Neuroscience and Biobehavioral Reviews 69 (2016) 216–238

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

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

Unconscious emotion: A cognitive neuroscientific perspective

Ryan Smith (Ph.D.)[a,*], Richard D. Lane[a,b,c]

[a] Department of Psychiatry, University of Arizona, Tucson, AZ, United States
[b] Department of Psychology, University of Arizona, Tucson, AZ, United States
[c] Department of Neuroscience, University of Arizona, Tucson, AZ, United States

ARTICLE INFO

Article history:
Received 21 September 2015
Received in revised form 6 July 2016
Accepted 9 August 2016
Available online 10 August 2016

Keywords:
Emotion
Cognition
Appraisal
Internal models
Emotion regulation
Consciousness
Unconscious processing
Interoception
Medial prefrontal cortex (MPFC)
Anterior cingulate cortex (ACC)
Insula

ABSTRACT

While psychiatry and clinical psychology have long discussed the topic of unconscious emotion, and its potentially explanatory role in psychopathology, this topic has only recently begun to receive attention within cognitive neuroscience. In contrast, neuroscientific research on conscious vs. unconscious processes within perception, memory, decision-making, and cognitive control has seen considerable advances in the last two decades. In this article, we extrapolate from this work, as well as from recent neural models of emotion processing, to outline multiple plausible neuro-cognitive mechanisms that may be able to explain why various aspects of one's own emotional reactions can remain unconscious in specific circumstances. While some of these mechanisms involve top-down or motivated factors, others instead arise due to bottom-up processing deficits. Finally, we discuss potential implications that these different mechanisms may have for therapeutic intervention, as well as how they might be tested in future research.

Contents

* Corresponding author at: Department of Psychiatry, University of Arizona, 1501 N. Campbell Ave., Tucson, AZ 85724-5002, United States.
E-mail address: rssmith@email.arizona.edu (R. Smith).

http://dx.doi.org/10.1016/j.neubiorev.2016.08.013
0149-7634/

1. The need for reconsideration of unconscious emotion

The concept of unconscious emotion may at first blush appear to be an oxymoron. What are emotions if not intense, conscious, subjective experiences that constitute our greatest joys and greatest sorrows? Emotion arguably makes living worthwhile: consider the value of life if emotions could not be experienced. Yet, with the advent of cognitive neuroscience and the foundational distinction between implicit and explicit processes that apply to all major areas of cognition, including perception, attention, memory and decision-making, coupled with the realization that the generation, expression, experience, and regulation of emotion all involve perceptual/cognitive mechanisms, it has been argued that the same implicit-explicit distinction that applies to cognition generally also applies to emotion (Kihlstrom et al., 2000; Lane et al., 2000; Smith and Lane, 2015). Indeed, evidence that unconscious emotion exists, at least in some forms, is now fairly strong. For example, emotionally relevant stimuli presented so briefly that perceptual awareness is not possible nevertheless reliably influence preferences (Zajonc, 1980), consummatory behavior (Winkielman and Berridge, 2004), and can also trigger other emotion-related physiological/behavioral reactions (Tamietto and de Gelder, 2010). A growing literature has also established that unconscious or implicit attitudes and beliefs have a profound effect on social behavior (McConnell and Leibold, 2001).

After reviewing a considerable body of such evidence from several research domains – including studies of implicit memory, the subliminal mere exposure effect, and deficit profiles in neurological and psychiatric patients, as well as studies of observed dissociations between the various components of an emotional response – it was suggested by Kihlstrom et al. (2000) that the unconscious emotional effects observed in these studies can be sub-divided into two broad categories of phenomena. The first category – which we will call "unconsciously generated emotion" – involves cases where emotional responses are themselves consciously experienced/recognized, but where those emotions are generated in response to unconscious processes (e.g., unconscious percepts, thoughts, or memories). In such cases, an individual will report feeling an emotion, but they will not be consciously aware of the internal/external event that caused the feeling. A slight variant that also falls within this broad category is a set of cases where one is conscious of both the emotional response and the eliciting cause, but where one remains unaware of the causal relation between them. For example, one might consciously perceive a desk and consciously experience becoming sad, and yet not be aware that the desk-percept caused the sadness response. In contrast to such cases, the second category suggested by Kihlstrom et al. (2000) – which he calls "implicit emotion" – is instead when an emotional response is generated but not consciously experienced/recognized. In this type of case, for example, a person might display an automatic fearful facial expression, exhibit an increased heart rate, and behave avoidantly in response to a stimulus, and afferent feedback would trigger unconscious representations of these changes in the brain – yet the person would not report consciously feeling fear. While Kihlstrom et al. (2000) identify many behavioral findings that are consistent with both categories, they identify very few instances in which the neural basis of such effects is examined, and this characterization remains largely true to date. Thus, although the behavioral reality of unconscious emotion has been fairly well established, a more detailed consideration of the origins, mechanisms, and maintenance of unconscious emotion from a cognitive neuroscientific perspective has not been undertaken, and it is the aim of this paper to attempt to fill this gap.

There are several reasons why a review of this topic is needed. First, advances in basic emotion theory point to the importance of unconscious emotion. In "Rethinking the Emotional Brain" (LeDoux, 2012), Ledoux addressed the challenges of linking animal and human research on emotion given that humans can report on their conscious experiences whereas other animals cannot. He proposed that emotions occur when survival circuits[1] are activated (in humans or other animals), leading to changes in various aspects of behavior, cognition, and physiology. Crucially, he argued that the activation of such circuits is not sufficient to generate a conscious feeling on its own. Instead, these activations must interact with other neural systems involved in conscious processing and awareness (i.e., if the organism in question possesses them), indirectly contributing to the generation of a subjective feeling. In the case of humans, we are learning a great deal about the neural basis of consciousness in relation to multiple domains of cognition, particularly visual and auditory perception (Dehaene, 2014). This work has revealed a great deal about the mechanisms of unconscious and conscious cognition, but, with few exceptions, these insights have not been applied to emotion. Addressing this topic will advance our understanding of how humans are and are not like our phylogenetic neighbors. With regard to humans, we have recently published a review of the hierarchical neural networks responsible for the generation, perception and regulation of conscious and unconscious emotion (Smith and Lane, 2015), which assumed that the full range of processing from unconscious to conscious would occur in each domain. In this paper we consider for the first time from the perspective of that model how emotion that is and remains unconscious (i.e., the "implicit emotion" category) may come about.

A second important reason for addressing this topic involves the clinical domain of psychotherapy. Traditional psychoanalytic concepts of affect held that unconscious emotions residing in the id pressed for discharge but were held in the unconscious by the forces of repression (Brenner, 1973). The advances in cognitive neuroscience alluded to above have led to some recognition within psychoanalysis that concepts about the unconscious should be updated. For example, Modell has called for a shift from traditional concepts of the unconscious as a cauldron of forbidden impulses to a cognitive and affective unconscious that is fundamentally adaptive (Modell, 2010, 2008), and Ginot has elaborated on the empirical foundation/justification and clinical implications of such a shift (Ginot, 2015).

Within psychoanalysis, alternative models of psychopathology focusing on dissociation (rather than conflict and repression) and the importance of the interpersonal relationship between therapist and client have been proposed – supported in part by findings in modern cognitive and affective neuroscience (Bucci, 2016). This perspective highlights the need to convert subsymbolic emotional responses to symbolic, conceptual representations of emotional experience. More generally, a fundamental principle of many psychotherapy modalities is that "emotion processing" is a necessary ingredient for therapeutic success. A quintessential example of this is Emotion Focused Therapy, which has a substantial record of empirical research supporting it, both in terms of outcome and process research (Greenberg, 2010). This form of therapy involves helping clients to experience their emotions, to become aware of them, to label them, understand them, and transform them. *However, the nature of the emotion prior to it being further processed in this manner is currently considerably less clear.* Recent work investigating the cognitive and neural processes underlying conscious

[1] While Ledoux believes that these survival circuits are responsible for generating the autonomic, cognitive, and behavioral reactions associated with the term "emotion," he does not believe that there is a different circuit for each of the "basic emotion" concepts often used in psychological research (e.g., sadness, happiness, fear). Instead such basic emotion terms are likely applied to the outputs of different circuits in different contexts, and their use is also likely dependent on previous learning.

and unconscious emotion suggests, however, that there may be many variants of unrecognized or cognitively inaccessible emotion (Brosch and Sander, 2013; Lambie and Marcel, 2002; Lane et al., 2015b; Smith and Lane, 2015; Winkielman and Berridge, 2004). Knowing what these different variants of implicit emotion are, and what processes keep these emotional responses from being recognized, might significantly assist clinical interventions. That is, a more detailed, up to date model of these processes could allow clinicians to tailor interventions to what each individual client needs. Thus, in this paper we aim to provide a basic taxonomy of different neuroscientifically plausible mechanisms that may be capable of accounting for implicit emotions, and to highlight their various implications for therapeutic intervention.

A third important domain in which unconscious emotion appears to be highly relevant is physical health. The association between depression or anxiety and early mortality in a variety of disease contexts is now unequivocal (Frasure-Smith and Lesperance, 2005; Friedman and Thayer, 1998; Grippo and Johnson, 2009; Gross and Levenson, 1997; Kemp et al., 2010; Thayer and Lane, 2007; Thayer et al., 2010). Although the latter findings are largely based on self-reported emotional experiences, recent research also indicates that the majority of instances in which mental stress results in detrimental physiological changes are not associated with self-reported emotion (Brosschot, 2010; Brosschot et al., 2010), particularly among people with lower trait emotional awareness (Verkuil et al., 2016). This implies that emotional arousal may often not be consciously recognized, and, combined with other findings, it appears clear that persistently activated (but consciously unrecognized) emotion can have prolonged physiological effects that are deleterious to health (Lane, 2008). An example (now replicated in several studies, reviewed in Slavich and Irwin, 2014) of such mechanisms is that negative emotional arousal can produce increases in circulating pro-inflammatory cytokines (mediated by increased sympathetic tone, reduced vagal tone, and related endocrine responses), and that such changes, when chronic, can increase risk of systemic disease. Therefore, a better understanding of the ways in which emotion can remain unconscious/unrecognized, and how this can be overcome, might therefore lead to improved ways to assess it, understand its underlying mechanisms, and intervene to prevent adverse health consequences.

A fourth domain, related to all of the preceding three, involves the phenomenon of somatization or the expression of emotional distress in the form of physical symptoms. It is estimated that at least one third of all medical visits involve somatic complaints that are not adequately explained by detectable physical disease processes (Kroenke, 2003), and many of these complaints may be attributable to unrecognized emotional causes (Konnopka et al., 2012; Sharpe and Carson, 2001). As impairments in affective theory of mind have been shown to contribute to this type of somatization (Stonnington et al., 2013; Subic-Wrana et al., 2010), and as somatization also represents a significant health care cost (Konnopka et al., 2012), better understanding in the area of unconscious emotion might lead to improvements in diagnosis, treatment, and potentially prevention.

