing hypothesis that, whatever else a brain is doing – thinking, feeling, perceiving, preparing for action – it is also implementing allostasis in the service of behavior – regulating your autonomic nervous system, your immune system, and your endocrine system – as well as representing interoceptio. .

Interoception is made available to consciousness as lower dimensional feelings of affect [71]. As a consequence, the properties of affective feelings—valence and arousal [72,73]—are basic features of consciousness [74–80] that, importantly, are not unique to instances of emotion. We hypothesize that emotions (as opposed to cognitions) are constructed when interoceptive sensations are intense or when the change in affect is large and foregrounded in awareness. This may help explain why, in mammals, the brain regions that are responsible for establishing and maintaining allostasis ("limbic" regions such as the amygdala, ventral striatum, insula, orbitofrontal cortex, anterior cingulate cortex, and medial prefrontal cortex) are usually assumed to contain the circuits for emotion – they are important for regulating metabolism and energy expenditures associated with intense affect; for a discussion, see [66]. In fact, many of these limbic regions happen to be some of the most highly connected regions in the cortex [81], and they exchange information with midbrain, brainstem, and spinal cord nuclei that coordinate autonomic, immune, and endocrine systems with one another, and with the systems that control skeletomotor movements [82].

This constructionist approach shifts the focus in neuroscientific investigations of emotion in two related ways. First, an identity relationship between an emotion state and its physical causes is no longer assumed. Second, there is the potential to unite cognitive, perceptual, social and affective domains of neuroscience into a unified framework that places metabolism and energy regulation at the core of all mental activity [65,66]. As a consequence, traditional issues surrounding reverse inference [83] become a feature instead of a problem (for a similar view, see [84,85]). All mental events appear to be constructed within a domain-general functional architecture of the brain [34,86]. What we learn about emotion may be generalizable to understanding the mechanisms for other mental phenomena (and vice versa).

3.3. A focus on functionally integrated brain systems

As a network, the brain has large-scale topographical and dynamic properties. Both structurally and functionally, it can be understood as a set of dynamic communities or subnetworks (e.g., [86,87]) which, confusingly, are themselves usually referred to as "networks" (and so we follow that tradition in this paper). Brain networks have the property of homeostasis, meaning that there is a shifting population of neurons that maintain a network over time (e.g., [88]. Accordingly, a fruitful approach may be to understand emotional events as dynamic configurations (i.e., "recipes) of networks and their shifting neural constituents [57,17]. The brain's network architecture is dynamic, not static, but it is reasonably conserved across different brain states (see [86,87]). A key direction of future research, then, is understanding the spatiotemporal dynamics of network integration that creates the brain states that implement individual emotional events.

3.4. A focus on meaning-making

Emotions are not reactions to the world. We hypothesize that they are constructions of the world (or more specifically, they are your brain's construction of your bodily sensations and movements in the immediate context). They are a brain's explanation for bodily sensations in relation to the surrounding situation. Therefore, the biology of meaning making should figure prominently in the neuroscientific understanding of what emotions are, how they are caused and how they work. The theory of constructed emotion, as a biopsychosocial approach, explicitly includes the neural architecture that implements meaning making [57,89,3,4]. For example, neuroimaging studies of emotion consistently reveal that the default mode network is engaged as people experience emotion and simulate behaviors that are labeled as emotional [21]; these findings are instructive, because they suggest that the same circuitry that is necessary for running a mental model of the world [90], and that is responsible for implementing semantic knowledge [91]; also see [92,56]), for mentalizing [93–98], and for creating multimodal concepts [99,100], including emotion concepts [101,102], is also important to constructing instances of emotion. Of course, more research is necessary to identify the mechanisms of meaning making in the emergence of emotions. Nonetheless, the neuroscientific study of emotion has something to teach us about the biology of meaning-making more generally.

3.5. A focus on prediction

Meaning-making within the brain is a predictive activity [103]. In contrast to traditional stimulus-response "feedforward" frameworks, recent predictive coding (a.k.a. active inference, belief propagation) models suggest that the brain functions as Bayesian filters for incoming sensory input, *guiding* action and constructing perception (e.g., [65,104–109,68].[16] Past experiences are reconstructed as partial neural patterns that serve as prediction signals (also known as "top-down" or "feedback" signals, and more recently as "forward" models) to continuously anticipate events in the sensory environment, in part, by planning for motor and visceromotor changes.[17] Without past experience as a guide, the brain cannot transform flashes of light into sights, chemicals into smells, and variable air pressure into music, or interoception into emotions. The result is experiential blindness.[18] Prediction signals are embodied, whole brain representations that anticipate (1) upcoming sensory events both inside the body and out as well as (2) the best action to deal with the impending sensory events. From this perspective, unanticipated information from the world (prediction error) functions as feedback that can be learned to modify predictions (also known as 'bottom-up' or, confusingly, 'feedforward' signals). Error signals track the difference between the predicted sensations and those that are incoming from the sensory world (including the body's internal milieu). Once these errors are minimized, predic-

[16] Notably, Buzsaíki wrote that "Brains are foretelling devices" Buzsaíki, 2006 and there is accumulating evidence that prediction and prediction error signals oscillate at different frequencies within the brain (e.g., [166–168]).

