Social Neuroscience

ISSN: 1747-0919 (Print) 1747-0927 (Online) Journal homepage: https://www.tandfonline.com/loi/psns20

Multiple scales of valence processing in the brain

Vincent Man & William A. Cunningham

To cite this article: Vincent Man & William A. Cunningham (2019): Multiple scales of valence processing in the brain, Social Neuroscience, DOI: 10.1080/17470919.2019.1692068

To link to this article: https://doi.org/10.1080/17470919.2019.1692068

Accepted author version posted online: 12 Nov 2019.
Published online: 21 Nov 2019.

SOCIAL NEUROSCIENCE
https://doi.org/10.1080/17470919.2019.1692068

Multiple scales of valence processing in the brain

Vincent Man [a,b] and William A. Cunningham [a]

[a]Department of Psychology, University of Toronto, Toronto, ON, USA; [b]Divisions of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, USA

ABSTRACT

Psychological theories posit that affective experiences can be decomposed into component constituents, yet disagree on the level of representation of these components. Affective experiences have been previously described as emerging from core dimensions of valence and arousal. However, this view needs to be reconciled with accounts of valence processing in appetitive and aversive circuits from the neuroscience literature. Here we offer an account of affect that allows for both perspectives but compares across levels of analysis. At one level of analysis, valence and arousal are observed already in the properties of encountered stimuli and the appetitive and aversive neural circuits that engage accordingly. At another level of analysis, the explicit experiential aspect of affective processes are compressed and appraised in a manner that allows these experiences to be organized along valence and arousal axes. We review both the behavioral neuroscience evidence on appetitive and aversive circuits as well as the cognitive neuroscience literature on compression in information coding across multiple domains of processing. We argue that these processes are domain-general and adapt these principles to provide a perspective on how valence can be represented at multiple scales in the brain.

ARTICLE HISTORY
Received 22 March 2019
Revised 28 October 2019
Published online 21 November 2019

KEYWORDS
Affect; valence; compression; appetitive; aversive

Introduction

Conceptualizing a framework to understand human affect has been challenging due in part to disagreements between both the nature of constituent components underlying affective phenomena as well as their causal versus consequential organization. Examples of disagreements in affective theory are myriad. Psychological descriptions of affective processes have emphasized the importance of studying underlying physiological and mental components (Cacioppo & Berntson, 1994; Russell, 1980), in contrast to perspectives that organize emotional states into distinct categories with unique neurobiological underpinnings (Ekman, 1992). Even within the domain of psychological approaches that emphasize interactions between domain-general affective components, there remains disagreement about what constitutes the fundamental axes along which affective processes should be organized. A classic debate pits proposals of a circumplex space in which valence and arousal constitute two orthogonal affective dimensions that can give rise to a multitude of affective phenomena under different configurations (Barrett & Russell, 1998; Russell, 1980), versus those arguing that positive and negative processes are better construed as separate dimensions which can vary in arousal and correspond to distinct motivational drives, such as the Evaluative Space model (ESM) and others (Cacioppo & Berntson, 1994; Gray, 1990; Lang, Bradley, & Cuthbert, 1998). One complication in resolving this debate is that support for one perspective over another can apparently shift according to the domain of the evidence surveyed. Focus on the effable experiential aspect of affect (i.e., feelings) seems to support the view that orthogonal axes of valence and arousal comprise a space onto which many described emotional labels can be projected (Russell, 1980) whereas focus on biological evidence tends to support the view that affective processes map onto distinct neural circuitry (e.g., related to stimulus-response associations) across animals and humans (LeDoux, 2012).

In this current perspective, we argue that we can indeed turn to both the neuroscience and psychology literatures to potentially reconcile this ambiguity. Here we are inspired by a neuroscientific perspective propagated by John Cacioppo and others, which places emphasis on the hierarchical nature of brain organization (Cacioppo & Berntson, 1992) and accordingly, the idea that mental processes can exist at multiple levels of analysis (i.e., different spatial and temporal scales, but also different levels of abstraction). Further, we are inspired by the general scientific approach of reconciliation, rather than further division, between existing debates on a topic, by drawing from multiple domains of study

CONTACT Vincent Man v.man@mail.utoronto.ca Department of Psychology, University of Toronto, Toronto, ON M5S 3G3, USA

(Cacioppo & Berntson, 1999). Thus, we adopt this attitude of trying to reconcile between competing accounts of affect, and in the current perspective we aim to do so by trying to understand one of the fundamental questions underlying any theory of affect: what role does valence (i.e., positivity vs negativity) play as a building block of affective processes, and how is it organized in the brain? We think about this question again using a recurrent insight across the empirical and theoretical work left by Cacioppo: that psychological processes and the neural systems that support these processes should be understood across multiple scales. Our current goal is thus not to provide an exhaustive new model of affect; rather it is to expand upon and add a contribution to an important issue throughout the literature.

Our main thesis is that the question of how valence serves as a fundamental building block behind affective processes differs as a function of the level of analysis. Here we focus on two levels of analyzes. The first concerns the neural circuits (e.g., circuits that distinguish aversive processing from appetitive) that engage in response to stimuli in the world that may already carry valenced properties (e.g., a noxious scent) and give rise to affectively charged behaviors (e.g., avoidance). We review the brain circuits that are plausibly involved in processing valence at this scale, drawing from the behavioral neuroscience literatures (Anderson & Adolphs, 2014; Dickinson & Dearing, 1979; LeDoux, 2012). Doing so emphasizes the heterogeneous nature of the construct of valence at this level of analysis, mirroring well described heterogeneity in neurobiological underpinnings of arousal systems (McNaughton & Gray, 2000). The second concerns the manner in which valence (along with arousal) has been construed as a fundamental building block behind explicit emotional experiences (Russell, 1980).

This level of analysis is clearly distinguished from the former in that the evidence largely comes from the human psychological literature rather than animal neuroscience, with many treatments of affect grounded in neurbiology eschewing this domain of affect (Anderson & Adolphs, 2014; LeDoux, 2012). Here we offer the alternative possibility that at this level of analysis, a valence (and arousal) construct can be considered as a consequence of felt affect, rather than a cause. To support this claim, we draw from the cognitive neuroscience literature in which the compression of information into sparse, efficient codes accounts for multiple psychological phenomena in vision, memory, and planning. We argue that this suggests the possibility that compression serves as a domain-general computational mechanism that is applied to complex internal affective states. The result of compressing these rich internal states is that feelings can be efficiently labeled and broadcast onto a common space organized by axes of valence and arousal. Finally, we consider the open questions of what purpose this neural mechanism serves and what comprises the rich internal states that are compressed.