Further, research within the past few decades appears to support the possibility that pain intensity within chronic pain syndromes can be amplified by suppressed or unrecognized emotional reactions. For example, carefully controlled experiments have provided evidence for the role of inhibited anger in increasing somatic pain (Burns et al., 2008), and current developmental theories seeking to explain somatoform pain have stressed the important contribution of both early life adversity and the unmet need for emotional closeness with others (Landa et al., 2012). As briefly mentioned above, because negative emotion can increase bodily inflammatory responses via descending autonomic pathways (Slavich and Irwin, 2014), and inflammatory responses are also known to amplify pain (e.g., Woolf et al., 1997), there is also a plausible physiological process whereby such effects may occur. More generally, "unspeakable" personal dilemmas, in which one's own needs and desires are perceived to conflict with internalized social norms and values, appear to represent a common context in which unrecognized emotional arousal contributes to a variety of physical symptoms (Griffith and Griffith, 1994).

There are many well-documented cases of such unspeakable dilemmas and their association with both amplified pain and clinical observations consistent with unconscious emotion (Anderson and Sherman, 2013; Anderson, 2017, 1998). However, most clinical cases of this kind are often unavoidably "messy" and complex, leading to difficulties in applying any idealized neuro-cognitive model of the type we present below with confidence. To provide a concrete example, we will therefore make use of a simpler case vignette that is representative of the basic elements of the types of clinical cases that we are interested in explaining. Specifically, we will use the case of "Walter" (Conenna, 2013; pgs 52–53)[2] who is described as currently mourning the loss of his wife of fifty years named Martha, and who said he felt both "grief and sadness." However, he had also simultaneously acquired unpleasant pain in his back that he did not understand. Walter further claimed to be angry at "life," because "it's not fair that good people die," and he also said he wished Martha was with him right now. Over the course of therapy, it was suggested that Walter was actually angry with Martha for dying. After a long period of silence, Walter was provoked to cry as a result of this suggestion. He eventually composed himself, and stated: "Yes. I can see that I've been angry with Martha. And I can see that this doesn't mean that I love her any less." At the end of the session, Walter is described as walking out of the room displaying signs of improved mood and reduced somatic pain.

As stated above, we intend to use this case vignette as a device for the purposes of illustration/exposition; we believe it serves as a simplified, yet concrete example representative of many others (Anderson and Sherman, 2013; Anderson, 2017, 1998), and it contains the basic elements of internal conflict, unrecognized emotion, and somatic symptoms with which many clinicians are regularly confronted. In this case, Walter was aware of his back pain, his sadness/grief due to the loss of his wife, and his anger at life. However, he did not have conscious access to the fact that he was angry with Martha, and this appears to be related to the fact that blaming (or being angry with) his wife for dying was "impermissible" (in the sense that it conflicted with his own explicit norms/values and seemed incompatible with his love for her). This case represents the type of clinical phenomena we seek to explain through consideration of plausible mechanisms delineated in pre-clinical cognitive neuroscience research. After introducing relevant background information directly below, we will return to this case and illustrate how some of the mechanisms we propose may be able to account for experiences like Walter's in a neurobiologically plausible manner.

[2] The author was a patient himself and had similar personal experiences with pain and emotion. Although the author is not a clinician or researcher himself, the facts and clinical observations as stated have been verified with the author [personal communication 6/23/16] as an accurate account of an actual clinical encounter with another patient witnessed by the author. Details such as whether Walter had a history of abandonment in childhood, or the degree to which the marriage with Martha was loving or ambivalent, would certainly be clinically relevant but are not known. Not knowing these details, we submit, makes the case more tractable for the analysis we intend, whereas to include such details would add even greater complexity, would require more specificity rather than generality, and thus would distract from the clarity and purpose of the exposition.

2. A cognitive neuroscientific approach to unconscious emotion

In spite of the evidence supporting the existence of unconscious emotion and the need to better understand it, the nature of unconscious emotion remains controversial. For example, the idea of "repression" – that complex, motivated unconscious factors can keep fully formed thoughts, feelings, and memories out of awareness – has been criticized in light of current conceptions within the cognitive and neural sciences (Kihlstrom, 2002; Rofé, 2008). On the other hand, a growing body of work within cognitive neuroscience has recently provided evidence for the ability to intentionally suppress the retrieval of both emotional and non-emotional memories (Anderson and Hanslmayr, 2014; Depue, 2012), lending support to the idea that unconscious motivational factors may be capable of modulating conscious access to mentally represented information. In a recent review we have also defended the possibility that, in some cases, the conceptual meaning of one's own emotional reactions may fail to be appropriately represented at all, independent of the question of conscious access (Lane et al., 2015b). Such ideas are also broadly consistent with long-standing theories that separate emotions into theoretically distinct components, including physiological (autonomic/endocrine/immune) responses, behavioral (skeletomotor) responses, and cognitive responses (Lang, 1988, 1968; Rachman, 1978). Specifically, because evidence suggests that these different components can become desynchronized from one another (Hodgson and Rachman, 1974; Rachman and Hodgson, 1974), and because part of the cognitive response can be understood to involve conscious awareness (Kihlstrom et al., 2000), it is plausible to imagine that conscious awareness of emotion could desynchronize from the physiological and behavioral responses (as well as from other aspects of the cognitive response). The result would be a lack of conscious awareness of one's physiological/behavioral reactions and/or a lack of awareness of their emotional meaning (e.g., that they signify an emotion like sadness), and thus a lack of reportable emotional feeling.

In this article, we will use both the term "unconscious" and the term "unrecognized" to jointly describe instances, such as that of "Walter," in which an emotional response is activated but not verbally reported or otherwise consciously understood (and thus not reportably experienced as a discrete emotional feeling). This falls into the "implicit emotion" category discussed in Section 1. Since Freud (and some of his predecessors), many have previously used the term "unconscious," but we believe that this term may at times lead to confusion; this is because emotional reactions have multiple components (Rachman, 1978; Shiota and Kalat, 2012), and, in the cases under consideration, individuals appear to be conscious of some aspects of their emotional response and not others. For example, they may be fully conscious of the bodily aspects of their emotional reaction (e.g., they may feel an increase in heart rate), they might be aware of strong impulses to act in a specific manner, and they might even recognize that they have certain types of attentional/memory biases in their present state. What is clear, however, is that in the types of cases we focus on here, such individuals often do not *consciously recognize* their reactions as being emotional in nature. That is, they do not have conscious access to the fact that their reaction is related to an emotion concept like anger or guilt. Alternatively, they might *misidentify* the nature of their emotional reaction. For example, they might misidentify anger as fear. This might be especially likely to occur in cases where a person also misidentifies the cause of their emotion, as with the phenomena discussed in Section 1 associated with the "unconsciously generated emotion" category (reviewed in Kihlstrom et al., 2000).

Given the paucity of available research that is specifically directed at unconscious emotion, our major strategy will involve taking broader lessons gleaned from the cognitive/neural sciences and applying them to emotion – something that has not been done previously. Our starting point is a focus on internal representations of situations and how their significance is automatically appraised (including representations of the self in relation to those situations). Emotion involves an assessment of the extent to which needs, goals, and values are met or not met in a given situation, and an adjustment of behavior, physiology, cognition and experience to adapt to that situation (Levenson, 1994). The way situations are internally represented (both perceptually and conceptually), and how the significance of such representations is then automatically appraised, are both major determinants of whether an emotional reaction is generated or not and, if so, what emotion is generated. Once the emotional reaction is generated, the different aspects of this reaction may or may not become consciously accessible. For example, according to our previous hierarchical model (Smith and Lane, 2015), if a person's attention were sufficiently distracted, they might not become aware of their bodily response (e.g., increased heart rate, change in posture, etc.). Further, a person might be aware of such bodily reactions, but still fail to consciously recognize this reaction as one of, for example, fear. Several additional factors will influence whether or not each aspect becomes conscious. As such, we will discuss four foundational processes that determine whether different aspects of an emotional reaction will become conscious or not in a given situation.

1. Constructing/maintaining a probabilistic, unconscious "internal model," which stores/represents acquired knowledge about one's self, the world and their interaction.
2. The automatic appraisal of the current situation as it is represented in this unconscious internal model.
3. Selective conscious access to elements of this (otherwise unconscious) internal model.
4. Cognitive control processes associated with both the automatic and goal-directed amplification/suppression (termed "dynamic filtering") of conscious access to representations within this internal model.

As detailed below, we will argue that neuroscientific work on automatic situational appraisal (Brosch and Sander, 2013) can be combined with recent work on probabilistic, hierarchical internal models (Friston, 2005; Hohwy, 2014; Moreno-Bote et al., 2011; Vul and Pashler, 2008; Vul et al., 2009) to explain the unconscious generation of an emotional reaction. By "emotional reaction" we mean a combination of both an automatically triggered autonomic/somatic reaction (e.g., changes in facial expression, posture, breathing, vasoconstriction, etc.) and an automatically triggered cognitive reaction (e.g., attentional/memory biases, strong motivations/impulses to decide to act in specific ways, etc.). We will argue, however, that recent models of consciousness, which involve many unconscious representations simultaneously competing for conscious access (Dehaene et al., 2006), can be combined with work on both top-down memory suppression (Anderson and Hanslmayr, 2014; Kuhl et al., 2007) and prefrontal cortex (PFC) dynamic filtering mechanisms (Shimamura, 2000), to explain how motivated factors could prevent specific aspects of this emotional response from being selected for conscious access after they are unconsciously perceived/recognized.

After reviewing current work in these areas, we will highlight plausible interactions between these processes, and we will illustrate how such interactions can be combined to provide multiple mechanisms capable of explaining the cases of unconscious/unrecognized emotion discussed above. In the taxonomy we provide below, we will first discuss various "top-down" mechanisms capable of keeping an emotional reaction from being consciously recognized. We will then discuss some alternative "bottom-up" mechanisms that are also able to account for phe-

notypically similar phenomena. Finally, we will discuss possible implications of these various mechanisms for future research and clinical practice.

3. Internal models and automatic appraisal

There have been several recent advances in the neural sciences with regard to the nature of unconscious mental processes. For example, when one perceives, remembers, or imagines their past, present, or future position in the world, (e.g., one's relation to the relevant objects, people, and contexts in question), the various perceptual/conceptual properties of these actual/potential situations are now thought to be represented across hierarchically organized sensory, memory, and motor systems within the brain (Danker and Anderson, 2010; Dayan and Daw, 2008; Friston, 2005; Hohwy, 2014; Kiefer and Barsalou, 2013; Kreiman et al., 2000; O'Craven and Kanwisher, 2000; Pezzulo et al., 2015; Pouget et al., 2000). Hierarchical internal representations are also held about one's self, including representations of both one's physical and mental attributes (Friston and Frith, 2015; Metzinger, 2003; Northoff et al., 2006). This large set of inter-related representations jointly comprises one's "internal model" of the world, the self, other people, and their relations to one another. However, as we will describe in more detail below, at any given moment the vast majority of the information represented within this internal model is unconscious; one only selectively gains conscious access to specific aspects of this internal model based on a selection process that takes various factors into account, such as salience, goal-relevance, and probability of accuracy (Dehaene, 2014; Dehaene et al., 2006; Moreno-Bote et al., 2011; Vul and Pashler, 2008; Vul et al., 2009).