[17] The term "feedback" derives from a time when the brain was thought to be largely stimulus driven Sartorius et al., 1993. Nonetheless, the history of science is laced with the idea that the mind drives perception, e.g., in the 11th century by Ibn al-Haytham (who helped to invent the scientific method), in the 18th century by Kant [169], and in the 19th century by Helmholtz. In more modern times, see Craik's concept of internal models (1943), Tolman's cognitive maps (1948), Johnson-Laird's internal models, and for more recent references, Neisser, [170] and [171]). The novelty in recent formulations can be found in (1) the hypothesis that predictions are *embodied* simulations of sensory motor experiences, (2) they are ultimately in the service of allostasis and therefore interoception is at their core, and, of course (3) the breadth of behavioral, functional, and anatomic evidence supporting the hypothesis that the brain's internal model implements active inference as prediction signals, including (4) the specific computational hypotheses implementing a predictive coding account.

[18] Simulations also constitute representations of the past (i.e., memories) and the future (i.e., prospections), and implement imagination, mind wandering, and daydreams ([172,155]).

tions also serve as inferences about the causes of sensory events and plans for how to move the body (or not) to deal with them [103,108]. By modulating ongoing motor and visceromotor actions to deal with upcoming sensory events, a brain infers their likely causes. Neuroscience evidence supports predictive coding accounts for every exteroceptive sensory system (see [104] for examples), and the same predictive dynamics may also be true for sensations arising from inside the body [65,66].

From this perspective, emotions are not organized reactions to the world. They are guesses about what to do next, rooted in prior experience, and the sensory consequences of those guesses (as well as the sights, sounds, smells and other experiences of the world). Emotions may be the Bayesian filters, predictions, or active inferences (what we have referred to as embodied concepts)—the representations that typically dominate as intrinsic brain activity. The neural representation of emotion, then, may be *created* from memory, rather than merely associated with or triggered by memory [3,4]; this hypothesis is consistent with meta-analytic evidence that the neural basis of emotions routinely involves traditional memory-related networks such as the default mode network [20,21,38,110]. An abundance of behavioral work, and a handful of neuroimaging studies (e.g. [111,112]), have emphasized the role of top-down knowledge in emotion construction (see reviews by [113,114]; [56] under review). But few, if any, studies properly test a predictive coding perspective on emotion (in part because experimental studies are designed as a sequence of independent stimulus-response trials that force the brain into a mode where prediction error dominates; [3]). Current models posit that predictions and prediction errors are interwoven throughout subcortical and cortical brain regions and organized by lamina [4,65,104], which scientists can now observe using high-resolution 7T scanning [162,163]. Moving beyond standard neuroimaging studies, a predictive coding inspired model of emotion requires a dynamic, high-resolution, and distributed approach.

3.6. A focus on variability, rather than typologies, biomarkers, affect programs, and central states

Because emotion categories do not appear to have a person-independent set of markers (in the autonomic nervous system, in behaviors, or in facial movements), scientific investigations should be explicitly designed to capture the variety within an emotion category, ideally using multiple measures in multiple contexts. This is an example of population thinking, which is one of the conceptual innovations found in Darwin's *On the Origin of Species* (for discussion, see [3]).[19]

Consistent with this approach, a growing number of neuroimaging studies now reveal that, when it comes to emotion, variety is the norm and attempt to study this variety in a situated, contextually variable way (e.g., [115,80]). A given emotion is a category of variable instances that are tuned to the situation at hand. As situations vary so do the instances of an emotion category. People smile when sad, cry when angry, and scream when happy. A person can tremble, jump, freeze, scream, gasp, hide, attack, and even laugh all in the face of fear. Just instances of the same emotion category can vary across contexts, they can also show variability within the same contexts. This form of variation – degeneracy [24] – has almost been ignored brain imaging experiments of emotion (although degenerate circuitry for behaviors can be found in some of the non-human animal work, as described above). Degeneracy is ubiquitous in biology (for examples, see [24,88,62]), and natural selection favors systems with degeneracy because they are high in complexity and robust to damage [116]. A context change reveals a hidden reservoir of variation for selection processes to act on immediately, without waiting for reproduction and mutation. When it comes to the brain, this translates into preserved emotional function in the face of damage and disease (for examples, see [3]). Thus, future neuroscience investigations of emotion might focus more on the capacity for dissimilar representations (e.g., different sets of neurons) to give rise to instances of the same mental category (e.g., anger) in the same context (i.e., many-to-one mappings of structure to function).

Understanding the neural basis of fear (or any emotion category) will benefit from models that are designed to explain all forms of variability. Such variation is unlikely to rely on a core "fear" circuit (i.e. neural essence) with minor fluctuations or modifications, nor be captured by a handful of narrow "fear" categories. The collection of instances referred to as "fear" is more likely to rely on a distributed architecture, no individual piece of which is essential across instances. Nor across individuals (given the likelihood of degeneracy). Future work would benefit from taking an idiographic (individual and instance dependent) approach is preferable to a nomothetic approach.