Appetitive and aversive brain circuits

We begin with a review of our first level of analysis: the neural circuits that support the processing of valenced input stimuli. Consistent with previous proposals (Gray, 1990; Lang et al., 1998; Watson & Tellegen, 1985), and following in the work of John Cacioppo (Cacioppo & Gardner, 1999), we underline the evidence supporting the independent processing of positive and negative incoming information along appetitive and aversive circuits, respectively. Furthermore, we expand on the possibility that despite the separability of processing between positive and negative information, there is some evidence that these processes occur in parallel along adjacent neural substrates. We focus in particular on the circuitry of the amygdala and mesolimbic dopaminergic pathways to highlight examples of separable processing of positive and negative information.

We start with the uncontroversial premise that encountered stimuli are perceived along multiple features which may include valence (positivity and negativity) and intensity. Whether these features are necessarily inherent to the external stimulus or whether a feature of an encountered stimulus arising from top-down perceptual processes (Frith & Dolan, 1997) is beyond the current scope; our focus is that inputs from sensory systems can already shape affective processing streams, such as appetitive or aversive circuits.

In affective circuits that involve the amygdala, cortical association regions that integrate information from primary sensory cortices project to the thalamus, which route information to the lateral amygdala (LA; see LeDoux (2000) for a review). From here, the amygdala has played a central role in aversive processes. In particular, the central nucleus of the amygdala has been found to be a critical structure in the neural circuit underlying an automatic system of threat detection (Davis & Whalen, 2001), as well as fear conditioning and its associated behavioral expressions (Anderson, Spencer, Fulbright, & Phelps, 2000; LaBar, Gatenby, Gore, LeDoux, & Phelps, 1998; LeDoux, 1998, 2000). In humans, early work on amygdala function emphasized its greater responses to negative affective states across stimulus modalities, such as faces (Anderson et al., 2000) and odors (Zald & Pardo, 1997), and at multiple stages of

processing from perception (Adolphs, Russell, & Tranel, 1999; Calder, Keane, Manes, Antoun, & Young, 2000) to evaluation (Adolphs & Tranel, 2004).

A critical mechanism underlying aversive processes, fear conditioning involves the binding and transmission of information about a conditioned stimulus (CS; e.g., a tone) and an unconditioned stimulus (US; e.g., a shock) within the amygdala, and the elicitation of fear reactions through output projections to downstream brainstem areas that regulate associated behavioral, autonomic and endocrine systems (LeDoux, 2000). As described above, stimulus properties across modalities are transmitted to the lateral nucleus of the amygdala through a quick thalamic pathway (LeDoux, Farb, & Ruggiero, 1990) as well as a cortical pathway that accumulates information more slowly (Quirk, Armony, & LeDoux, 1997). At the lateral nucleus, information from the CS and US are integrated at a cellular level (LeDoux, 1998). Moreover, the context in which a conditioned cue is embedded plays a central role in learning processes; as such, projections form the CA1 and subiculum of the hippocampus are essential for contextual conditioning (Canteras & Swanson, 1992; Maren & Fanselow, 1995). This is mediated by hippocampal projections to the amygdala, which regulates the flexible switching between valence-specific neurons in the basolateral nucleus (Redondo et al., 2014). Output signals from the central nucleus of the amygdala form an important pathway that supports the expression of affective responses to threat as well as associated physiological outputs. Projection sites from the central nucleus include the periaqueductal gray (PAG) and the lateral and paraventricular nuclei of the hypothalamus (Rizvi, Ennis, Behbehani, & Shipley, 1991), which regulate defensive fight or flight behavior (LeDoux & Daw, 2018; LeDoux, Iwata, Cicchetti, & Reis, 1988), as well as cholinergic systems of the forebrain (Everitt & Robbins, 1997; Holland & Gallagher, 1999; LeDoux, 2000). One underlying mechanism through which the amygdala processes fear is by directing attention toward salient information, supported by CeA-mediated lowering of neuronal thresholds in the basal forebrain via acetylcholine modulation (Bucci, Holland, & Gallagher, 1998).

Conditioning paradigms (both instrumental and Pavlovian) similarly form the bedrock of the literature of appetitive processes. We focus again on amygdala-dependent circuits in these processes, in part to highlight the how appetitive and aversive processes can operate in parallel using apparently similar neighboring circuits. Nevertheless, amygdala-dependent circuits that support reward processing and corresponding approach behavior can be distinguished from aversive pathways described above.[1] Here, the focus within the amgydala shifts to the basolateral complex (BLA) rather than the lateral and central nuclei (Baxter & Murray, 2002; Davis & Whalen, 2001). Again, the amygdala plays a role as a hub in these processes. Specifically, it has been shown that the BLA plays a critical function in representing current stimulus-outcome associations, and encoding representations of reinforcer value in the appetitive domain. That the current value representation depends on the BLA has been shown in paradigms in which the reward value associated with a stimulus changes over time (e.g., reinforcer devaluation; Adams & Dickinson, 1981). Critically, devaluation sensitivity is affected by neurotoxic lesions to the BLA, but not to the central amygdala (CeA; Hatfield, Han, Conley, Gallagher, & Holland, 1996). Selective satiation paradigms that similarly manipulate the value of a reinforcer are affected by lesions to the BLA, with evidence showing that monkeys do not avoid choosing objects associated with satiated outcomes (Málková, Gaffan, & Murray, 1997).

Its role in updating the current value associated with a stimulus means that the BLA is important for flexibility in goal-directed behavior, which will become more apparent when considering the connectivity between the amygdala, nucleus accumbens (NAcc), and orbitofrontal cortex (OFC). One example of a functional role of this circuit is apparent in second-order conditioning, in which the current value of the initial CS (first-order CS) is able to invoke conditioned responses in another neutral stimulus (second-order CS); this effect disappears with ablations to the BLA, but not the CeA (Hatfield et al., 1996). Critically, this process depends on interactions with striatal regions such as the NAcc (Cador, Robbins, & Everitt, 1989; Everitt, Cador, & Robbins, 1989; Setlow, Holland, & Gallagher, 2002). Furthermore, an important distinction has been made between the predictive and affective properties of a secondary reinforcer (Baxter & Murray, 2002; Parkinson et al., 2001), which importantly contributes to our understanding of neural implementations of learned representations. After BLA ablations, extra-amygdalar cortical regions are able to support the continued predictive properties of second-order CSs, in that they continue to guide behavior, despite losing their reinforcing properties (Parkinson et al., 2001). As such, the reciprocal connectivity between the BLA and NAcc to the OFC is critical for guiding choice behavior and response selection.