Counterintuitively, unlike the single, discrete percepts and beliefs that are consciously experienced, recent work suggests that the representations held within this larger unconscious internal model are instead probabilistic; that is, as opposed to distinct all-or-none representations, the brain instead simultaneously unconsciously represents many different possible interpretations (e.g., perceptions, beliefs) about the world, self, and their relation, along with the probability that each interpretation is correct (Moreno-Bote et al., 2011; Pouget et al., 2000; Vul and Pashler, 2008; Vul et al., 2009). For example, recent studies have shown that when shapes are embedded within the "ground" side of a visual figure, participants do not report consciously recognizing them; yet, both neural and behavioral evidence can be found that the brain unconsciously and automatically represents possible semantic/conceptual interpretations of the shape (Cacciamani et al., 2014; Sanguinetti et al., 2014). However, being on the "ground" side lowers the probability that this interpretation is the right one, and this reduces the chances that it will be selected for conscious access. Such studies suggest, therefore, that many interpretations of one's situation, at both abstract and concrete levels of description, are automatically represented prior to conscious processing. These internally represented probability distributions across multiple perceptual/conceptual interpretations are continually updated based on new sensory input, likely based on an iterative process that attempts to minimize the error between this sensory input and the input predicted by the internal model in its current form (Hohwy, 2014; Pezzulo et al., 2015).

The unconscious information within such an internal model can also take multiple forms (Dehaene et al., 2006). One way in which information can be unconscious is if it takes the form of activated, content-bearing neural states, which are not consciously accessed. In such cases, informational content is represented, but the individual is unaware of it. This can occur during the presentation of subliminal stimuli, or when attentional resources are sufficiently taxed by a separate ongoing task (Simons and Chabris, 1999). As discussed above, it can also happen if the representation in question is estimated to have a lower probability of being correct compared to other competing interpretations.

In contrast, a further way in which information can remain unconscious is if it is not represented by an active neural state at all, but *only implicitly represented* within the brain's structure – such as within the pattern of synaptic connections between neurons (and the specific strengths of these connections). When a given neural state is activated, for example, this structure will determine how connected regions of the brain and body will respond. Hence, it can embody specific implicit rules about which stimulus representations are linked to which responses, or which representations predict which other representations (Pezzulo et al., 2015). For example, due to either innate or learning-based factors, the representation of a certain perceived movement pattern might be synaptically "wired-up" to directly trigger a fight-or-flight response; in such cases, the perceived movement need not first activate a neural representation of the concept "threat" to trigger this response – and thus there might not be any unconsciously activated "threat" representation to gain conscious access to (LeDoux, 1996). Similar phenomena have been described by Tomkins in which basic expressions of affect in infants (such as startle or laughter) are induced by simple changes in the intensity of sensory stimulation (Tomkins, 1995). These may be good examples of what Zajonc referred to when he argued that "preferences need no inferences" (Zajonc, 1980). Relatedly, similar synaptic connection strength patterns are also thought to play a critical role in long-term memory storage. For example, during memory retrieval, specific patterns of connections (acquired due to past experience) allow for the recreation of a pattern of neural activity similar to the one that was generated during the perception of an event – and this recreated activity pattern represents the various aspects of the memory of that event (Danker and Anderson, 2010; Lynch, 2004).

Finally, it is important to highlight that internal models within the brain are thought to simultaneously represent descriptions of the self and the world (including other people) at many levels of abstraction (Hohwy, 2014). For example, while neurons in primary sensory regions appear to generate representations associated with very concrete, low-level regularities (such as the presence of a specific auditory tone), higher cortical levels generate representations associated with more abstract regularities (such as whether a series of tones signifies a specific word/concept). Each such cortical region might be thought of as representing one particular hypothesis space (e.g., the space of possible tones, or the space of possible words, etc.), and assigning a probability value to each possibility within its respective space in response to a new wave of sensory input. While these different hierarchical regions interact and inform one another (e.g., hearing certain tones might be more consistent with hearing certain words), it is important to keep in mind that they represent distinct layers of information that can be present within experience, and one might lose conscious access to one layer but not the other (e.g., as when one is conscious of automatically clenching one's fists, but is not conscious of being angry).

As we have characterized it, this unconscious internal model that is maintained in the brain can be thought of as descriptive in nature. For example, its content (when verbally described) could include the following as one represented high-probability interpretation: "I am currently sitting in a classroom, there is a green chalkboard in front of me, and I feel bored." However, the significance of this description to the person whose brain is representing it still requires evaluation (with regard to their own needs, goals, values, etc.). "Automatic appraisal" refers to a further process within the brain whereby distinct mechanisms receive this probabilistic, largely unconscious "description" held within the internal model, and then evaluate its significance in the present moment for the person along various dimensions. It is referred to as "automatic"

because at least many of the appraisal mechanisms in question appear capable of operating outside of awareness, and may not require attention to the specific properties of one's internal model that are being evaluated (Brosch and Sander, 2013).

We have recently proposed and defended a neural model of the generation, perception, and regulation of one's own emotional states (Smith and Lane, 2015), and in that model, appraisal also operates on a hierarchical, iterative basis. Some appraisal dimensions, such as "novelty" and "concern relevance," can be evaluated quickly, whereas others, such as "goal-congruence," "agency/control," and "compatibility with norms/values," may require greater computational resources (and hence more time). Appraising one's internally represented situation along each of these dimensions also appears to involve distinct brain regions. In our model, for example, the initial evaluation of concern relevance recruits the amygdala, whereas evaluations of goal-congruence recruit the dorsal anterior cingulate (dACC) and dorsolateral prefrontal cortex (DLPFC). Evaluating agency/control instead involves a broad network of regions, including dorsomedial prefrontal cortex (DMPFC), the temporoparietal junction (TPJ), and sensory-motor cortices, among other regions. Evaluating norm/value compatibility (described further below) recruits anterior temporal and dorsolateral frontal regions. Finally, the ventromedial prefrontal cortex (via interaction with several other appraisal-related brain regions) may eventually arrive at a global evaluation of the affective meaning of one's represented situation, after integrating broader information about one's current context and one's present goals (Roy et al., 2012). This region may also be necessary for appraising, and unconsciously generating emotional reactions, while entertaining possible, as opposed to actual, situations (Gupta et al., 2011).

Although the exact number and type of appraisal dimensions posited varies between theories (for example, see Frijda, 1986; Lazarus, 1991; Reisenzein, 2006; Roseman et al., 1990; Scherer, 1984), cross-cultural research supports the idea that unique patterns in self-reported judgments across the appraisal dimensions in such theories are associated with different self-reported basic emotions (Moors et al., 2013; Scherer, 1997). In general, the function of each of these appraisal mechanisms is to (1) detect/evaluate features of one's internal model that are significant to one's current concerns, needs, goals, and values, and (2) initiate a set of appropriate cognitive, autonomic, neuroendocrine, and behavioral changes in response. Given the iterative, hierarchical nature of these mechanisms, however, it is possible that one appraisal mechanism might quickly initiate this type of "emotional" cognitive/bodily reaction, whereas another might adjust that reaction shortly thereafter as more information becomes available.

Crucially, in our model these appraisal mechanisms do *not* only evaluate the unconscious representations that are estimated to be the most likely (and which typically become conscious); instead, *they will evaluate the full probability distribution that is unconsciously represented across possible interpretations*. In other words, before initiating the cognitive/bodily responses described above, appraisal mechanisms will take into account the full range of possible interpretations that are represented unconsciously – and the probability of correctness assigned to each one. This means that, for example, an emotional reaction might be generated by automatic appraisal mechanisms in response to the unconsciously represented possibility of "threat," even if another competing interpretation of "non-threat" was represented as having a higher probability of accuracy within the internal model (and was consciously accessible). Further, as at least many of these mechanisms can operate outside of awareness, these emotional reactions can be generated without being consciously perceived or understood. For example, say that sensory input updated one's internal model such that it now represented a person as being "disrespectful" (i.e., it represented this interpretation as now having a high probability of accuracy). If so, this could, via these appraisal mechanisms, initiate (1) a set of changes in one's autonomic state, facial expression, and body posture (Friedman and Kreibig, 2010; Kragel and Labar, 2013; Nummenmaa et al., 2014), as well as (2) a set of biases in attention, memory retrieval, and action selection (Gupta et al., 2011; Huntsinger, 2013; Lewis et al., 2005; Mitchell, 2011), all without a person having conscious access to the reason for this reaction, or to the fact that their reaction was one of anger. Whether a person became aware of these further facts would depend on other processes that influence the competition for conscious access (described below).

The idea that unconscious interpretations are still evaluated by automatic appraisal mechanisms is supported by studies that have shown that bodily emotional reactions, and biases in cognition and action selection, can be generated (e.g., by subcortical structures such as the amygdala) via subliminal perception of emotional stimuli (reviewed in Kihlstrom et al., 2000; Tamietto and de Gelder, 2010). Such effects can be explained by the idea that emotional reactions are generated in response to represented interpretations within the internal model that are not selected for conscious access. In the case of subliminal perception, such an interpretation would remain unconscious due to the fact that sensory input is too weak to provide sufficient evidence in favor of that interpretation, and hence its estimated probability would remain too low to become conscious (see Section 4 on conscious access below). However, the relationship between unconscious probabilistic representation, automatic appraisal, and the generation of an emotional reaction remains largely unexplored. *One interesting hypothesis, however, is that the degree to which an emotional reaction is generated by an unconscious interpretation may be proportional to the estimated probability that the interpretation in question is correct.* Thus, for example, if an unconscious interpretation of "threat" was estimated to have a 30% probability of being correct, appraisal mechanisms might generate less intense autonomic arousal than if it was instead represented to have a 40% probability of accuracy. This could be the case, even if another interpretation of "non-threat" had the highest represented probability estimate in both cases, and was the interpretation selected for conscious access. Future research should test this hypothesis empirically, as it could play an important role in the bottom-up mechanisms for keeping emotion unconscious that we will describe later in this paper.

While we have thus far described appraisal mechanisms as simply "evaluating" a represented description within one's probabilistic internal model, certain learning processes also appear capable of directly modifying the circuitry within appraisal-related structures such that the same represented description can provoke a different appraisal. This may often involve both classical and operant conditioning processes (Cushman, 2013; Daw and Shohamy, 2008; LeDoux, 2012, 2013; Mitchell, 2011; Pessoa and Adolphs, 2010); for example, classic work on the amygdala has shown that synaptic changes within specific amygdalar nuclei underlie the acquisition of conditioned fear (LeDoux, 2012, 1996). However, recent considerations of the role of emotions in moral psychology have also stressed the potential "statistical intelligence" of the appraisal and emotion generation process (Cushman, 2013; Railton, 2014). That is, while many findings within moral psychology support the idea that automatic, intuition-based emotional reactions guide moral judgments (Greene and Haidt, 2002; Haidt, 2001; Schnall et al., 2008; Wheatley and Haidt, 2005), it has also been argued that these moral emotional reactions are the result of fairly sophisticated implicit statistical rule learning mechanisms (Dwyer et al., 2010; Railton, 2014). Traditional studies of implicit statistical rule learning, for example, have found that participants shown several letter sequences following an "artificial grammar" (a complicated set of rules stating what letters may follow what others) can learn, at above chance levels, to detect when the rules of that

grammar are broken (Pothos, 2007). They do not have conscious access to the grammar's rules (e.g., they could not verbally state them); instead different sequences simply begin to automatically feel right or wrong when they are presented. In a similar manner then, it is suggested that, in the emotional domain, automatic appraisal mechanisms may be highly sensitive to statistical trends within experience that signify when specific situational elements are predictive of the need for a given emotional response (Dwyer et al., 2010; Railton, 2014). Such implicitly learned "rules" about which representations should evoke which emotional responses represent an important example of how unconscious information can be stored "structurally" (in the pattern of synaptic connections between neurons), as was described above; as conscious access requires that information be represented within active, content-bearing neural states, this structural storage format also explains why these rules cannot be consciously reported (Dehaene, 2014; Dehaene et al., 2006).