4. Conclusion

In all areas of natural science – physics, chemistry, and biology – progress has involved a shift away from the essentialism of classical views towards a more dynamic, contextual and constructionist approach to the physical world. Neuroscience, with advances in neuroimaging methods and analysis, are following a similar path, achieving astounding discoveries, some of which transform our basic understanding of how brain creates mind. From this perspective, emotions should be studied as dynamic, highly variable whole brain constructions of *what bodily sensations mean* in the context of the immediate environment. The mechanisms that implement these constructions are not specific to the domain of emotion but operate across cognitions, perceptions, and action, so that the scientific knowledge of one domain is of value for understanding the others. The next generation of neuroscience investigations, with a focus on whole-brain network dynamics along with time-varying, multivariate analytic approaches, is well-positioned to make landmark discoveries about the neural basis of emotion.

Acknowledgements

This paper was prepared with support from the National Institutes on Aging (R01 AG030311) and the National Cancer Institute (U01 CA193632) to LFB.

References

[1] L.F. Barrett, Navigating the science of emotion, in: H. Meiselman (Ed.), Emotion Measurement, Elsevier, Oxford, England, 2016, pp. 31–63.
[2] J.J. Gross, L.F. Barrett, Emotion generation and emotion regulation: one or two depends on your point of view, Emot. Rev. 3 (2011) 8–16.

[19] Before *On the Origin of Species*, a "species" was defined as biological type (i.e., with a set of unchanging physical characteristics or features that are passed down through the generations). This typological characterization fundamentally underestimates within-category variation (in its phenotypic features and in it's gene pool) and overestimates between category variation (and borderline cases are often encountered; [143]. One of Darwin's greatest conceptual innovations in *Origin* was to revolutionize the concept of a species as a biopopulation of highly variable individuals (instead of a group of highly similar creatures who share a set of co-occurring biological features) [143]. Since then, the concept of a "species" has been characterized on the basis of what category members do (i.e., functionally), not on the basis of a shared gene pool or a set of physical features: A species is a reproductive community (sometimes, members of different species are reproductively incompatible; sometimes they don't, such as when they are geographically isolated). Fundamentally, this translates into the insight that a biological category (a "species") is a conceptual category, rather than a typological one: a species is a population of physically unique individuals who similarities are defined functionally, not physically.

Please cite this article in press as: L.F. Barrett, A.B. Satpute, Historical pitfalls and new directions in the neuroscience of emotion, Neurosci. Lett. (2017), http://dx.doi.org/10.1016/j.neulet.2017.07.045