We further consider how the neural heterogeneity within a region accounts for functional diversity and flexibility. We review evidence of neural coding that allows for

[1]Circuit level distinctions are complicated by complex aversive processes beyond reflexive behavioral outputs. See LeDoux and Daw (2018) for a review of goal-directed and deliberative defensive processes.

the representation of both positive and negative information simultaneously, and argue that parallel processes contribute to apparent heterogeneity of function within certain brain regions (see also Berridge, 2019 for a review). For example, parcellation of amygdala sub-regions reveal that modularized functional descriptions begin to break down (Bickart, Dickerson, & Barrett, 2014), especially when considering intra-regional connectivity within the amygdala. Despite individually being implicated in the literature on appetitive and aversive processes respectively, intra-regional circuits allow interactions between the LA and CeA, in which incoming multimodal sensory and associated affective information are bound (LeDoux, 2015). A more nuanced story of intra-amygdala function considers the modulatory role of the intercalacted layer (ITC), an important layer of GABAergic inhibitory neurons situated between the CeM and BLA (Amano, Unal, & Paré, 2010; Zikopoulos, John, García-Cabezas, Bunce, & Barbas, 2016). Considering the excitatory and inhibitory neuronal dynamics via the ITC can account for the many flexible functional processes often attributed to the amygdala. For example, excitatory neurons from the LA synapse onto inhibitory neurons in the ITC and subsequent projections from the ITC terminate at mostly inhibitory neurons in the CeA; thus activation of LA neurons ultimately leads to increased output from the CeA due to disinhibition, allowing for the expression of behavioral and physiological responses via brainstem and hypothalamic targets (see LeDoux, 2007 for a review).

Critically, conditioning paradigms that reverse the cues associated with rewarding and aversive outcomes demonstrate that these selective neurons are not encoding the sensory properties of the cue but rather the outcomes with which the cues are associated, indicating that these neurons are specific to valence features (Paton, Belova, Morrison, & Salzman, 2006; Schoenbaum, Chiba, & Gallagher, 1999). These two sub-populations of valence-specific neurons are spatially interspersed in the BLA (Shabel & Janak, 2009), where neurons sensitive to the same valence show greater correlation in activation than neurons sensitive to the opposing valence (Zhang et al., 2013). One study tested between two different models of processing in the amygdala (Shabel & Janak, 2009). The "different circuits" model predicted an inverse relationship between activity during appetitive and aversive arousal and is analogous to the bipolar perspective on valence organization in the psychological literature. On the contrary, in the "same circuits" model, a generalized circuit was not specific to appetitive or aversive stimuli and predicted a positive relationship between activity during appetitive and aversive arousal. Critically, despite a design biased toward the "different circuits" model, the study found that a subset of amygdala neurons held quantitatively similar changes in activation to stimuli paired with appetitive and aversive reinforcement independently, supporting the "same circuits" model. These changes in neural activity accompany stimuli paired with reinforcement, suggesting that the neurons are tracking the motivational significance of incoming information, rather than particular sensory features.

Similar principles of parallel appetitive and aversive processes have been proposed in the NAcc. Functional subdivisions within the NAcc are involved in separate circuits and support different aspects of motivated behavior (Berridge, 2019; Berridge & Robinson, 2003). For example, the NAcc has been linked to mesolimbic dopaminergic circuitry that supports appetitive learning (Berridge & Robinson, 1998; Ito & Hayen, 2011), and these processes can be distinguished from 'liking' aspects of reward that are mediated through opioid systems, particularly in the rostrodorsal aspect of the NAcc shell (Berridge, Robinson, & Aldridge, 2009). Recently, rostrocaudal functional gradients have been delineated along the NAcc shell, with appetitive processes (liking and wanting) represented in the rostral shell and aversive processes (fear and disgust) in the caudal shell (see Berridge (2019) for a review). Nevertheless, divisions between appetitive and aversive neuronal populations in the NAcc are similarly complicated by evidence that the degree to which populations along the rostrocaudal gradient in the NAcc uniquely support appetitive and aversive processes are affected by modulatory influences, both via neurotransmitters (e.g., glutamate-sensitive AMPA receptors Richard and Berridge (2011)) and environmental contingencies (Reynolds & Berridge, 2008).

Together, these findings support the notion that appetitive and aversive processes are not necessary opponent, but rather can operate in parallel and share similar neural substrates. The biological evidence suggests that positive and negative processes have distinct substrates at the neuronal and circuit level, and that pathways supporting these processes are able to operate in parallel, are amenable to modulatory influences, and output to distinct downstream processes such as value-based learning and defensive behavior. It should be apparent that this evidence is therefore highly consistent with a multivariate space of positivity and negativey, well described in the ESM (Cacioppo & Berntson, 1994), which allows for independent activation schemes as well as the coactivation of positive and negative dimensions of affect.

Of course, the Evaluative Space Model also allowed for a reciprocal relationship between positive and negative

processes, and is thus not incompatible with bipolar conceptualizations of valence (Russell, 1980). Instead, it argued that a bipolar conceptuatlisation was not strictly necessary to account for multiple affective phenomena, and that the precise relationship between positivity and negativity was contingent upon the conditions under which this relationship is examined. In this current review, our argument is that the level of analysis, and accordingly the literature respective to that level of analysis, emphasizes either separable positive and negative processing or the organization of affective experiences along a continuum from positivity to negativity. Above and previously (Man, Nohlen, Melo, & Cunningham, 2017) we argued that by zooming in on the micro/meso-level of analysis, the neuroscience evidence largely supports separable and coactive positive and negative processing. Nevertheless, at the macro level of analysis, the psychological data demonstrates the utility of a single dimension of valence that comprises, along with an orthogonal axis of arousal, a space from which affective experience are emergent. Critically, we propose a mechanism by which these affective experience emerge and are organized.

Compression and valence

One of the principal lingering questions left by a narrow focus on appetitive and aversive processes is the disconnect with psychological descriptions of affective states (LeDoux, 2012). Psychological theories have focused on expressed evidence of varying emotional phenomena such as self-reported affective labels (Russell, 1980). In doing so they have also placed emphasis on the role of valence in affective processes, albeit in a different manner in which valence constitutes a 'core' component underlying affect (Barrett, 2006). This has led to debate on the organization of valence and the manner in which it contributes to more complex affective processes. We argue that apparent disagreements amongst affective theorists on this issue can potentially be reconciled by recognizing that the evidence respective to neural affective circuits, and the perspectives stemming from psychological descriptions of emotions, are operating at different levels of analysis. Here we turn to a coarser and more abstract level of analysis, that of complex, felt emotional experiences, and posit a novel perspective on how valence might be implicated.