The fact that these automatic appraisal mechanisms may learn a sort of "implicit emotional grammar," reflecting statistical regularities in the relationship between situational elements and appropriate emotional responses, highlights the potential intelligence of unconsciously generated emotional reactions (Bargh and Morsella, 2008). Studies using the Iowa Gambling task (Bechara et al., 1997; Buelow and Suhr, 2009; Gupta et al., 2011), for example, have found evidence that, by causing automatic unpleasant bodily reactions (or at least internal representations of such reactions, Wiens, 2005), implicit statistical learning mechanisms can lead participants to avoid choosing card decks that tend to result in long-term losses, even when they do not consciously understand why they are avoiding them. More generally, these considerations also highlight the important distinction between (1) processes underlying "emotion generation" (many of which operate outside of awareness) and (2) processes underlying the subsequent perception of the cognitive and bodily reactions generated (and the recognition of their conceptual emotional meaning). Even if certain processes generate an emotional reaction (e.g., the survival circuits discussed by LeDoux, 2012), this does not guarantee that this reaction will subsequently be perceived and recognized appropriately, or that it will enter conscious awareness. In what follows, we will now consider the factors that may contribute to these further stages of emotional processing.

4. Conscious access

The above discussion of automatic appraisal highlights the need to distinguish between conscious and unconscious aspects of emotion. However, to do so requires a theory of consciousness. One leading neuroscientific account of consciousness is the "Global Neuronal Workspace" (GNW) model (Baars, 2005; Dehaene and Naccache, 2001; Dehaene, 2014; Dehaene et al., 2003; Del Cul et al., 2009; Kouider et al., 2007). As we have previously proposed a means of extending this model to conscious/unconscious emotion (Smith and Lane, 2015), we will also draw on the insights of the GNW framework here. According to the GNW model, unconscious perceptual processes operate in parallel, and the vast majority of the information that is represented unconsciously fails to enter consciousness (and hence will not be reportable). As discussed in the previous section, in contrast to "discrete" conscious conclusions, unconsciously represented information within the brain's internal model takes the form of probability distributions (that simultaneously describe the likelihood that several different possible percepts/beliefs about the self and the world may be correct); according to the GNW framework (Dehaene, 2014), when one experiences a single conscious percept/belief, this is the result of top-down control mechanisms that "sample" from one out of the many interpretations that are unconsciously represented – typically the one represented as having the highest likelihood of being correct (Friston, 2005; Hohwy, 2014; Moreno-Bote et al., 2011; Vul and Pashler, 2008; Vul et al., 2009). So while automatic appraisal can evaluate the larger set of possible interpretations that are unconsciously represented (and their estimated probabilities), consciousness can only access one interpretation at a time. This means that emotion could be generated by one of the many parts of the internal model's "description" that have lower represented probabilities, and hence the emotion's cause would remain unconscious and not understood (See Fig. 1).

The GNW model suggests that these many unconsciously represented pieces of information (of the "activated neural state" type discussed above) are each held in locally reverberating neural "buffers" (largely in posterior cortical regions), and that they compete with one another for the ability to be "sampled" and gain access to conscious awareness (Dehaene et al., 2006). When one of these representations gains sufficient strength to outcompete the others (based on factors such as attention, salience, goal-relevance, strength of the perceptual input signal, and estimated probability of accuracy), this initiates a top-down signal from the prefrontal cortex (and associated posterior parietal regions), which amplifies the "winning" representation and allows it to be "globally broadcast" within a broad frontal-parietal network associated with controlled, sequential cognition and goal-directed action selection (Andersen and Cui, 2009; Sackur and Dehaene, 2009; Zylberberg et al., 2011). This global broadcasting function allows the content of the winning representation to be sufficiently "noticed" by these downstream control systems that its content can now be held in mind, sequentially manipulated, and used to inform deliberative action selection. In essence, this frontal-parietal control network is choosing to "bet on" a specific subset of the probabilistic hypotheses held within the larger, unconscious internal model – typically a subset taken to be goal-relevant and to have the highest estimated likelihood of being correct – and using these selected representations to guide the deliberative action selection process. As verbal reports represent one class of deliberative actions that are guided by this process, this explains why selection for global broadcasting is a necessary condition for verbal reportability.

As mentioned above, we have previously applied the GNW framework to the conscious perception/recognition of one's own emotional responses (Smith and Lane, 2015). According to that model, after one's bodily reaction is generated in response to automatic appraisal mechanisms, these autonomic/somatic changes are subsequently detected/represented within various cortical regions that subserve interoception and somatosensation, including primary and secondary somatosensory cortices within the parietal cortex and the left and right insular cortices. Within our model, regions of the lateral anterior temporal lobe (LATL), as well as the rostral anterior cingulate cortex (rACC) and adjacent medial prefrontal cortex (MPFC), are then proposed to assign conceptual significance to these bodily reactions, such that they are represented as signifying one or more emotion concepts (e.g., sadness, fear, anger etc.). Critically, *these sensory/conceptual representations with regard to one's own bodily/emotional state are also a part of the internal model, and will be unconsciously represented in terms of the same probability distributions across possible interpretations discussed above.*

Further, modulatory influences from appraisal mechanisms, as well as other background expectations within one's current context, may also contribute to what emotion concept is "assigned" to these perceived bodily reactions as the most likely interpretation. For example, the same felt bodily reaction might be more likely recognized as "sadness" when at a funeral than when at a birthday party. Finally, the cognitive biases in attention, memory retrieval, and decision-making that are initiated by automatic

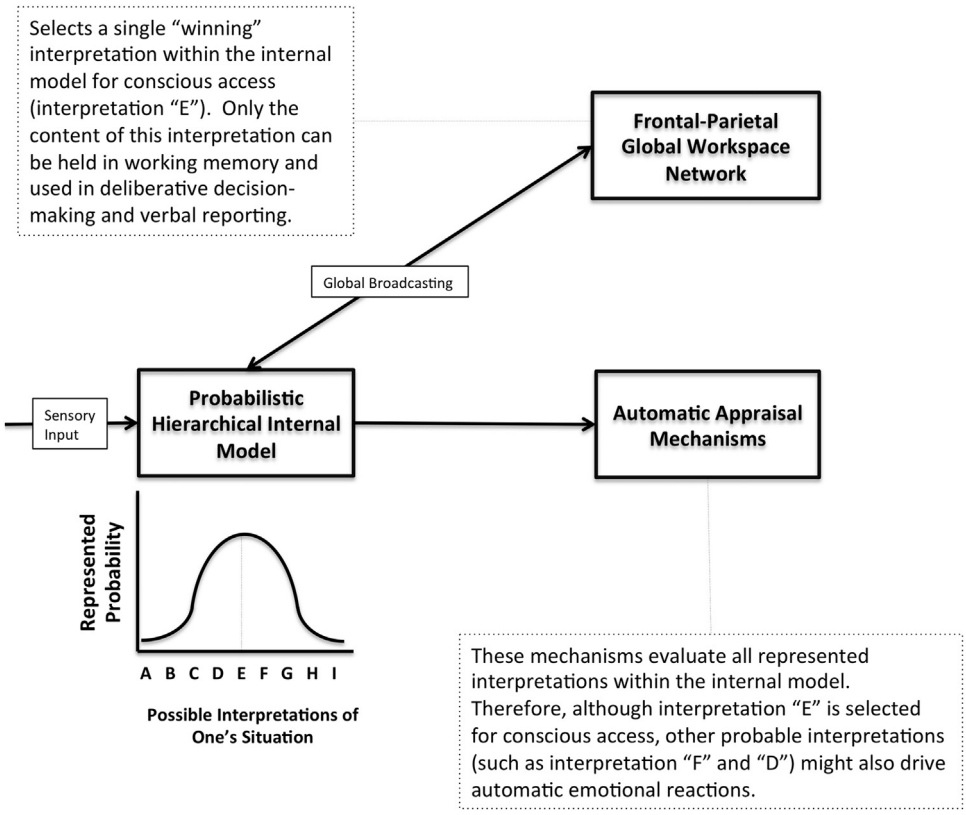

Fig. 1. Conscious vs. Unconscious Processing. This figure illustrates how unconscious probabilistic representations within an internal model can result in both 1) non-probabilistic, "all-or-none" conscious percepts/beliefs and 2) automatic emotional reactions (that may or may not relate to those consciously experienced percepts/beliefs). Note that while, for simplicity, the figure illustrates a single probability distribution over a single space of possible interpretations, the internal model we describe in the text includes a large number of separate (but hierarchically linked) hypothesis spaces – each representing a different type of information (e.g., visual vs. auditory) and/or a different level of description of it (e.g., perceptual vs. conceptual). As also described in the text, only representations within some of these spaces will be selected for conscious access at any given time, based on a range of factors (e.g., salience, goal-relevance, etc.).

appraisal mechanisms will each play a role in determining whether these percept- and concept-level representations of one's own emotional reaction are selected for global broadcasting, and hence whether they are consciously accessible to the individual from moment to moment. For example, while certain psychological states might promote attention to one's own emotions, others might preferentially direct attention outward (Huntsinger, 2013). We have suggested that a medial frontal-parietal network, involving the dorsomedial prefrontal cortex (DMPFC), and the posterior cingulate (PCC) and precuneus, may be primarily involved in the attentional selection and maintenance of emotion-related representations, such that they can be held in mind and used within deliberative cognition.

In summary, the unconscious internal model we have described includes several hierarchically linked hypothesis spaces that describe both one's external and one's internal situation. The parts of the model that pertain to one's external situation will involve probabilities assigned to different possible perceptual/conceptual descriptions of the world outside of one's body and one's position in it (e.g., "I see sand and water; I am on a beach"). The parts of the model that pertain to one's internal situation will involve probabilities assigned to different possible perceptual/conceptual descriptions of the world inside of one's body (e.g., "my face feels warm; I am happy"). Many interpretations/probabilities are represented in each part of the model, and only a small subset of this information is selected for conscious access at any given moment via frontal-parietal mechanisms (discussed further in the following section). However, even if some representations within the internal model remain unconscious, they may still be capable of influencing the generation of future emotional responses, which could lead one to be unaware of these emotional responses, their causes, or both.

5. Cognitive control

The final preliminary topic we will discuss before introducing our proposed taxonomy of mechanisms underlying unconscious emotion is the broad set of processes associated with "cognitive control." Cognitive control processes have been largely linked to distinct regions of the prefrontal cortex, and serve the overarching function of representing goals, and coordinating the interactions between distinct neural systems in a context- and goal-specific manner (Braver, 2012; Gazzaniga et al., 2014, ch. 12). The attentional selection, amplification/maintenance, and deliberation functions discussed in the previous section on conscious access are examples of cognitive control, but many other important mechanisms also fall within this broader category. One mechanistic conception of the nature of these prefrontal functions is the idea that the PFC acts as a "dynamic filter" (Shimamura, 2000). Dynamic filtering theory suggests that the PFC selects different top-down signaling patterns, or "filters," in different contexts, which amplify/maintain the strength of some representations over others. Crucially these PFC "filters" also *suppress competing representations*, such that they are hindered from becoming consciously accessible, and hence prevents them from interfering with conscious, goal-directed cognition (Thompson-Schill et al., 2005, 1999). For example, studies in both the visual and auditory domain have provided evidence that goal-related PFC activation may cause representations of irrelevant or potentially interfering perceptual

stimuli within sensory cortices to drop below baseline activation levels (Druzgal and D'Esposito, 2003; Knight and Grabowecky, 1995). Similarly, we have previously suggested that DMPFC may represent goals of a social/emotional nature; when activated, these goal representations may function to suppress/amplify representations of the thoughts and emotions of self and other that are irrelevant/relevant to those goals, respectively (Lane et al., 2015b; Smith and Lane, 2015).