[3] L.F. Barrett, How Emotions Are Made: The Secret Life the Brain, Houghton-Mifflin-Harcourt, New York, NY, 2017.
[4] L.F. Barrett, The theory of constructed emotion: an active inference account of interoception and categorization, Soc. Cogn. Affect. Neurosci. (2017), http://dx.doi.org/10.1093/scan/nsw154.
[5] W.B. Cannon, Bodily Changes in Pain, Hunger, Fear, and Rage: An Account of Recent Researches into the Function of Emotional Excitement, D. Appleton and company, 1915.
[6] J.W. Papez, A proposed mechanism of emotion, Arch. NeurPsych. 38 (1937) 725–743.
[7] P.D. MacLean, The Triune Brain in Evolution: Role in Paleocerebral Functions, New York, NY, Plenum Press, 1990.
[8] S. Tomkins, Affect Imagery Consciousness: Volume i: The Positive Affects, Springer Publishing, New York, NY, 1962.
[9] S. Tomkins, Affect Imagery Consciousness: Volume II: The Negative Affects, Springer Publishing, New York, NY, 1963.
[10] P. Ekman, Universals and cultural differences in facial expressions of emotion, in: J. Cole (Ed.), Nebraska Symposium on Motivation, 1971, vol 19, University of Nebraska Press, Lincoln, 1972, pp. 207–282.
[11] P. Ekman, D. Cordaro, What is meant by calling emotions basic, Emot. Rev. 3 (4) (2011) 364–370.
[12] J. Panksepp, Toward a general psychobiological theory of emotion, Behav. Brain Sci. 5 (1981) 407–467.
[13] J. Panksepp, Affective Neuroscience: The Foundations of Human and Animal Emotions, Oxford University Press, New York, 1998.
[14] J. Panksepp, D. Watt, What is basic about basic emotions? Lasting lessons from affective neuroscience, Emot. Rev. 3 (4) (2011) 387–396.
[15] K.R. Scherer, Emotion as a multicomponent process: a model and some cross-cultural data, Rev. Personal. Soc. Psychol. (1984).
[16] K.R. Scherer, Emotions are emergent processes: they require a dynamic computational architecture, Philos. Trans. R. Soc. Lond.: Ser. B. Biol. Sci. 364 (2009) 3459–3474, http://dx.doi.org/10.1098/rstb.2009.0141.
[17] K.A. Lindquist, L.F. Barrett, A functional architecture of the human brain: insights from the science of emotion, Trends Cogn. Sci. 16 (2012) 533–540 (PMC3482298).
[18] J.L. Tracy, D. Randles, Four models of basic emotions: a review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt, Emot. Rev. 3 (4) (2011) 397–405.
[19] F.C. Murphy, I. Nimmo-Smith, A.D. Lawrence, Functional neuroanatomy of emotions: a meta-analysis, Cognit. Affect. Behav. Neurosci. 3 (2003) 207–233.
[20] H. Kober, L.F. Barrett, J. Joseph, E. Bliss-Moreau, K.A. Lindquist, T.D. Wager, Functional grouping and cortical-subcortical interactions in emotion: a meta-analysis of neuroimaging studies, Neuroimage 42 (2008) 998–1031 (PMC2752702).
[21] K.A. Lindquist, T.D. Wager, H. Kober, E. Bliss-Moreau, L.F. Barrett, The brain basis of emotion: a meta-analytic review, Behav. Brain Sci. 35 (2012) 121–143 (PMC4329228).
[22] J.S. Feinstein, R. Adolphs, D. Tranel, A tale of survival from the world of Patient S.M, in: D.G. Amaral, R. Adolphs (Eds.), Living Without an Amygdala, Guilford, New York, 2016, pp. 1–38.
[23] J.S. Feinstein, R. Adolphs, A. Damasio, D. Tranel, The human amygdala and the induction and experience of fear, Curr. Biol. 21 (2011) 1–5.
[24] G.M. Edelman, J.A. Gally, Degeneracy and complexity in biological systems, Proc. Natl. Acad. Sci. 98 (24) (2001) 13763–13768.
[25] B. Becker, Y. Mihov, D. Scheele, et al., Fear processing and social networking in the absence of a functional amygdala, Biol. Psychiatry 72 (1) (2012) 70–77.
[26] N.H. Kalin, Mechanisms underlying the early risk to develop anxiety and depression: a translational approach, Eur. Neuropsychopharmacol. 27 (2017) 543–553.
[27] E. Bliss-Moreau, M. Bauman, D.G. Amaral, Neonatal amygdala lesions result in globally blunted affect in adult Rhesus macaques, Behav. Neurosci. 125 (2011) 848–858.
[28] E. Bliss-Moreau, G. Moadab, D.G. Amaral, Lifetime consequences of early amygdala damage in Rhesus monkeys, in: D.G. Amaral, R. Adolphs (Eds.), Living Without an Amygdala, Guilford, New York, 2016, pp. 149–185.
[29] Bliss-Moreau, Lavenex, Amaral, (under review). Fear learning persists following early Amygdala damage in nonhuman primates. Manuscript under review.
[30] J.L. McGaugh, Consolidating memories, Annu. Rev. Psychol. 66 (2016) 1–24.
[31] S.A. Guillory, K.A. Bujarski, Exploring emotions using invasive methods: review of 60 years of human intracranial electrophysiology, Soc. Cogn. Affect Neurosci. (2014) nsu002.
[32] L.F. Barrett, K. Lindquist, E. Bliss-Moreau, S. Duncan, M. Gendron, J. Mize, L. Brennan, Of mice and men: natural kinds of emotion in the mammalian brain? Perspect. Psychol. Sci. 2 (2007) 297–312.
[33] L.F. Barrett, B. Mesquita, K.N. Ochsner, J.J. Gross, The experience of emotion, Annu. Rev. Psychol. 58 (2007) 373–403.
[34] L.F. Barrett, A.B. Satpute, Large-scale brain networks in affective and social neuroscience: towards an integrative functional architecture of the brain, Curr. Opin. Neurobiol. 23 (3) (2013) 361–372, http://dx.doi.org/10.1016/j.conb.2012.12.012.
[35] K.S. Kassam, A.R. Markey, V.L. Cherkassky, G. Loewenstein, M.A. Just, Identifying emotions on the basis of neural activation, PLoS One (2013).