Our main proposition is that at this higher-order level of analysis, multifaceted interactions (e.g., recurring, Barrett, 2017; Cunningham, Zelazo, Packer, and Van Bavel, 2007; Scherer, 1984) between myriad processes contribute to some representation of complex, felt affective experiences in the brain. However, these feelings as we describe them are not merely an emergent construct (Barrett, 2009) from complicated internal dynamics: there is a critical stage of processing in which these complex representations are compressed into more efficient codes, and that compression step allows our internal emotional experiences to be communicable. We suggest that one result of compressing ineffable internal experiences into something that can be expressed in body and language is the broadcasting of resulting codes (i.e., expressed feelings) onto a space that is well described by valence and arousal axes (Barrett & Russell, 1998; Russell, 1980). In this way valence is organized at this higher-order level of analysis: as a structural consequence of neural computations that allow us to usefully categorize and place boundaries on our rich internal states.

What evidence is there to suggest that this kind of compression process is taking place? To substantiate our proposal, we move beyond the descriptive level above and propose mechanistic and biological descriptions of how a compression process might occur, and how it might apply to the domain of affect. To describe what we mean by compression at a mechanistic level, we turn to the information theory literature below and discuss how dense information can be efficiently coded (Barlow, 1961) into sparser representations, thereby 'compressing' the information and reducing representational demands. To offer the possibility that the brain might be doing this kind of computation, we then review the evidence that this kind of information transformation, into relatively more compact representations, occurs across the cognitive neuroscience literature in domains such as visual processing, memory, and decision-making. Finally, we offer some predictions that this account might make for the domain of affect and bring in existing literature that seemingly supports these predictions. We end our discussion with open questions of why this compression process might occur, and how meta-cognitive processes might contribute to this process.

One well-described computation related to the concept of compression is *efficient coding*, in which multidimensional inputs are transformed into an "efficient" representation that maximizes the amount of information retained while minimizing resource demands (such as representational capacity; Attneave, 1954; Barlow, 1961). An intuitive way of understanding the efficient coding principle is by considering the transformation of raw input stimuli S into outputs O: $O \propto f(S)$, such that O compactly retains as much of the original information initially conferred by the stimuli. The goal of efficient coding is to specify the function f in such a manner as to maximize the information shared between the stimuli and the output: $I(O; S)$ (Zhaoping, 2014). Given this objective, what might serve as candidate transformations (i.e., how is f specified)? One manner in

which raw input stimuli can be efficiently represented is by transforming them into a set of outputs which each carry independent aspects of information. In other words, if there is overlapping information shared between multiple input stimuli (i.e., $I(S1; S2) \neq 0$), an efficient function would be to transform the inputs into outputs that are maximally decorrelated, so that each output can capture as much independent variance as possible. Doing so would minimize the amount of redundant information between input stimuli in the encoding process. A useful analogy is a latent variable decomposition of data such as the (infomax) independent components analysis (Bell & Sejnowski, 1995), a common signal processing step often used in fMRI and EEG research.

Alternatively, imagine a case in which the set of input stimuli is noisy (low signal-to-noise ratio [SNR]). In this case, much of the information conferred by *S* is not useful information to retain. Specifying *O* in such a way as to maximally capture independent sources of variance in *S* would no longer be the most efficient way of compactly representing the signal in *S*, since much of the variance in *S* is driven by noise. In these low SNR cases, an efficient encoding transformation of the noisy *S* would instead to be smooth or average over the inputs, and represent the smoothed (e.g., mean) signal in *O* (Zhaoping, 2014). In this way, the resulting code would still be more compact, and would capture the *relevant* variance across *S* (i.e., its signal). It is then apparent that despite different encoding strategies, which may be contingent on the nature of the input stimuli (e.g., its SNR), in both cases this principle reflects a form of information compression in that a larger set of raw input stimuli are transformed into a smaller set of output codes. As such, this principle serves as a useful, albeit not exhaustive, computational description of the compression process here.

Another utility of using the efficient coding principle to demonstrate compression is that its been successfully employed to describe neural mechanisms underlying psychological phenomena across multiple cognitive domains. An early example includes the well characterized receptive fields of the primary visual cortex. Natural images are good examples to demonstrate an efficient coding transformation in that there is high spatial correlation between adjacent 'pixels' of the image: adjacent places on an image are more likely to be depicting the same object, similar color and luminance, etc. Previous work has shown that already the distribution of cones in the retina are arranged to give maximum information about color in natural scenes and do so following the efficient coding principle of reducing redundancy in the scene (Lewis & Zhaoping, 2005). Furthermore, an encoding scheme which maximizes the sparseness of its outputs results in modeled receptive fields that carry properties similar to biological receptive fields in primary visual cortex, such as spatial localization and orientation (Olshausen & Field, 1996). Finally, it has been shown that receptive fields in human primary visual cortex reflect preplay activity in anticipation of spatially distributed sequences of events; this preplay activity is elicited by partial sequence cues and shows compression in the temporal domain (Ekman, Kok, & Lange, 2017).

Compression in the temporal domain has also been shown to be a feature of memory consolidation. Neurophysiological evidence has since shown that consolidation processes involve the replay of experienced events during sleep (Hoffman, 2002) and is accompanied by corresponding neural activity that preserves the temporal order of the experienced event (Louie & Wilson, 2001). Critically, replay of the encoded sequence during sleep in neocortex (medial prefrontal cortex) is temporally compressed (Euston, Tatsuno, & McNaughton, 2007). This neural evidence has been corroborated with psychological evidence demonstrating temporal compression during the retrieval of spatial memory under navigational demands (Arnold, Iaria, & Ekstrom, 2016; Bonasia, Blommesteyn, & Moscovitch, 2016; Jafarpour & Spiers, 2017).

A central tenant of memory consolidation theory is that as episodic memories are encoded in long term storage, there is a transfer of the memory trace from the hippocampus to the neocortex (Scoville & Milner, 1957) alongside accompanying trace transformation (Moscovitch & Nadel, 1998; Winocur & Moscovitch, 2011). An emerging body of work has detailed how the consolidation process transforms the initial encoding of an experience into a compact form by stripping away perceptual details of the experience and retaining relevance-based sparse information (Winocur & Moscovitch, 2011). This is accompanied by changes in the neural substrates underlying different aspects of the memory trace whereby gist information is encoded in the anterior hippocampus and schemas, generalized knowledge structures extracted from co-occuring experiences, are represented in the ventromedial prefrontal cortex (vmPFC; Gilboa and Marlatte, 2017; Robin and Moscovitch, 2017). One critical function of the representation of abstracted knowledge structures is that they mediate the fast acquisition of new information such as in few-shot learning (e.g., Tse et al., 2007). In these cases, new information can be compared against and quickly assimilated into existing schemas. This consolidation transformation holds similarities with efficient coding in the visual system by describing a function through which information is transformed into sparser representations. Perhaps one way of regarding schemas and gist information is as 'codes' that capture summarized versions of the initial experienced events; the

rich perceptual details in the initial episodic memory can often be redundant and this redundancy is reduced by retaining the relevant knowledge structure.