Recently, the same general result has been extended to declarative memory processes, in order to explain both incidental and motivated forgetting. With regard to explaining incidental forgetting (i.e., accidental forgetting), one evidentially supported mechanism is referred to as "retrieval-induced inhibition." Based on phenomena such as part-set cueing impairment (Anderson, 2003; Chan, 2009), which illustrate that retrieving some items from a list makes it more difficult to retrieve other items from the same list, this explanation suggests (somewhat ironically) that incidental forgetting can result from successfully retrieving related information. That is, cognitive control networks within the PFC appear to automatically suppress the accessibility of memories related to what has been retrieved, in order to minimize future interference (Kuhl et al., 2007). The adaptive assumption underlying this type of automatic dynamic filtering appears to be that since the information currently being retrieved is very likely goal-relevant (and hence "high priority"), the accessibility of related representations (that are currently not in need of retrieval, and hence relatively "low priority") should be inhibited so as to avoid possible future interference with high priority information (see Baddeley et al., 2015). Central to the topic of the present paper, this type of selective suppression of conscious access to low priority information occurs entirely outside of awareness, and since it appears to be based on a type of heuristic assumption about future goal-relevance, it can often frustrate an individual's present goals of retrieving pieces of information stored in long-term semantic or episodic memory.

Also directly relevant to the topic of this paper, recent research has found that people can cause themselves to forget certain information in a "voluntary" or "motivated" fashion as well, through both "thought substitution" and "retrieval suppression" strategies (Anderson and Hanslmayr, 2014; Benoit and Anderson, 2012; Depue, 2012). When presented with a retrieval cue, thought substitution involves the intentional retrieval of a different thought/memory, in order to keep the to-be-suppressed piece of information out of awareness. Retrieval suppression instead involves attempting to more directly keep the to-be-suppressed item out of mind, without replacing it with anything else. Several studies have now shown that, using either method, later recall performance is reduced for voluntarily suppressed items compared to control items that were never retrieved or suppressed (reviewed in Anderson and Hanslmayr, 2014; also see Benoit and Anderson, 2012). While thought substitution activates left prefrontal regions that appear to increase hippocampal activity (i.e., memory retrieval for a competing item), retrieval suppression instead appears to involve reduced hippocampal and sensory cortical activation as a result of increases in right prefrontal regions that are also implicated in top-down behavioral inhibition (Aron and Poldrack, 2006; Aron et al., 2007; Benoit and Anderson, 2012). Interestingly, studies have found similar right prefrontal activations associated with forgetting in psychogenic amnesia (Kikuchi et al., 2010; Tramoni et al., 2009), and other studies have also found evidence that similar types of prefrontal inhibition can be (at least partially) initiated by unconscious stimuli (Hughes et al., 2009; van Gaal et al., 2010). These findings both suggest that information need not be consciously accessible in order to trigger this type of "motivated" suppressive filtering. Finally, recent work has provided evidence that motivated retrieval suppression even reduces the later visual priming effects associated with implicit memory (Gagnepain et al., 2014), suggesting potentially broad-sweeping long-term effects on unconscious processing. It is important to highlight, however, that this work on voluntary retrieval suppression has been the topic of multiple critiques (Kihlstrom, 2002; Schacter, 2001), and it also remains largely unexamined at present whether cues to successfully suppressed emotional memories can trigger emotional responses or related priming effects (but for some recent evidence, see Smith et al., 2016).

The types of cognitive control mechanisms described here are typically understood to be domain general; thus, for example, the same processes invoked in the non-emotional areas of cognition described above are also posited to implement top-down emotion regulation strategies in recent models supported by neuroimaging studies (Buhle et al., 2014). In addition, within recent large-scale neural network models (e.g., Barrett and Satpute, 2013) emotion is also understood to be implemented by a range of domain-general networks. For example, emotion concept representations like "sad" can be understood in similar terms to any other distributed concept representation in semantic memory networks (Pobric et al., 2010; Wilson-Mendenhall et al., 2011). Interoceptive/somatic perception in emotion can also be understood to operate in relevantly similar ways to other cortical sensory systems. Further, studies have found state-dependent changes in conscious access to one's own heart beat and respiration in emotional contexts (Khalsa et al., 2016, 2009). Thus, it is highly plausible that the same top-down control mechanisms discussed above should also act to amplify/suppress conscious access to representations of emotional bodily responses and emotion concept representations. However, to date there have been no studies (of which we are aware) that have directly tested suppression of conscious access to information about one's own emotional state. Thus this also represents a potentially fruitful avenue for future research.

6. Modeling unconscious emotion

Having reviewed the relevant background information, we will now illustrate how interactions between the processes just discussed can provide mechanisms that explain the types of clinically observed cases of unconscious emotion described above. Within the neuro-cognitive framework of the literature introduced above, we suggest that cases of unconscious/unrecognized emotion, such as Walter's unrecognized anger in the introduction, can ultimately arise in two different ways: one due to top-down, motivated factors, and the other due to a bottom-up processing deficit. The "top-down" variant involves mechanisms whereby PFC-controlled dynamic filtering processes may prevent concept-level representations of emotion from being selected for global broadcasting, despite the fact that these representations are both possessed and activated appropriately at an unconscious level. In contrast, the "bottom-up" variant – what we have previously termed "affective agnosia" (Lane et al., 2015b) – can occur if the concept-level representation of one's own emotional reaction is never activated, or if misrecognition otherwise occurs in the absence of suppressive, top-down influences. We will describe mechanisms that could bring about both of these variants below.

Before doing so, however, it will be important to further clarify what is meant by the term "unconscious emotion." In this paper we mainly consider ways in which a person can have a specific emotional reaction without consciously recognizing it (i.e., the "implicit emotion" category described in the introduction). *This clearly implies that there are criteria that determine the identity of an emotional reaction that are independent of conscious recognition.* Several possible criteria could be used. First, a person might be said to have an unconscious emotion because the pattern of automatic appraisals that were unconsciously triggered was specific to that

emotion. According to this, a person might be said to have unconscious anger because the "anger appraisal pattern" was triggered by the contents within their internal model. Secondly, a person might be said to enter a certain emotional state because the specific cognitive/bodily reaction that was triggered is specific to that emotion. According to this, a person might have unconscious fear because the combination of their bodily reaction (e.g., heart racing, trembling) and their current cognitive/behavioral tendencies (e.g., desire to run and hide, attentional biases toward threatening stimuli) are jointly specific to fear. Third, even if a person's actual bodily reaction was not specific to a certain emotion, the brain might unconsciously represent their bodily reaction as if it were specific to that emotion (e.g., the brain might represent an increase in heart rate even if actual heart rate remains slow; Wiens, 2005). Fourth, a person's internal model might unconsciously represent one concept-level emotional interpretation of their bodily/cognitive reaction as most probable (e.g., the brain may represent "anger" as the most probable self-description, but this representation is not selected for conscious access).

It is currently controversial which of these factors can specify the identity of an unconscious emotion. For example, some work suggests that appraisal patterns and/or bodily reactions could be emotion-category specific (Kreibig, 2010; Moors et al., 2013; Scherer, 1997), whereas other work suggests that emotion categories are conceptual categories without specific bodily correlates (Barrett, 2006; Lindquist and Barrett, 2008). It is also possible that emotion concepts, like "sadness" or "anger," have specific ranges of bodily correlates within individuals, but that such patterns do not generalize across individuals (Thayer and Faith, 1994). In this paper, we have simply assumed that factors such as one or more of those described above allow it to be true that a person is in a specific emotional state without consciously recognizing it. States like "unconscious anger" would not be conceptually possible in the absence of such identity-determining criteria. However, we remain neutral on which criteria are correct, and simply highlight this as an important area for future research.

6.1. Top-down recognition failures

We will now illustrate, in detail, how, in a top-down manner, a combination of the neuro-cognitive processes described above, may be capable of accounting for cases like that of Walter detailed in the introduction. We will then generalize from this example case to describe the model elements and their interactions.

First, in Walter's case there was a precipitating event – his wife Martha's death. As a result of this precipitating event, Walter's unconscious internal model of the world – the probabilistic, hierarchical perceptual/conceptual "description" of Walter's situation in life represented across his brain's sensory/memory systems – was updated so as to now include her death and current absence from his life, as well as what that predicted about his future. Second, this updated, internally represented description of his situation was received and automatically evaluated by the hierarchical appraisal mechanisms introduced above. As Walter's internal model likely conceptualized this event as signifying "the loss of someone highly valued," the pattern of reactions across appraisal mechanisms (specifically those within the amygdala and dACC, involving concern-relevance and goal-congruence respectively), and the resulting perceived bodily responses (represented in somatosensory cortex and insula), would likely have combined to activate the representation of the concept "sadness" (within rACC/MPFC and LATL) (Brosch and Sander, 2013; Smith and Lane, 2015). As Walter's sadness was sufficiently salient, and relevant to his current goals, this concept-level representation was selected for global broadcasting, and was therefore verbally reportable, able to be held in working memory, and subsequently able to be used within conscious deliberation when Walter was intentionally deciding how he should plan to navigate through his present and future life. Thus, Walter consciously recognized his sadness.

For the sake of example in the present context, we will assume Walter's internally represented description of Martha's death also triggered a pattern of automatic appraisals consistent with resentment/anger. This assumption is consistent with the fact that he was able to consciously assign the resulting reaction to "anger at life"; however, another represented possible interpretation – "anger at Martha" – did not reach conscious awareness until after therapeutic intervention. The latter conceptualization could be generated if his internal model unconsciously associated Martha's leaving him with the concept of "voluntary abandonment." Perhaps, as is often observed in clinical cases (Mahler et al., 2008), Walter had a history in which abandonment was perceived as intentional and voluntary, in which case he may have also implicitly learned to associate abandonment with blameworthiness (either of the person doing the abandoning or the person abandoned). Such a conclusion could have also been reached in childhood when causal explanations such as "she left because I was bad" are common.

Walter also clearly possessed the concept of anger, and his internal model would have likely unconsciously represented "anger" as one of the possible interpretations of his perceived reaction that had a high likelihood of being correct. Indeed, he even had conscious access to his "anger at life." Yet, unlike Walter's sadness, the conceptualization that he was "angry at Martha" was not consciously accessible; he may also have been unaware of the aspects of his internal model that involved Martha or himself being "responsible" or "to blame" for her leaving him. If, for the sake of example, we take the above description of Walter's case at face value, can the elements of cognitive/affective neuroscience described above aid in understanding it? Here we suggest that a combination of additional unconscious appraisals, and dynamic prefrontal filtering, may be able to account for the elements of such cases (See Fig. 2).