[36] P.A. Kragel, K.S. LaBar, Multivariate neural biomarkers of emotional states are categorically distinct, Soc. Cogn. Affect. Neurosci. (2015) nsv032.
[37] P.A. Kragel, K.S. LaBar, Decoding the nature of emotion in the brain, Trends Cogn. Sci. 20 (6) (2016) 444–455.
[38] T.D. Wager, J. Kang, T. Johnson, T. Nichols, A.B. Satpute, L.F. Barrett, A bayesian model of category-specific emotional brain responses, PLoS Comput. Biol. (2015).
[39] R. Adolphs, How should neuroscience study emotions? By distinguishing emotion states, concepts and experiences, Soc. Cogn. Affect. Neurosci. (2017) nsw153.
[40] D.J. Anderson, R. Adolphs, A framework for studying emotion across species, Cell 157 (2014) 187–200.
[41] L.F. Barrett, Functionalism cannot save the classical view of emotion (short version), Soc. Cogn. Affect. Neurosci. (2017), http://dx.doi.org/10.1093/scan/nsw156 (There's also an extended version of the paper).
[42] D.L. Medin, A. Ortony, Psychological essentialism, Simil. Analogic. Reason. 179 (1989) 195.
[43] H. Saarimäki, A. Gotsopoulos, I.P. Jääskeläinen, J. Lampinen, P. Vuilleumier, R. Hari, M. Sams, L. Nummenmaa, Discrete neural signatures of basic emotions, Cereb. Cortex 26 (6) (2016) 2563–2573.
[44] E. Clark-Polner, T.D. Johnson, L.F. Barrett, Multivoxel pattern analysis does not provide evidence to support the existence of basic emotions, Cereb. Cortex (2016) bhw028.
[45] E. Clark-Polner, T.D. Wager, A.B. Satpute, L.F. Barrett, The brain basis of affect, emotion, and emotion regulation: current issues, in: M. Lewis, J. Haviland-Jones, L.F. Barrett (Eds.), Handbook of Emotions, 4th ed., The Guilford Press, New York, NY, 2016.
[46] C. Herry, J.P. Johansen, Encoding of fear learning and memory in distributed neuronal circuits, Nat. Neurosci. 17 (12) (2014) 1644–1654.
[47] P. Tovote, J.P. Fadok, A. Luthi, Neuronal circuits for fear and anxiety, Nat. Rev. Neurosci. 16 (2015) 317–331.
[48] R.C. Bolles, M.S. Fanselow, A perceptual-defensive-recuperative model of fear and pain, Behav. Brain Sci. 3 (2) (1980) 291–301.
[49] M.S. Fanselow, Neural organization of the defensive behavior system responsible for fear, Psychonomic Bull. Rev. 1 (4) (1994) 429–438.
[50] A.J. Shackman, A.S. Fox, Contributions of the central extended amygdala to fear and anxiety, J. Neurosci. 36 (2016) 8050–8063.
[51] W.A. Cunningham, K.A. Dunfield, P. Stillman, Emotional states from affective dynamics, Emot. Rev. 5 (2013) 344–355.
[52] Joseph E. LeDoux, Anxious: Using the Brain to Understand and Treat Fear and Anxiety, Penguin, New York, 2015.
[53] J.E. LeDoux, D.S. Pine, Using neuroscience to help understand fear and anxiety: a two-system framework, Am. J. Psychiatry 173 (2017) 1083–1093.
[54] J.E. LeDoux, R. Brown, A higher-order theory of emotional consciousness, Proc. Natl. Acad. Sci. (2017), http://dx.doi.org/10.1073/pnas.1619316114.
[55] K.A. Lindquist, Emotions emerge from more basic psychological ingredients: a modern psychological constructionist approach, Emot. Rev. 5 (2013) 356–368.
[56] A.B. Satpute, K.L. Lindquist (under review). At the neural intersection between language and emotion: Support for a constitutive view.
[57] L.F. Barrett, The future of psychology: connecting mind to brain, Perspect. Psychol. Sci. 4 (2009) 326–339, http://dx.doi.org/10.1111/j.1745-6924.2009.01134.x.
[58] L.F. Barrett, Emotions as natural kinds? Perspect. Psychol. Sci 1 (2006) 28–58.
[59] S. Deneve, C.K. Machens, Efficient codes and balanced networks, Nat. Neurosci. 19 (2016) 375–382.
[60] A.M. Sillito, The contribution of inhibitory mechanisms to the receptive field properties of neurones in the striate cortex of the rat, J. Physiology 250 (1975) 305–309.
[61] E. Bullmore, O. Sporns, The economy of brain network organization, Nat. Rev. Neurosci. 13 (5) (2012) 336–349.
[62] G. Tononi, O. Sporns, G.M. Edelman, Measures of degeneracy and redundancy in biological networks, Proc. Natl. Acad. Sci. 96 (6) (1999) 3257–3262.
[63] P. Sterling, S. Laughlin, Principles of Neural Design, MIT Press, 2015.
[64] P. Sterling, Allostasis: a model of predictive regulation, Physiol. Behav. 106 (1) (2012) 5–15.
[65] L.F. Barrett, W.K. Simmons, Interoceptive predictions in the brain, Nat. Rev. Neurosci. 16 (2015).
[66] I.R. Kleckner, J. Zhang, A. Touroutoglou, L. Chanes, Chengie Xia, W.K. Simmons, K.S. Quigley, B.C. Dickerson, L.F. Barrett, Evidence for a large-scale brain system supporting allostasis and interoception in humans, Nat. Hum. Behav. 1 (2017).
[67] G. Pezzulo, F. Rigoli, K. Friston, Active inference: homeostatic regulation and adaptive behavioural control, Prog. Neurobiol. 134 (2015) 17–35.
[68] A.K. Seth, Interoceptive inference, emotion, and the embodied self, Trends Cogn. Sci. 17 (11) (2013) 565–573.
[69] A.K. Seth, K. Suzuki, H.D. Critchley, An interoceptive predictive coding model of conscious presence, Front. Psychol. 2 (2012) 395.
[70] M.P. van den Heuvel, O. Sporns, An anatomical substrate for integration among functional networks in human cortex, J. Neurosci. 33 (2013) 14489–14500.
[71] L.F. Barrett, E. Bliss-Moreau, Affect as a psychological primitive, Adv. Exp. Soc. Psychol. 41 (2009) 167–218.
[72] J.A. Russell, L.F. Barrett, Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant, J. Pers. Soc. Psychol. 76 (5) (1999) 805.