The psychological process of extracting abstracted structure from highly correlated information has also been demonstrated as an important component of decision-making and planning in complex, multidimensional environments. In the real world and in naturalistic tasks, decisions are often hierarchical (e.g., progress toward an overarching goal can be accomplished by finishing relevant subgoals Botvinick, 2008), and shared structure exists in the problem set faced by the decision maker (Botvinick, Weinstein, Solway, & Barto, 2015). For example, naturalistic tasks are structured in that subroutines of a task can be divided by boundaries (Botvinick, Niv, & Barto, 2009). Furthermore, neural substrates underlying the segmentation and representation of task events have been delineated along the lateral prefrontal cortex (Zacks, Speer, Swallow, Braver, & Reynolds, 2007march), with hierarchical brain organization from sensory to association cortices such as the angular gyrus corresponding to structured information along short and long timescales, respectively (Baldassano et al., 2017). A critical direct test of the implication of efficient coding mechanisms in a hierarchical decision task involved the navigation of a spatial map with bottlenecked transition states (e.g., doorways between rooms; Solway et al., 2014). The authors found that the optimal navigational route through this kind of structured problem demanded a hierarchical representation of action subsequences to find the bottleneck at the end of the room rather than a search through all adjacent spaces within a room. Following the efficient coding hypothesis, this optimal representation yielded a maximally efficient code by compressing redundant actions (within-room traversals) into a sparser set of relevant actions (across-room traversals).

Inspired by this pattern of information compression across multiple domains of cognition, we now turn back to understanding affective experiences. Specifically, we consider how affective processes at a more fine-grained level of analysis, such as the aforementioned interacting appetitive and aversive neural circuits, might translate to felt, expressible feelings at a more abstract resolution. We propose that a similar compression process occurs in the case of overt emotions: the brain pushes a raw input through a compression filter to result in a sparser, less redundant, representation. The resulting sparse representation comprises our emotional expressions which, relative to the dynamic and complex affective processes from which they arise, are bounded, labeled, and associated with distinct physical and verbal expressions. In this way emotional experiences remain 'emergent' (Barrett, 2009) from lower-order processes; what we offer here is a candidate emergence mechanism. We suggest also that the nature of the compression process results in codes that can be easily projected onto a space structured by dimensions of valence and arousal. As such, at the fine-grained level of analysis valence constitutes a key constituent feature which weighs the relative engagement of appetitive versus aversive neural circuits, well supported by the behavioral neuroscience literature. At a broader level of analysis, valence is a descriptive dimension that results from the compression of interacting affective processes into expressed affective experiences, and is consistent with early psychological data that overt emotional tags could be broadcast and organized around a circumplex comprising valence and arousal axes (Russell, 1980).

We argue that internal affective processes are transformed into expressed feelings through an efficient compression function. If it is the case that this transformation follows the principles of efficient coding, we would expect several downstream predictions. First, it should be the case that the resulting sparse codes capture a large part of the variance of underlying affective processes but do so by maximally reducing redundant information. This means that resulting emotional feelings might capture most of the types of interactions between affective processes but fail to retain high levels of detail, analogous to the transformation of an initially encoded experience in the hippocampus to schematized and abstracted knowledge structures in the neocortex. As such, the class of emotional phenomena described as ineffable (e.g., nonconscious, nonreflective; Frijda, 2009) may constitute the 'residual' of the compression processes: dynamics and variance in affective processes that are not encoded into sparser codes, and so are not easily described, categorized, and expressed. Relatedly, it is interesting to note that the dictionary of emotional words (as well as the physical manifestations of emotional expressions) differs across languages and cultures (Lomas, 2018). One possibility is that the particular encodings deemed 'efficient' are modulated by the cultural context in which an individual is embedded. The nature of the efficient sparse codings that make up emotional language can be both generalizable and context-specific, and in contexts where certain kinds of affective expressions are more relevant (e.g., for communication), different sets of sparse codes emerge from the affective inputs.

If a compression process occurs in the affective domain, one open question is concerned with the nature of the input signal: what is being compressed? Recent advances have postulated the critical role of internal affective states (Anderson & Adolphs, 2014; Barrett, 2017; Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012), with various accounts across perspectives of how to best characterize

these internal states. One interesting recent proposal (Barrett, 2017) draws from predictive coding theories of neural function (Clark, 2012; Friston, 2005; Rao & Ballard, 1999). In this account, internal representations provide predictive signals via interoceptive processes. Predictive signals from internal representations originate in deep layers of agranular cortex, interacting with relatively more laminated cortical regions to compute prediction errors to compute prediction errors (see Barbas, 2015; Barrett and Simmons, 2015; Chanes and Barrett, 2016, for more in-depth reviews on these mechanisms). One potential example of this predictive framework in affective processing can be seen in reciprocal connections between the amygdala and infralimbic cortex, with recent evidence showing that that the pathways to and within the amygdala can support this flexibility. For example, long-range excitatory neurons from the infralimbic region synapse onto inhibitory interneurons in the BLA as well as the GABAergic inhibitory neurons of the ITC (Cho, Deisseroth, & Bolshakov, 2013; Janak & Tye, 2015; LeDoux, 2007), providing a potential structural mechanism underlying the top-down biasing of amygdala response to incoming information from prefrontal regions. One such top-down signal might come from expectations, which modulates responses to positive and negative-encoding amygdala neurons (Belova, Paton, Morrison, & Salzman, 2007). Within local neuronal populations, computational approaches have described how attractor landscapes might capture dynamic mental states, including 'affective' states of varying valence, via interactions between persistently firing neurons (Salzman & Fusi, 2010). This sort of landscape could also capture the manner in which internal affective states shift under internal and external contextual perturbations (Rigotti, Ben-Dayan Rubin, Wang, & Fusi, 2010).