First, based on Walter's own upbringing and learning history, it is plausible to assume he internalized a certain sort of "unconscious emotional grammar" that contained specific implicit statistical rules. As discussed above, these "rules" are embedded in the structural connections between neurons that determine how strongly one representation predicts another (and therefore also determine stimulus-response relationships); hence Walter need not have any conceptual representation of these rules at all (conscious or unconscious). The implicit rule mentioned in the previous paragraph that "abandonment predicts blameworthiness" may be one example. With regard to the appraisal dimension of "compatibility with norms/values," another important implicit rule might be something like the following:

> *When someone blames or becomes angry with others for tragic events outside of their own control, especially due to their own resulting losses that are trivial in comparison,* this predicts negative evaluations, and loss of social support, from those who become aware of it.

Essentially, the events that Walter has witnessed throughout his life would have combined so as to adjust the synaptic connections between representations within his internal model to the point where a combination of the (italicized) features on the input end of the above stated rule is structurally represented as a statistically significant predictor of the (underlined) consequences. Recent computational models of cortical function and learning suggest the brain is fully capable of abstracting this sort of implicit rule from trends in past experience (reviewed in Hohwy, 2014). Further, studies have shown that accessing information regarding one's personal values has been linked to medial prefrontal and dorsal striatal activation (Brosch et al., 2012), whereas knowledge of

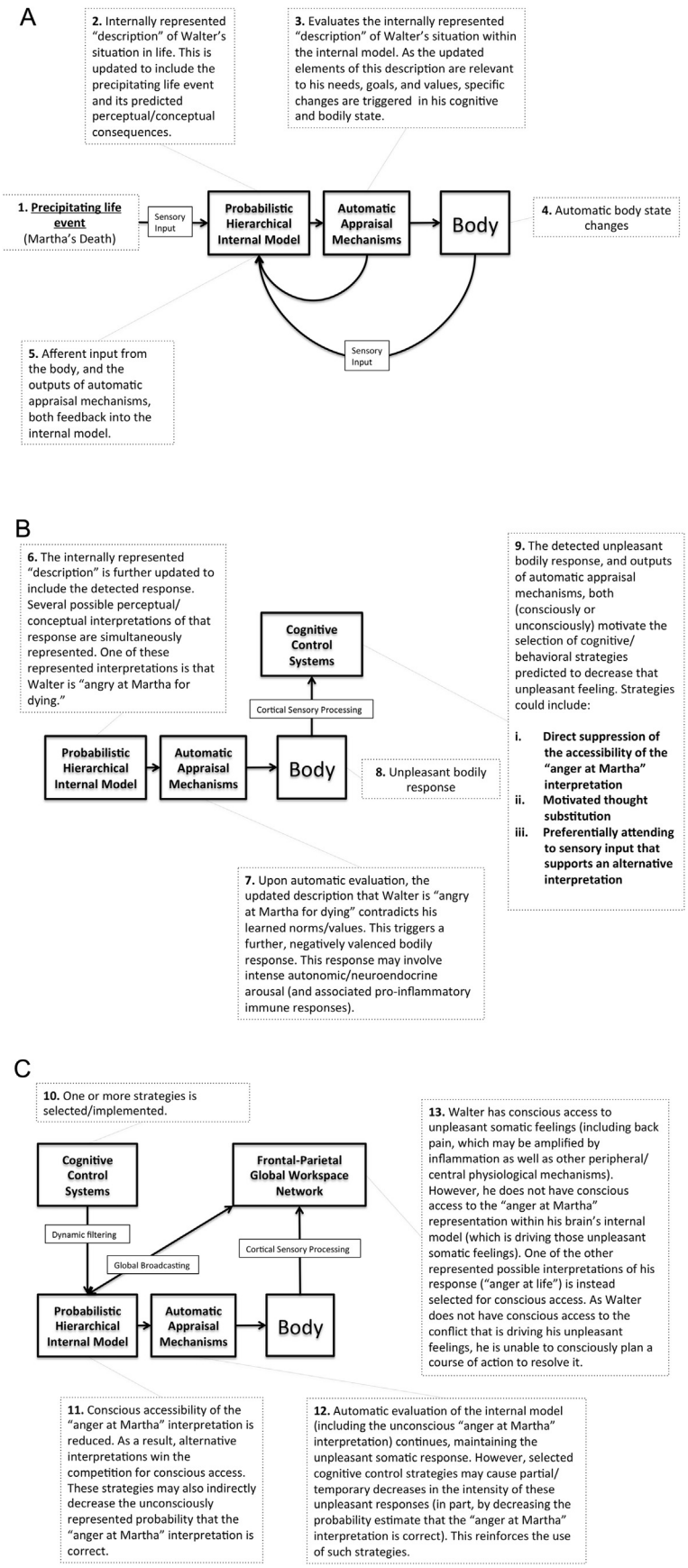

Fig. 2. Top-down Mechanisms. (A-C) In step-by-step fashion, using the example case of Walter (described in the text) to provide a concrete illustration, this figure illustrates how the top-down mechanisms we describe can be initiated, and how they can keep an individual unaware of their own emotional state.

social norms may be represented (at least in part) within superior anterior temporal regions (Zahn et al., 2007). Other work has also shown that top-down control mechanisms within DLPFC may be required when social norms disagree with one's personal values (Knoch et al., 2006).

We suggest therefore, that when Walter's internal model is updated to include the unconscious representations of "Martha is to blame" and "anger at Martha" as possible interpretations, this would initiate a further round of automatic appraisal. Since the unconscious description of the possibility that "I am angry with Martha for dying" likely contradicted Walter's norms/values, the relevant appraisal mechanisms described above would trigger a further negatively valenced response. This response would be perceived as an unpleasant bodily feeling (represented in part within the anterior insula), and would also trigger cognitive biases and motivate the selection of cognitive/behavioral strategies designed to diminish this negative response, at least partially through interactions with regions of dACC (Medford and Critchley, 2010). This is the same sort of unpleasant feeling that may motivate avoidant decision-making (in the absence of conscious understanding) within the Iowa Gambling task discussed above (Bechara et al., 1997; Buelow and Suhr, 2009; Gupta et al., 2011). We suggest that once this motivation is in place, multiple dynamic filtering mechanisms could plausibly be triggered in order to suppress the accessibility of Walter's unconscious representations of "Martha is to blame" and "anger at Martha" as possibilities. We will now discuss each of the ways this might happen, but the end result of each will be that, despite the fact that such representations may remain present/active within Walter's internal model (i.e., within the reverberating neural buffers described by the GNW framework), they will fail to win out in the competition for global broadcasting and remain consciously inaccessible (Dehaene et al., 2006).

6.1.1. Dynamic filtering as a result of motivated retrieval biases

One way in which this might occur is through a mechanism akin to retrieval-induced forgetting. For example, perhaps Walter could have initially been briefly aware of the possible interpretation that he was angry with Martha. If so, however, the automatic negative reaction to this possible interpretation would have quickly motivated the search for other interpretations. The competing interpretation of "anger at life" would be one example. When Walter voluntarily and repeatedly retrieved (or "sampled") these related, alternative interpretations from his probabilistic internal model, this could cause retrieval-induced forgetting mechanisms to label the "anger" interpretation as "lower priority," and suppress its future accessibility; doing so might also be reinforced by the resulting reductions in the intensity of his unpleasant response. Given that what are being suppressed/retrieved are concept-level interpretations of Walter's reaction, this is fairly similar to standard semantic memory retrieval processes. As described, this is also very similar to the work discussed above that has illustrated reduced conscious accessibility of specific memories as a result of "thought substitution" (Benoit and Anderson, 2012). That is, Walter might be substituting the thought "I am angry at life" in order to avoid accessing the thought that he could be angry with Martha. Alternatively, such a negatively valenced response to the possibility of his anger at Martha might also motivate Walter's use of right DLPFC-mediated retrieval suppression mechanisms (Anderson and Hanslmayr, 2014).

Both of these scenarios would result in long-term suppression of the "anger at Martha" representation by PFC-mediated dynamic filtering mechanisms, putting the representation of this possible interpretation at a distinct disadvantage in the competition for selection by global broadcasting mechanisms. Further, while we have suggested the possibility that Walter may have at least initially consciously entertained the possibility that he was angry at Martha, the work discussed above on psychogenic amnesia (Kikuchi et al., 2010; Tramoni et al., 2009) and unconsciously triggered inhibition (Hughes et al., 2009; van Gaal et al., 2010) both suggest that this type of motivated suppression could be activated, even if the "anger at Martha" interpretation was never consciously accessed. That is, even if the "anger at life" interpretation was represented with a higher probability right from the start, the unconscious (lower probability) interpretation of "anger at Martha" may still have been strong enough to drive appraisal mechanisms to trigger an unpleasant bodily response.

At a minimum, patients such as Walter might be motivated to reduce the resulting negatively valenced bodily response through means of trial-and-error, even if no explicit knowledge of the reason for that negative reaction were accessible. In other words, even if such patients do not understand why they feel better when they happen to retrieve one thought as opposed to another, if they found that retrieving certain thoughts always reduced their unpleasant feelings more than others, such thought patterns would very likely increase in frequency. In Walter's case, as retrieving and focusing on the "anger at life" interpretation would result in suppression of the strength of the competing unconscious "anger at Martha" interpretation, this could provide temporary reductions in his negative affect without him understanding why. This is based on the assumption that this suppression of the strength of the "anger at Martha" interpretation would also decrease the strength of the unpleasant response it was driving appraisal mechanisms to generate.

6.1.2. Dynamic filtering as a result of motivated attentional biases

A third top-down mechanism whereby conscious accessibility could be reduced to Walter's unconscious representation of the "anger at Martha" interpretation is through voluntary attention. Within the brain, attentional modulation appears to represent a mechanism whereby certain streams of information can be weighted more heavily than others in their ability to update one's internal model (Feldman and Friston, 2010). If Walter preferentially and repeatedly attended to information that supported the conclusion that he was *not* angry with his wife Martha, and supported the conclusion that he was "angry at life" instead, this would therefore have two effects. First, it would have the effect of shifting the probability distributions within his internal model in a biased manner, making the "anger at Martha" interpretation decrease in its represented probability of being accurate (and it would also make the represented probability of the "anger at life" interpretation increase). Second, as a result Walter would find that this tended to cause reductions in his negative affect, because the resulting shift in represented probabilities would also reduce how intensely automatic appraisal mechanisms reacted to the possibility of "anger at Martha." These reductions in negative affect would therefore reinforce Walter's "avoidant" attentional bias.

As the global workspace network appears to preferentially sample from "high probability" interpretations (Moreno-Bote et al., 2011; Vul and Pashler, 2008; Vul et al., 2009), this would also decrease the chances of the "anger at Martha" representation being selected for global broadcasting. Perhaps one of the functions of a therapist raising the possibility of "anger at Martha" is therefore to draw attention to the plausibility of the thought, thus shifting the represented probability distributions in a way that would increase the chances that this interpretation would be selected for global broadcasting. As we will discuss more below, however, guiding attention in this way would need to be done with caution, because the intensity of arousal generated by the "anger at Martha" interpretation (or any other "impermissible" interpretation) would increase as that interpretation became represented as more and more probable. As we discuss below in more detail,

if arousal becomes too high, medial prefrontal regions involved in mentalization and concept-level emotion representation may become deactivated, hindering a person from accurately recognizing their emotional state (Lane et al., 2015a,b). Thus a therapist may need to foster a calm, soothing context, or actively promote use of calming self-regulation interventions, to counteract these increases in arousal in order to facilitate accurate conscious emotion recognition. The value that a client places on a therapist's intervention such as this might well be a function of the strength of the therapeutic alliance and the level of trust established in this relationship.