[73] P. Kuppens, F. Tuerlinckx, J.A. Russell, L.F. Barrett, The relation between valence and arousal in subjective experience, Psychol. Bull. 139 (4) (2013) 917–940, http://dx.doi.org/10.1037/a0030811.
[74] A.R. Damasio, The somatic marker hypothesis and the possible functions of the prefrontal cortex, Philosop. Trans. R. Soc. Lond. Ser. B: Biol. Sci. 351 (1346) (1996) 1413–1420.
[75] G.M. Edelman, G. Tononi, A Universe of Consciousness: How Matter Becomes Imagination, Basic books, Chicago, IL, 2000.
[76] W. James, The Principles of Psychology, vol 1, Henry Holt and Company, New York, NY, 1890.
[77] W. James, The Principles of Psychology, vol 2, Henry Holt and Company, New York, NY, 1905.
[78] J.R. Searle, The Rediscovery of the Mind, MIT press, 1992.
[79] J.R. Searle, Mind: A Brief Introduction, Oxford University Press, 2004.
[80] C.D. Wilson-Mendenhall, L.F. Barrett, L.W. Barsalou, Variety in emotional life: within-category typicality of emotional experiences is associated with neural activity in large-scale brain networks, Soc. Cognit. Affect. Neurosci. 10 (1) (2015) 62–71, http://dx.doi.org/10.1093/scan/nsu037.
[81] M.P. van den Heuvel, O. Sporns, Rich-club organization of the human connectome, J. Neurosci. 31 (44) (2011) 15775–15786.
[82] D.J. Levinthal, The motor cortex communicates with the kidney, J. Neurosci. 32 (2012) 6726–6731.
[83] R.A. Poldrack, Can cognitive processes be inferred from neuroimaging data? Trends Cogn. Sci. 10 (2) (2006) 59–63.
[84] A.S. Fox, R.C. Lapate, R.J. Davidson, A.J. Shackman, Epilogue—The nature of emotion: a research agenda for the 21 st century, in: A.S. Fox, R.C. Lapate, A.J. Shackman, R.J. Davidson (Eds.), The Nature of Emotion: Fundamental Questions, 2nd ed., Oxford University Press, New York, 2017 (in press).
[85] H. Okon-Singer, D.M. Stout, M.D. Stockbridge, M. Gamer, A.S. Fox, A.J. Shackman, The interplay of emotion and cognition, in: A.S. Fox, R.C. Lapate, A.J. Shackman, R.J. Davidson (Eds.), The Nature of Emotion: Fundamental Questions, 2nd ed., Oxford University Press, New York, 2017 (in press).
[86] M.G. Mattar, M.W. Cole, S.L. Thompson-Schill, D.S. Bassett, A functional cartography of cognitive systems, PLoS Comput. Biol. 11 (2015) e1004533 (10.1371).
[87] R. Ciric, J.S. Nomi, L.Q. Uddin, A.B. Satpute, Contextual connectivity: a framework for understanding the intrinsic dynamic architecture of large-scale functional brain networks, Sci. Rep. (2017) (in press).
[88] E. Marder, A.L. Taylor, Multiple models to capture the variability in biological neurons and networks, Nat. Neurosci. 14 (2) (2011) 133–138.
[89] L.F. Barrett, Emotions are real, Emotion 12 (3) (2012) 413–429, http://dx.doi.org/10.1037/a0027555.
[90] R.L. Buckner, The serendipitous discovery of the brain's default network, Neuroimage 62 (2) (2012) 1137–1145.
[91] J.R. Binder, R.H. Desai, The neurobiology of semantic memory, Trends Cognit. Sci. 15 (11) (2011) 527–536, http://dx.doi.org/10.1016/j.tics.2011.10.001.
[92] A.B. Satpute, D. Badre, K.N. Ochsner, Distinct regions of prefrontal cortex are associated with the controlled retrieval and selection of social information, Cereb. Cortex 24 (2014) 1269–1277.
[93] B.T. Denny, H. Kober, T.D. Wager, K.N. Ochsner, A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex, J. Cogn. Neurosci. 24 (8) (2012) 1742–1752.
[94] A.B. Satpute, C.D. Wilson-Mendenhall, I.R. Kleckner, L.F. Barrett, Emotional experience, in: A.W. Toga (Ed.), Brain Mapping: An Encyclopedic Reference, Academic Press, Waltham, MA, USA, 2015.
[95] R.P. Spunt, A.B. Satpute, M.D. Lieberman, Identifying the what, why, and how of an observed action: an fmri study of mentalizing and mechanizing during action observation, J. Cogn. Neurosci. 23 (1) (2011) 63–74.
[96] R.P. Spunt, R. Adolphs, Folk explanations of behavior a specialized use of a domain-general mechanism, Psychol. Sci. (2015).
[97] A.B. Satpute, M.D. Lieberman, Integrating automatic and controlled processes into neurocognitive models of social cognition, Brain Res. 1079 (1) (2006) 86–97.
[98] D.M. Amodio, C.D. Frith, Meeting of minds: the medial frontal cortex and social cognition, Nat. Rev. Neurosci. 7 (4) (2006) 268–277.
[99] A.E. Skerry, R. Saxe, Neural representations of emotion are organized around abstract event features, Curr. Biol. 25 (15) (2015) 1945–1954.
[100] L. Fernandino, C.J. Humphries, L.L. Conant, M.S. Seidenberg, J.R. Binder, Heteromodal cortical areas encode sensory-motor features of word meaning, J. Neurosci. 36 (38) (2016) 9763–9769.
[101] M.V. Peelen, A.P. Atkinson, P. Vuilleumier, Supramodal representations of perceived emotions in the human brain, J. Neurosci. 30 (30) (2010) 10127–10134.
[102] C.D. Wilson-Mendenhall, L.F. Barrett, W.K. Simmons, L.W. Barsalou, Grounding emotion in situated conceptualization, Neuropsychologia 49 (5) (2011) 1105–1127.
[103] T. Lochmann, S. Deneve, Neural processing as causal inference? Curr. Opin. Neurobiol. 21 (5) (2011) 774–781.
[104] L. Chanes, L.F. Barrett, Redefining the role of limbic areas in cortical processing, Trends Cogn. Sci. 20 (2) (2016) 96–106.
[105] A. Clark, Whatever next? Predictive brains, situated agents, and the future of cognitive science, Behav. Brain Sci. 36 (03) (2013) 181–204.
[106] S. Denève, R. Jardri, Circular inference: mistaken belief: misplaced trust, Curr. Opin. Behav. Sci. 11 (2016) 40–48.