At a more macro level of neural analysis, there is some evidence for a set of core, general regions that allow for flexibility in processing valenced information (Lindquist et al., 2012; Liu, Hairston, Schrier, & Fan, 2011). According to this view, neural substrates underlying a general network can be functionally selective to positive and negative processes, dynamically fluctuating according to top-down influences. Contrary to specific networks for discrete emotional states, there is much spatial overlap across brain correlates of different types of affective processes (Touroutoglou, Hollenbeck, Dickerson, & Barrett, 2012). At both the meso-scale of reciprocal interactions between top-down and bottom-up affective processes, and the macro-scale of distributed brain networks that support a generalizable set of core processes, these affective dynamics serve as potential candidate inputs into the compression process.

Indeed, the features of neural substrates previous found to be involved in affective processes, that of being flexible and generalized across multiple types of emotional experiences, are consistent with predictions made by a compression hypothesis that describes a transformation of multidimensional input signal into output codes that optimizes for minimal redundancy, when the SNR is high. Such cases might include straightforward univalent processes, and a distributed brain network that recruits processes across cognitive domains can be compressed into a more defined, unambiguous expression (e.g., "I'm feeling good"). However, the hypothesis suggests that when the SNR is low the inputs are not transformed to reduce redundancy, but smoothed over to boost the signal. This might be in the case of mixed emotions, when the internal affective state comprises a mix of positive and negative processes (Man et al., 2017). Here the problem of compression is to delineate an explicit compact representation of these complex interactions by mixing over the input bivalent processes (e.g., perhaps by taking a context relevance-weighted average). The resulting compacted code should reflect this averaging, with corresponding overt expressions (e.g., "bittersweet").

This compression hypothesis is not without its own open questions. What processes might organize the compressed outputs into a space defined by valence and arousal? We raise the possibility that explicit evaluations of felt emotions play an important role in our ability to carve out and make sense of our experiences. This perspective is influenced by psychological work on second-order cognition (i.e., "thoughts about thoughts/feelings") which involve assessments of thoughts and feelings as good or bad, appropriate, desirable, etc., as well as appraisal of the uncertainty of our cognitive processes (Petty, Briñol, Tormala, & Wegener, 2007). One possibility is that second-order appraisals of emergent, felt emotions might organize the experience into summary dimensions, which allow for both understanding different feelings within an individual and expressing feelings between individuals. We put forward the possibility that the historical delineation of affect along orthogonal axes of valence and arousal might actually describe dimensions onto which compressed affective experiences are projected. In other words, introspective processes graft out affective experiences onto a common space that is well-described by the valence-arousal circumplex (Barrett & Russell, 1998; Russell, 1980).

What this suggests is that, rather than a causal constituent, at this higher-order level of analysis "core affect" constructs of valence and arousal may be a downstream

consequence of a series of hierarchical neural processes that include the compression, and perhaps even meta-cognitive organization, of information. By emphasizing multiple levels of analyzing affective processes, we hope to offer a potential reconciliation between contrasting accounts of valence. We hope to provoke further empirical research to delineate the neural and computational mechanisms underlying evaluative transformations of complex experiences.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Vincent Man http://orcid.org/0000-0001-5380-5521
William A. Cunningham http://orcid.org/0000-0002-7063-5859

References

Adams, C. D., & Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *The Quarterly Journal of Experimental Psychology Section B*, *33*(2b), 109–121.

Adolphs, R., Russell, J. A., & Tranel, D. (1999). A role for the human amygdala in recognizing emotional arousal from unpleasant stimuli. *Psychological Science*, *10*(2), 167–171.

Adolphs, R., & Tranel, D. (2004). Impaired judgments of sadness but not happiness following bilateral amygdala damage. *Journal of Cognitive Neuroscience*, *16*(3), 453–462.

Amano, T., Unal, C. T., & Paré, D. (2010). Synaptic correlates of fear extinction in the amygdala. *Nature Neuroscience*, *13*(4), 489.

Anderson, A., Spencer, D. D., Fulbright, R. K., & Phelps, E. A. (2000). Contribution of the anteromedial temporal lobes to the evaluation of facial emotion. *Neuropsychology*, *14*(4), 526.

Anderson, D., & Adolphs, R. (2014). A framework for studying emotions across species. *Cell*, *157*(1), 187–200.

Arnold, A. E. G. F., Iaria, G., & Ekstrom, A. D. (2016, December). Mental simulation of routes during navigation involves adaptive temporal compression. *Cognition*, *157*, 14–23.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*(3), 183.

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron*, *95*(3), 709–721.e5.

Barbas, H. (2015). General cortical and special prefrontal connections: Principles from structure to function. *Annual Review of Neuroscience*, *38*, 269–289.

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. *Sensory Communication*, *1*, 217–234.

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science*, *1*(1), 28–58.

Barrett, L. F. (2009, July). The future of psychology: Connecting mind to brain. *Perspectives on Psychological Science*, *4*(4), 326–339.

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1–23.

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, *16*(7), 419.

Baxter, M. G., & Murray, E. A. (2002). The amygdala and reward. *Nature Reviews Neuroscience*, *3*(7), 563.

Bell, A. J., & Sejnowski, T. J. (1995, November). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*(6), 1129–1159.

Belova, M. A., Paton, J. J., Morrison, S. E., & Salzman, C. D. (2007). Expectation modulates neural responses to pleasant and aversive stimuli in primate amygdala. *Neuron*, *55*(6), 970–984.

Berridge, K. C. (2019). Affective valence in the brain: Modules or modes? *Nature Reviews Neuroscience, 20(4), 225-234.*

Berridge, K. C., & Robinson, T. E. (1998). What is the role of dopamine in reward: Hedonic impact, reward learning, or incentive salience? *Brain Research Reviews*, *28*(3), 309–369.

Berridge, K. C., & Robinson, T. E. (2003). Parsing reward. *Trends in Neurosciences*, *26*(9), 507–513.

Berridge, K. C., Robinson, T. E., & Aldridge, J. W. (2009). Dissecting components of reward: 'liking', 'wanting', and learning. *Current Opinion in Pharmacology*, *9*(1), 65-73.

Bickart, K. C., Dickerson, B. C., & Barrett, L. F. (2014). The amygdala as a hub in brain networks that support social life. *Neuropsychologia*, *63*, 235–248.

Bonasia, K., Blommesteyn, J., & Moscovitch, M. (2016). Memory and navigation: Compression of space varies with route length and turns. *Hippocampus*, *26*(1), 9–12.

Botvinick, M. (2008). Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences*, *12*(5), 201–208.

Botvinick, M., Niv, Y., & Barto, A. G. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*, *113*(3), 262–280.

Botvinick, M., Weinstein, A., Solway, A., & Barto, A. (2015). Reinforcement learning, efficient coding, and the statistics of natural tasks. *Current Opinion in Behavioral Sciences*, *5*, 71–77.