6.1.3. Top-down model summary

To now generalize from Walter's example case, our summarized model of top-down, motivated unconscious/unrecognized emotion includes the following. First, our model makes use of evidence suggesting that the brain maintains a probabilistic, hierarchical internal model with a description of the world and our position within it (Friston, 2010, 2005; Hohwy, 2014), and suggests that this probabilistic description is iteratively evaluated by several interacting automatic appraisal mechanisms (Brosch and Sander, 2013). Second, it suggests that when these appraisal mechanisms trigger a cognitive/bodily emotional response, this response is subsequently unconsciously processed within multiple cortical regions (including the insula, rACC/MPFC, and LATL) *and used to update one's internal model further* (this updating likely occurs across many cortical regions, via iterative, hierarchical error-minimzation processes; see Friston, 2005; Gu et al., 2013; Seth, 2013). However, when this updated internal model includes a description of one's self that is incompatible with one's own internalized norms/values, this can trigger another round of automatic appraisal, a further unpleasant bodily feeling, and a related motivation to select a cognitive/behavioral strategy that is predicted to minimize the intensity of that unpleasant feeling. Once a strategy is found that causes partial reductions in the intensity of this unpleasant feeling, its continued use can become habitual via reinforcement learning. Crucially, while these strategies may keep an unconsciously represented interpretation from being consciously accessed, and also reduce how probable it is represented to be, they will typically not stop the influence that such a representation has on automatic appraisal mechanisms. One reason for this is that a person will have to continually avoid evidence supporting the avoided interpretation, and thus its unconsciously represented probability will be unlikely to drop to negligible levels. Hence, chronic, misunderstood emotion can continue to be generated – including the associated bodily autonomic/endocrine/immune system responses that may amplify pain and explain other related somatic/health complaints (reviewed in Irwin and Cole, 2011; Schultze-Florey et al., 2012; Slavich and Irwin, 2014).

Our model remains neutral about whether one initially has some conscious access to such "impermissible" interpretations of their emotional responses, or whether these aspects of one's updated model, and the resulting motivational influence, remain fully unconscious; both appear plausible, and may be variably applicable to different individual cases. However, once this (conscious or unconscious) motivation is in place, we suggest that multiple related mechanisms could then prevent one's "impermissible," unconsciously represented interpretation of their emotional reaction from being selected for global broadcasting within GNW frontal-parietal networks. First, the strategies of motivated thought substitution or direct retrieval suppression could be reinforced, and the repeated use of either would result in long-term suppression of the accessibility of the impermissible interpretation, through PFC-mediated dynamic filtering mechanisms. Second, a strategy involving the motivated use of selective attention could be reinforced; if so, this would cause a biased updating of the probability distributions within one's internal model, leading to lower represented probability estimates that the impermissible interpretation was correct. This would also decrease the likelihood that GNW frontal-parietal networks would "sample" (i.e., consciously access) this unconsciously represented interpretation.

6.2. Bottom-up, perceptual emotion recognition failures

In contrast to the top-down mechanisms described above, the bottom-up variant of unconscious/unrecognized emotion can occur if either (1) an individual, as a result of either brain damage or unhealthy childhood developmental/learning processes, does not possess concept-level emotion representations (anger, fear, guilt etc.) relevant to the current situation, or if (2), despite possessing such concepts, these representations are not activated appropriately in response to perceiving the cognitive/bodily reactions triggered by automatic appraisal mechanisms (Lane et al., 2015b). While this second variant could be due to stable connectivity impairments between specific brain regions (such as the anterior insula and rACC/MPFC), it could also occur in a state-dependent manner during periods of very high arousal, due to the fact that high arousal causes inhibition of activity within the rACC/MPFC regions implicated in emotion concept representation (Lane et al., 2015a; Thayer et al., 2012; see also Arnsten and Robbins, 2002; Arnsten, 1998; Robbins and Arnsten, 2009). Thus, if automatic appraisal mechanisms trigger a sufficiently strong, high arousal emotional reaction (perhaps as in the case of the "unspeakable dilemmas" mentioned above), the resulting inhibition of rACC/MPFC could plausibly result in a decreased ability to recognize the conceptual emotional meaning of one's current state (See Fig. 3). When this high arousal state is transient, a person might be able to recognize/infer his or her own emotions after the fact (e.g., after calming down and reflecting). However, if a description involving stable threats to one's needs, goals, and values persisted within one's internal model, automatic appraisal mechanisms could, as a result, maintain a person in a state of chronically high arousal, resulting in a potentially chronic deficit in recognizing one's own emotions. This is one way in which unconsciously represented dilemmas could promote high arousal, and decrease both unconscious recognition of one's own emotional responses and conscious understanding of its underlying causes.

Based on our hypothesis described above – that the degree to which an emotional reaction is generated may be proportional to the estimated probability of the represented interpretation driving it – there is also an interesting dynamic that could arise. That is, as an individual comes progressively closer to gaining conscious awareness of an "impermissible" interpretation of their emotional state (within psychotherapy, for example), this could be understood to involve that interpretation coming to be represented with a higher and higher unconscious probability estimate (within rACC/MPFC, at least in part). Typically, this would culminate in that interpretation winning the competition for conscious access. However, as that interpretation came to be represented as more and more probable, it would also drive more and more intense arousal. If arousal became too intense, rACC/MPFC may be inhibited, preventing this emotion-interpretation process from continuing. Intense unpleasant arousal may therefore prevent conscious emotion recognition in a bottom-up manner; this suggests that therapeutic interventions may need to involve the creation of a safe, comforting context to counteract this type of intense arousal, such that this type of emotion recognition process can proceed to completion without the rACC/MPFC being deactivated.

Thus, while dynamic suppressive filtering as a result of the mechanisms described in the previous section (thought substitution, retrieval suppression, and motivated attention) represents one means of suppressing emotional awareness, there are also bottom-up mechanisms for doing so (Lane et al., 2015b). In Walter's

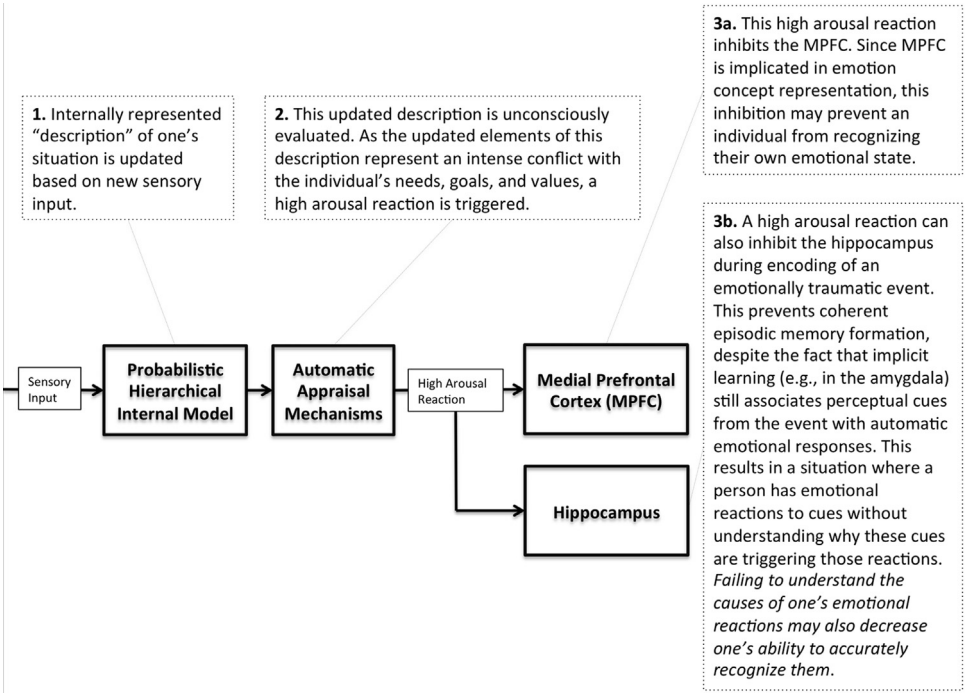

Fig. 3. Bottom-up Mechanisms. This figure illustrates two possible ways in which a bottom-up processing deficit, due to high arousal, can also keep an individual unaware of their own emotional state.

case, for example, if his negatively valenced reaction (resulting from the unconscious appraisal of norm/value incompatibility) involved sufficiently high arousal, this would result in an inhibition of the rACC/MPFC regions implicated in emotion concept representation. This inhibition could prevent both the formation and accessibility of conceptual interpretations of his emotional responses in a state-dependent manner, and if this high arousal remained chronic (due to unresolved dilemmas within his internal model), then Walter's emotions could also remain chronically unrecognized. Thus, although Walter did recognize both his anger at life and his sadness, the appraisal leading to "anger at Martha" could conceivably have induced a higher arousal response than the other two. Perhaps as the possibility of "anger at Martha" began to be unconsciously assessed, the induced arousal prevented this interpretation from ever reaching conscious awareness because the structures required for such concept-level functioning became inactivated.

As we have described elsewhere in detail (Lane et al., 2015b), this mechanism can also result in a sort of "short circuit" or "positive feedback loop," in which one's perceived (but incorrectly understood) bodily reactions are interpreted as dangerous/threatening, and hence drive even stronger further reactions from automatic appraisal mechanisms. As this cycle of "automatic appraisal-threatening bodily reaction-intensified automatic appraisal-intensified bodily reaction" continues, arousal (and related peripheral physiological responses) can be driven to intensely unpleasant levels (and fully capture one's attention – potentially amplifying them even further). The resulting bodily sensations, devoid of understood emotional meaning, can themselves become the object of concern and be a reason for seeking evaluation for a medical problem. In fact, this is exactly what happened in Walter's case as he sought clinical treatment for back pain. In contrast, when a healthy individual appropriately recognizes their bodily reactions as being emotional in nature, this can "quiet down" the reactivity of automatic appraisal mechanisms, due to the fact that the possibilities of physical danger or the threat of an undiagnosed systemic disease process are no longer estimated as likely. As one understands their emotional origin, one will typically also have a better sense of what one needs to do in order to control/adjust such reactions, which would also cease to drive appraisals of "lack of control." For example, once one understands that one's stomach pain is related to sadness, one can look for, and attempt to adjust, aspects of one's life that are common causes of sadness; whereas without that understanding one is left with unpleasant bodily symptoms and no understanding of how to deal with them except perhaps to seek medical care.