[107] K. Friston, The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. 11 (2) (2010) 127–138.
[108] J. Hohwy, The Predictive Mind, Oxford University Press, 2013.
[109] R.P. Rao, D.H. Ballard, Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects, Nat. Neurosci. 2 (1) (1999) 79–87.
[110] W.M. Wundt, Grundriss Der Psychologie, A. Kröner, Berlin, Germany, 1913.
[111] K.N. Ochsner, R.R. Ray, B. Hughes, K. McRae, J.C. Cooper, J. Weber, J.D. Gabrieli, J.J. Gross, Bottom-up and top-down processes in emotion generation: common and distinct neural mechanisms, Psychol. Sci. 20 (11) (2009) 1322–1331.
[112] A.B. Satpute, E.C. Nook, S. Narayanan, J. Weber, J. Shu, K.N. Ochsner, Emotions in black or white or shades of gray?: How we think about emotion shapes our perception and neural representation of emotion, Psychol. Sci. 27 (2016) 1428–1442.
[113] K. Lindquist, M. Gendron, A.B. Satpute, Language and emotion, in: M. Lewis, J. Haviland-Jones, L.F. Barrett (Eds.), Handbook of Emotions, 4th ed., The Guilford Press, New York, NY, 2016.
[114] K. Lindquist, A.B. Satpute, M. Gendron, Does language do more than communicate emotion? Curr. Direct. Psychol. Sci. 24 (2015) 99–108.
[115] S. Oosterwijk, K.A. Lindquist, K. Adebayo, L.F. Barrett, The neural representation of typical and atypical experiences of negative images: comparing fear: disgust and morbid fascination, Soc. Cogn. Affect. Neurosci. 11 (2016) 11–22.
[116] J. Whitacre, A. Bender, Degeneracy: a design principle for achieving robustness and evolvability, J. Theor. Biol. 263 (1) (2010) 143–153.
[118] L.F. Barrett, Solving the emotion paradox: categorization and the experience of emotion, Pers. Soc. Psychol. Rev. 10 (1) (2006) 20–46.
[120] L.F. Barrett, Psychological construction: the Darwinian approach to the science of emotion, Emot. Rev. 5 (2013) 379–389.
[121] L.F. Barrett, Construction as an integrative framework for the science of emotion, in: L.F. Barrett, J.A. Russell (Eds.), The Psychological Construction of Emotion, Guilford, New York, 2015, pp. 448–458.
[128] C.T. Gross, N.S. Canteras, The many paths to fear, Nat. Rev. Neurosci. 13 (9) (2012) 651–658.
[139] L.F. Barrett, Categories and their role in the science of emotion, Psychol. Inquiry 28 (2017) 20–26.
[140] E.H. Siegel, M.K. Sands, P. Condon Y., Chang J., Dy K.S., Quigley L.F. Barrett, (under review). Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories. Manuscript under review.
[141] L. Cosmides, J. Tooby, Evolutionary psychology and the emotions, in: M. Lewis, J.M. Haviland-Jones (Eds.), Handbook of Emotions, 2nd ed., Guilford, New York, 2000, pp. 91–115.
[142] D. Keleman, J. Rottman, R. Seston, Professional physical scientists display tenacious teleological tendencies: purpose-based reasoning as a cognitive default, J. Exp. Psychol. Gen. 142 (2013) 1074–1083.
[143] E. Mayr, What Makes Biology Unique? Considerations on the Autonomy of a Scientific Discipline, Cambridge University Press, New York, 2004.
[144] W. McDougall, An Outline of Psychology, Methuen, London, 1923.
[145] J.T. Cacioppo, G.G. Berntson, J.T. Larsen, K.M. Poehlmann, T.A. Ito, The psychophysiology of emotion, in: M. Lewis, J.M. Haviland-Jones (Eds.), Handbook of Emotions, 2, Guilford, New York, NY, 2000, pp. 173–191.
[146] C. Crivelli, M. Gendron, Facial expressions and emotions in indigenous societies, in: J.A. Russell, J.-M. Fernández-Dols (Eds.), The Science of Facial Expression, Oxford University Press, Oxford, UK, 2017, pp. 497–515.
[147] M. Gendron, D. Roberson, J.M. van der Vyver, L.F. Barrett, Perceptions of emotion from facial expressions are not universal: evidence from a remote culture, Emotion 14 (2014) 251–262 (PMC4752367).
[148] M. Gendron, D. Roberson, L.F. Barrett, Cultural variation in emotion perception is real: a response to Sauter et al, Psychol. Sci. 26 (2015) 357–359.
[149] K.S. Quigley, L.F. Barrett, Is there consistency and specificity of autonomic changes during emotional episodes?: Guidance from the Conceptual Act Theory and psychophysiology, Biol. Psychol. 98 (2014) 82–94.
[150] J.A. Russell, J.-A. Bachorwoski, J.-M. Fernandez-Dols, Facial and vocal expressions of emotion, Annu. Rev. Psychol. 54 (2003) 329–349.
[151] C. Izard, Basic emotions, natural kinds, emotion schemas: and a new paradigm, Perspectives in Psychological Science 2 (2007) 260–280.
[152] D. Keltner, D.T. Cordaro, Understanding multimodal emotional expressions: recent advances in basic emotion theory, in: J.A. Russell, J.-M. Fernández-Dols (Eds.), The Science of Facial Expression, Oxford University Press, Oxford, UK, 2017, pp. 57–75.
[153] J. Panksepp, Neurologizing the psychology of affects: how appraisal-based constructivism and basic emotion theory can coexist, Perspect. Psychol. Sci. 2 (2007) 281–296.
[154] M.I. Posner, S.W. Keele, On the genesis of abstract ideas, J. Exp. Psychol. 77 (3p1) (1968) 353.
[155] D.L. Schacter, D.R. Addis, The cognitive neuroscience of constructive memory: remembering the past and imagining the future, Philos. Trans. R. Soc. Lond. B Biol. Sci. 362 (1481) (2007) 773–786, http://dx.doi.org/10.1098/rstb.2007.2087.
[156] M. Gendron, L.F. Barrett, Reconstructing the past: a century of ideas about emotion in psychology, Emotion Rev. 1 (4) (2009) 316–339.
[157] C.L. Stephens, I.C. Christie, B.H. Friedman, Autonomic specificity of basic emotions: Evidence from pattern classification and cluster analysis, Biol. Psychol. 84 (3) (2010) 463–473.