Bucci, D. J., Holland, P. C., & Gallagher, M. (1998). Removal of cholinergic input to rat posterior parietal cortex disrupts incremental processing of conditioned stimuli. *Journal of Neuroscience*, *18*(19), 8038–8046.

Cacioppo, J. T., & Berntson, G. G. (1992). Social psychological contributions to the decade of the brain: Doctrine of multilevel analysis. *American Psychologist*, *47*(8), 1019–1028.

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, *115*(3), 401.

Cacioppo, J. T., & Berntson, G. G. (1999). The affect system: Architecture and operating characteristics. *Current Directions in Psychological Science*, *8*(5), 133–137.

Cacioppo, J. T., & Gardner, W. L. (1999). Emotion. *Annual Review of Psychology*, *50*(1), 191–214.

Cador, M., Robbins, T., & Everitt, B. (1989). Involvement of the amygdala in stimulus-reward associations: Interaction with the ventral striatum. *Neuroscience*, *30*(1), 77–86.

Calder, A. J., Keane, J., Manes, F., Antoun, N., & Young, A. W. (2000). Impaired recognition and experience of disgust following brain injury. *Nature Neuroscience*, *3*(11), 1077.

Canteras, N., & Swanson, L. (1992). Projections of the ventral subiculum to the amygdala, septum, and hypothalamus:

A phal anterograde tract-tracing study in the rat. *Journal of Comparative Neurology*, *324*(2), 180–194.
Chanes, L., & Barrett, L. F. (2016). Redefining the role of limbic areas in cortical processing. *Trends in Cognitive Sciences*, *20* (2), 96–106.
Cho, J.-H., Deisseroth, K., & Bolshakov, V. Y. (2013). Synaptic encoding of fear extinction in mpfc-amygdala circuits. *Neuron*, *80*(6), 1491–1507.
Clark, A. (2012). Embodied, embedded, and extended cognition. *The Cambridge Handbook of Cognitive Science*, 275–291.
Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The iterative reprocessing model: A multilevel framework for attitudes and evaluation. *Social Cognition*, *25*(5), 736–760.
Davis, M., & Whalen, P. J. (2001). The amygdala: Vigilance and emotion. *Molecular Psychiatry*, *6*(1), 13.
Dickinson, A., & Dearing, M. F. (1979). Appetitive-aversive interactions and inhibitory processes. *Mechanisms of Learning and Motivation: A Memorial Volume to Jerzy Konorski*, 203–231.
Ekman, M., Kok, P., & Lange, F. P. D. (2017). *Time-compressed Preplay of Anticipated Events in Human Primary Visual Cortex 8* (1), 1–9.
Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3–4), 169–200.
Euston, D. R., Tatsuno, M., & McNaughton, B. L. (2007). Fast-forward playback of recent memory sequences in prefrontal cortex during sleep. *Science*, *318*(5853), 1147–1150.
Everitt, B., Cador, M., & Robbins, T. (1989). Interactions between the amygdala and ventral striatum in stimulus-reward associations: Studies using a second-order schedule of sexual reinforcement. *Neuroscience*, *30*(1), 63–75.
Everitt, B. J., & Robbins, T. W. (1997). Central cholinergic systems and cognition. *Annual Review of Psychology*, *48*(1), 649–684.
Feldman Barrett, L., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology*, *74*(4), 967.
Frijda, N. H. (2009). Emotion experience and its varieties. *Emotion Review*, *1*(3), 264–271.
Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360* (1456), 815–836.
Frith, C., & Dolan, R. J. (1997). Brain mechanisms associated with top-down processes in perception. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*. *352*(1358), 1221–1230.
Gilboa, A., & Marlatte, H. (2017). Neurobiology of schemas and schema-mediated memory. *Trends in Cognitive Sciences*, *21* (8), 618–631.
Gray, J. A. (1990). Brain systems that mediate both emotion and cognition. *Cognition & Emotion*, *4*(3), 269–288.
Hatfield, T., Han, J.-S., Conley, M., Gallagher, M., & Holland, P. (1996). Neurotoxic lesions of basolateral, but not central, amygdala interfere with pavlovian second-order conditioning and reinforcer devaluation effects. *Journal of Neuroscience*, *16*(16), 5256–5265.
Hoffman, K. L. (2002). Coordinated reactivation of distributed memory traces in primate neocortex. *Science*, *297*(5589), 2070–2073.
Holland, P. C., & Gallagher, M. (1999). Amygdala circuitry in attentional and representational processes. *Trends in Cognitive Sciences*, *3*(2), 65–73.
Ito, R., & Hayen, A. (2011). Opposing roles of nucleus accumbens core and shell dopamine in the modulation of limbic information processing. *Journal of Neuroscience*, *31*(16), 6001–6007.
Jafarpour, A., & Spiers, H. (2017). Familiarity expands space and contracts time. *Hippocampus*, *27*(1), 12–16.
Janak, P. H., & Tye, K. M. (2015). From circuits to behaviour in the amygdala. *Nature*, *517*(7534), 284.
LaBar, K. S., Gatenby, J. C., Gore, J. C., LeDoux, J., & Phelps, E. A. (1998). Human amygdala activation during conditioned fear acquisition and extinction: A mixed-trial fmri study. *Neuron*, *20*(5), 937–945.
Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1998). Emotion, motivation, and anxiety: Brain mechanisms and psychophysiology. *Biological Psychiatry*, *44*(12), 1248–1263.
LeDoux, J. (1998). Fear and the brain: Where have we been, and where are we going? *Biological Psychiatry*, *44*(12), 1229–1238.
LeDoux, J. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, *23*(1), 155–184.
LeDoux, J. (2007). The amygdala. *Current Biology*, *17*(20), R868–R874.
LeDoux, J. (2012). Rethinking the emotional brain. *Neuron*, *73* (4), 653–676.
LeDoux, J. (2015). *Anxious: Using the brain to understand and treat fear and anxiety*. New York, NY: Penguin.
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*.
LeDoux, J., Farb, C., & Ruggiero, D. A. (1990). Topographic organization of neurons in the acoustic thalamus that project to the amygdala. *Journal of Neuroscience*, *10*(4), 1043–1054.
LeDoux, J., Iwata, J., Cicchetti, P., & Reis, D. J. (1988). Different projections of the central amygdaloid nucleus mediate autonomic and behavioral correlates of conditioned fear. *Journal of Neuroscience*, *8*(7), 2517–2529.
Lewis, A., & Zhaoping, L. (2005). Cone tuning curves and natural color statistics. *Journal of Vision*, *5*(8), 268.
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*(3), 121–143.
Liu, X., Hairston, J., Schrier, M., & Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: A meta-analysis of functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, *35*(5), 1219–1236.
Lomas, T. (2018). Experiential cartography and the significance of âĂIJuntranslatableâĂİ words. *Theory & Psychology*, *28*(4), 476–495.
Louie, K., & Wilson, M. A. (2001). Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron*, *29*(1), 145–156.
Málková, L., Gaffan, D., & Murray, E. A. (1997). Excitotoxic lesions of the amygdala fail to produce impairment in visual learning for auditory secondary reinforcement but interfere with reinforcer devaluation effects in rhesus monkeys. *Journal of Neuroscience*, *17*(15), 6011–6020.
Man, V., Nohlen, H. U., Melo, H., & Cunningham, W. A. (2017). Hierarchical brain systems support multiple representations of valence and mixed affect. *Emotion Review*, *9*(2), 124–132.

Maren, S., & Fanselow, M. S. (1995). Synaptic plasticity in the basolateral amygdala induced by hippocampal formation stimulation in vivo. *Journal of Neuroscience*, *15*(11), 7548–7564.

McNaughton, N., & Gray, J. A. (2000). Anxiolytic action on the behavioural inhibition system implies multiple types of arousal contribute to anxiety. *Journal of Affective Disorders*, *61*(3), 161–176.

Moscovitch, M., & Nadel, L. (1998). Consolidation and the hippocampal complex revisited: In defense of the multiple-trace model. *Current Opinion in Neurobiology*, *8*(2), 297–300.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*(6583), 607–609.

Parkinson, J. A., Crofts, H. S., McGuigan, M., Tomic, D. L., Everitt, B. J., & Roberts, A. C. (2001). The role of the primate amygdala in conditioned reinforcement. *Journal of Neuroscience*, *21*(19), 7770–7780.

Paton, J. J., Belova, M. A., Morrison, S. E., & Salzman, C. D. (2006). The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature*, *439*(7078), 865.

Petty, R. E., Briñol, P., Tormala, Z. L., & Wegener, D. T. (2007). The role of meta-cognition in social judgment. *Social Psychology: Handbook of Basic Principles*, *2*, 254–284.

Quirk, G. J., Armony, J. L., & LeDoux, J. (1997). Fear conditioning enhances different temporal components of tone-evoked spike trains in auditory cortex and lateral amygdala. *Neuron*, *19*(3), 613–624.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79.

Redondo, R. L., Kim, J., Arons, A. L., Ramirez, S., Liu, X., & Tonegawa, S. (2014). Bidirectional switch of the valence associated with a hippocampal contextual memory engram. *Nature*, *513*(7518), 426.

Reynolds, S. M., & Berridge, K. C. (2008). Emotional environments retune the valence of appetitive versus fearful functions in nucleus accumbens. *Nature Neuroscience*, *11*(4), 423.

Richard, J. M., & Berridge, K. C. (2011). Nucleus accumbens dopamine/glutamate interaction switches modes to generate desire versus dread: D1 alone for appetitive eating but d1 and d2 together for fear. *Journal of Neuroscience*, *31*(36), 12866–12879.

Rigotti, M., Ben-Dayan Rubin, D. D., Wang, X.-J., & Fusi, S. (2010). Internal representation of task rules by recurrent dynamics: The importance of the diversity of neural responses. *Frontiers in Computational Neuroscience*, *4*, 24.

Rizvi, T. A., Ennis, M., Behbehani, M. M., & Shipley, M. T. (1991). Connections between the central nucleus of the amygdala and the midbrain periaqueductal gray: Topography and reciprocity. *Journal of Comparative Neurology*, *303*(1), 121–131.

Robin, J., & Moscovitch, M. (2017). Details, gist and schema: HippocampalâĂŞneocortical interactions underlying recent and remote episodic and spatial memory. *Current Opinion in Behavioral Sciences*, *17*, 114–123.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161.

Salzman, C. D., & Fusi, S. (2010). Emotion, cognition, and mental state representation in amygdala and prefrontal cortex. *Annual Review of Neuroscience*, *33*, 173–202.

Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. *Approaches to Emotion*, *2293*, 317.

Schoenbaum, G., Chiba, A. A., & Gallagher, M. (1999). Neural encoding in orbitofrontal cortex and basolateral amygdala during olfactory discrimination learning. *Journal of Neuroscience*, *19*(5), 1876–1884.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, *20*(1), 11.

Setlow, B., Holland, P. C., & Gallagher, M. (2002). Disconnection of the basolateral amygdala complex and nucleus accumbens impairs appetitive pavlovian second-order conditioned responses. *Behavioral Neuroscience*, *116*(2), 267.

Shabel, S. J., & Janak, P. H. (2009). Substantial similarity in amygdala neuronal activity during conditioned appetitive and aversive emotional arousal. *Proceedings of the National Academy of Sciences*, *106*(35), 15031–15036.

Solway, A., Diuk, C., Córdova, N., Yee, D., Barto, A. G., Niv, Y., & Botvinick, M. (2014). Optimal behavioral hierarchy. *PLoS Computational Biology*, *10*(8), e1003779.

Touroutoglou, A., Hollenbeck, M., Dickerson, B. C., & Barrett, L. F. (2012). Dissociable large-scale networks anchored in the right anterior insula subserve affective experience and attention. *Neuroimage*, *60*(4), 1947–1958.

Tse, D., Langston, R. .F, Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., ... Morris, R. G. M. (2007). Schemas and memory consolidation. *Science*, *316*(5821), 76–82.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, *98*(2), 219.

Winocur, G., & Moscovitch, M. (2011). Memory transformation and systems consolidation. *Journal of the International Neuropsychological Society*, *17*(5), 766–780.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind/brain perspective. *Psychological Bulletin*, *133*(2), 273–293.

Zald, D. H., & Pardo, J. V. (1997). Emotion, olfaction, and the human amygdala: Amygdala activation during aversive olfactory stimulation. *Proceedings of the National Academy of Sciences*, *94*(8), 4119–4124.

Zhang, W., Schneider, D. M., Belova, M. A., Morrison, S. E., Paton, J. J., & Salzman, C. D. (2013). Functional circuits and anatomical distribution of response properties in the primate amygdala. *Journal of Neuroscience*, *33*(2), 722–733.

Zhaoping, L. (2014). *Understanding vision: Theory, models, and data* . Oxford, UK: Oxford University Press.

Zikopoulos, B., John, Y. J., García-Cabezas, M. Á., Bunce, J. G., & Barbas, H. (2016). The intercalated nuclear complex of the primate amygdala. *Neuroscience*, *330*, 267–290.