[158] A. Vazdarjanova, J.L. McGaugh, Basolateral amygdala is involved in modulating consolidation of memory for classical fear conditioning, J. Neurosci. 19 (15) (1999) 6615–6622.

[159] S.M. Reynolds, K.C. Berridge, Emotional environments retune the valence of appetitive versus fearful functions in nucleus accumbens, Nat. Neurosci. 11 (4) (2008) 423.

[160] R.J. Blanchard, D.C. Blanchard, Attack and defense in rodents as ethoexperimental models for the study of emotion, Prog. Neuro-Psychopharmacol. Biol. Psychiatry 13 (1989) S3–S14.

[161] S.C. Motta, M. Goto, F.V. Gouveia, M.V. Baldo, N.S. Canteras, L.W. Swanson, Dissecting the brain's fear system reveals the hypothalamus is critical for responding in subordinate conspecific intruders, Proc. Natl. Acad. Sci. 106 (12) (2009) 4870–4875.

[162] A.B. Satpute, J. Shu, J. Weber, M. Roy, K.N. Ochsner, The functional neural architecture of self-reports of affective experience, Biol. Psychiatry 73 (2012) 631–638.

[163] J. Goense, H. Merkle, N.K. Logothetis, High-resolution fmri reveals laminar differences in neurovascular coupling between positive and negative bold responses, Neuron 76 (3) (2012) 629–639.

[164] L.F. Barrett, K.A. Lindquist, E. Bliss-Moreau, S. Duncan, M. Gendron, J. Mize, L. Brennan, Of mice and men: Natural kinds of emotions in the mammalian brain? A response to panksepp and izard, Perspect. Psychol. Sci. 2 (3) (2007) 297–311.

[165] K. Strimbu, J.A. Tavel, What are biomarkers? Current Opinion in HIV and AIDS 5 (6) (2010) 463.

[166] L.H. Arnal, A.-L. Giraud, Cortical oscillations and sensory predictions, Trends Cognit. Sci. 16 (7) (2012) 390–398.

[167] S.L. Bressler, C.G. Richter, Interareal oscillatory synchronization in top-down neocortical processing, Curr. Opin. Neurobiol. 31 (2015) 62–66.

[168] A. Brodski, G.-F. Paasch, S. Helbling, M. Wibral, The faces of predictive coding, J. Neurosci. 35 (24) (2015) 8997–9006.

[169] E. Kant, 1781. Critique de la raison pure, trad. A. Tremesaygues et B. Pacaud, Paris, puf, 4.

[170] U. Neisser, Cognition and reality: Principles and implications of cognitive psychology, WH Freeman/Times Books/Henry Holt & Co., 1976.

[171] R.L. Gregory, Perceptions as hypotheses, Philos. Trans. R. Soc. Lond. Ser. B, Biol. Sci. (1980) 181–197.

[172] R.L. Buckner, D.C. Carroll, Self-projection and the brain, Trends Cogn. Sci. 11 (2) (2007) 49–57.

Please cite this article in press as: L.F. Barrett, A.B. Satpute, Historical pitfalls and new directions in the neuroscience of emotion, Neurosci. Lett. (2017), http://dx.doi.org/10.1016/j.neulet.2017.07.045