A different type of bottom-up emotion recognition failure appeals more directly to implicit statistical learning mechanisms. Specifically, we suggest that learned expectations within particular contexts could also prevent appropriate bottom-up recognition processes. In the visual domain, for example, it has been shown that congruent visual contexts facilitate object recognition, likely because context representations predict the presence of some objects and not others (Bar, 2004). Thus, it would be easier to recognize a cactus in the desert than in a rainforest. Similarly many studies suggest some contexts may facilitate or inhibit emotion recognition (reviewed in Barrett et al., 2011). With regard to Walter's case, we suggest that the context of "Martha's death" might be structurally represented as incongruent with him being angry with her. This might be the case, for example, if he and Martha had a loving relationship, and if he associated death and loss of the relationship as involuntary. If so, when the "Martha's death" representation was activated, it would inhibit the "anger at Martha" representation. This would decrease the represented probability that the "anger at Martha" interpretation was correct, making it harder for Walter to recognize it. As discussed above, however, Martha's absence from his life may also simultaneously predict abandonment for Walter, which unconsciously triggers his bodily anger response. Therefore two different associations with Martha's death might simultaneously trigger Walter's anger and prevent its recognition. Further, these associations would only be present in the pattern of structural connections between the representations in Walter's brain (as a result of his past experiences), and therefore he may not have any concept-level representation of them at all (conscious or unconscious).

In these circumstances, a therapist's approach to helping Walter gain conscious access to his anger at Martha may involve a bottom-up approach. Consistent with Gendlin's method of "Focusing" (Gendlin, 1982), for example, it may involve drawing Walter's attention to the patterns in his automatic emotional/bodily responses, so that he could figure out (i.e., conceptualize for the first time) what the implicit "rules" are that underlie them and what emotions the bodily sensations reflect. Once Walter gains a conceptual representation of the fact that he associates her absence with abandonment, for example, the rest of his internal model could be updated to reflect that; this would allow him to more accurately interpret both his past and present behavior. If he also gained a conceptual representation of the implicit "rule" that "anger at Martha" was inconsistent with her death and their loving relationship, a similar beneficial result might follow. That is, his internal model could be updated such that his anger was no longer represented as inconsistent with his love for her, and this would facilitate his ability to fully experience/recognize his anger. Overall therefore, coming to understand the implicit rules causing one's emotional reactions can help provide disambiguating information, making it more likely that one will interpret those reactions correctly. This also might relate to previous work on a type of dissociation (Bucci, 2016) in which hippocampal deactivation during high arousal can lead one to have strong emotional reactions to perceptual cues without forming the episodic memory that would allow one to understand why one has such reactions (Nadel and Jacobs, 1998; but see Kihlstrom, 2006). Here we highlight the fact that failing to understand why one has such reactions will also make it harder to recognize their correct emotional meaning. One will be much more likely to correctly understand their emotional reactions if they can correctly identify what is causing them and why. Thus the two categories of unconscious emotion discussed in the introduction – unconsciously generated emotion and implicit emotion – can plausibly interact with each other in important ways.

In summary, we suggest that bottom-up emotion recognition failures can also involve multiple mechanisms. First, chronically high arousal, due to persistent conflicts represented within one's own internal model, can inhibit the neural systems associated with concept-level emotion representation. This can result in a global inability to represent/recognize the identity of one's own emotions. If this were the case during development, one may also not appropriately acquire emotion concept representations to begin with, or one may inappropriately link them to one's perceived bodily states. This type of severe, global inability to recognize/understand emotions corresponds well to the construct of alexithymia, or what we have more recently referred to as affective agnosia (Lane et al., 2015b; Taylor, 2000). Second, one's learning history can lead to specific expectations embedded into the structural connections within one's internal model, and these implicit expectations can also hinder emotion recognition in particular contexts. Particularly, when one has implicitly learned that a certain situation is inconsistent with a certain emotional response, this can make it difficult to recognize that emotional response. Finally, sometimes emotion recognition requires that one recognize the implicit "rules" underlying automatic emotional/bodily responses. In such cases, one is attending to, and conceptualizing these implicit associations for the first time, and doing so can update one's internal model in ways that facilitate accurate recognition and awareness of one's own emotions.

7. Discussion

7.1. Limitations and opportunities

The model we propose here offers a broad and nuanced account of a range of mechanisms related to unconscious emotion. Our model is consistent with a large body of work on unconscious social cognition (reviewed in Bargh and Morsella, 2008), which suggests that sensory stimuli, whether consciously perceived or not, can trigger unconscious evaluative processes – and that these evaluative processes can influence motivation and action selection outside of awareness.

When considering potential implications of the present model, however, it is important to first highlight its current limitations, which all stem from a need for more direct evidential support. While we have appealed to many related cognitive/computational theories of brain function that are meant to be domain general, and which enjoy considerable support, their specific application to emotion remains insufficiently explored. For example, while the notions that the brain maintains an internal probabilistic model and that conscious access is a function of selective sampling/global broadcasting mechanisms have both received considerable evidential support (Bastos et al., 2012; Dehaene and Naccache, 2001; Dehaene, 2014; Del Cul et al., 2009; Friston, 2010, 2005; Hohwy, 2014; Moreno-Bote et al., 2011; Vul and Pashler, 2008; Vul et al., 2009), most of this work has involved exteroceptive sensory systems, and its applications to interoception/emotion are just beginning to receive exploration (Gu et al., 2013; Seth and Critchley, 2013; Seth, 2013; Smith and Lane, 2015). Further, much of what the model proposes regarding PFC-mediated dynamic filtering, and motivated suppression, rely heavily on an extrapolation from research in both declarative memory and visual/auditory perception. Finally, considerable research is still required to better understand the exact nature of, and interactions between, the various appraisal mechanisms we have discussed (Brosch and Sander, 2013; Moors et al., 2013). We have suggested, for example, that appraisal mechanisms can respond to the unconsciously represented probabilities of multiple interpretations of one's situation, and that the strength of an emotional reaction may be directly related to the probabilities represented. However, how strong such unconscious representations need to be in order to trigger detectable emotional reactions remains largely unexplored, and future research should address this interesting topic. Despite these limitations, as there is no inherent reason why the theories we appeal to should not apply to either interoceptive perception or emotion concept representations (and their deployment in the interpretation of one's own emotional responses), an important reason for presenting the proposed model is to stimulate further research to determine whether in fact this extrapolation to emotion is justified. At present, the absence of evidence in support of such an extrapolation is due to the fact that the issues have not yet (to our knowledge) been addressed. Indeed, without articulating the cognitive neuroscientific implications for unconscious emotion as we have, the likelihood that such associations would be discovered spontaneously through empirical studies alone is low.

Some authors (Solms and Panksepp, 2012) have argued that emotions are different from other cognitive processes and are always conscious, whereas one of us (RDL) and his colleagues have argued elsewhere that the foundational distinction in cognitive neuroscience between implicit and explicit processes applies to emotion as well (Lane et al., 2000). Moreover, consistent with LeDoux (2012) and others (Kihlstrom et al., 2000), the latter perspective holds that emotional responses begin as implicit/automatic responses and that explicit/conscious representations of those responses may or may not be added. The mechanisms outlined above can be viewed as an elaboration of how implicit emotion processing works and is maintained, as well as what is needed to make the transition to explicit conscious awareness. In the framework we have described, implicit/unconscious emotion is actually related to multiple processes. First, it is linked with unconscious sensory processing, which updates the multiple probabilistic interpretations that are held within the internal

model and that describe one's situation in perceptual and conceptual terms. Second, it involves the automatic appraisal processes that ultimately generate an emotional reaction based on the internal model's description, where this reaction can include cognitive, visceromotor, somatomotor and behavioral expressions of emotion in the absence of a consciously perceived and/or consciously understood feeling. Third, it includes the further unconscious perceptual processing of one's bodily emotional reaction (that updates the internal model further, and contributes to its description of the self).

To our knowledge this paper is the first in which the concept of an internal model has been applied to the phenomenon of unconscious emotion, and the first in which this internal model has been explicitly related to neural mechanisms associated with the dimensional appraisal process preceding the generation of an emotional response. Here it is important to highlight that both the unconscious internal model's description of the self/world, and the way it is automatically appraised, play an important role in unconscious emotion processing. For example, it appears plausible that, in evaluating the emotional significance of one's unconsciously represented situation, appraisal mechanisms may draw on a complex set of emotion schemas within the internal model that have been learned in development. These schemas would include predictive information regarding which sorts of actions typically occur in which situations, and which features of one's situation typically predict things like threat, loss, disappointment, and so forth. Thus, the same description of one's present situation may provoke different appraisals (and hence different emotional reactions) in different people, *because different learning histories – the effects of which are often only implicit – may lead people to associate the same situation with different predicted outcomes.* These different emotional reactions would subsequently also motivate different decisions/behaviors.

Thus, one opportunity offered by our model is that it highlights potential avenues for further exploration of the way the mechanisms we have described may interact with development to explain cases of emotional pathology. For example, if an abusive or distressing situation exists that is recurrent and unavoidable, a child may come to learn the circumstances in which such adverse experiences are likely to occur; the child may subsequently make cognitive and behavioral adjustments to increase predictability and minimize distress in such contexts. Repeated experiences of having one's optimistic expectations dashed, for example, could lead to a reduction of optimism and an increase in pessimism; that is, the interpretation that "bad things will continue to happen" would become represented as more and more probable. Further, if one expects bad things to happen, distress may be lessened when they occur (i.e., relative to those with optimistic expectations). Such adjustments could be highly adaptive in childhood. In fact, one of the major functions of such implicit learning may be to navigate life while minimizing emotional distress (and maximizing predictability). Importantly, however, in adulthood these adjustments may nonetheless act as a basis for recurring difficulties associated with inaccurately interpreting and responding to the world. Essentially, statistical regularities learned in a traumatic childhood environment will no longer be true in one's adult life, and thus the automatic emotional reactions that "assume" those regularities are true will promote maladaptive responses. They may also be very difficult to unlearn because they were learned through statistical regularities that became deeply ingrained through repetition, and because unlearning may often require gaining explicit understanding of the fact that this implicit learning has occurred, as well as why it has occurred.

Another way development may interact with the model we have described is via interactions between attention and learning. To see how, consider that learning efficiency in a given domain is modulated by attention (Feldman and Friston, 2010) such that one learns more efficiently from a given input signal if it is attended to. This is important in the context of the abusive/distressing childhood environments discussed above because, in such highly uncertain/threatening environments, one's attention will plausibly be directed exteroceptively the majority of the time (i.e., one will learn to constantly monitor for possible threats, including signs of threat in the emotional expressions of others; Paivio and Laurent, 2001). This means that one will attend relatively less to their own internal states, and therefore have less opportunity to learn about the patterns in their own interoceptive reactions. As a result, a person would also have less opportunity to learn to understand their own emotions, because emotion concepts themselves refer to patterns in one's bodily reactions and how those reactions covary with changes in the external world (Smith and Lane, 2015). Therefore, just as patterns of habitual attention can maintain a lack of emotional awareness in adulthood (i.e., what we have called "top-down attentional mechanisms in Section 6), similar reinforced attentional biases can lead to reduced emotional learning in childhood – potentially leading to the more 'bottom-up' emotion recognition problems in adulthood associated with affective agnosia (Lane et al., 2015b). This can clearly also further add to the recurring difficulties in later adulthood described in the previous paragraph.

The recurring patterns of attention and behavior described above, which lead to these recurring difficulties, can be more generally understood as expressions of what a given person's internal model has learned to predict in various situations based on past experience, and the appraisals and emotional reactions that result. Importantly, such learning processes can occur without a person consciously understanding why they are having the emotional reactions that they are having, and, as described above, this could also lead one to misidentify their emotions in such situations. Thus, one important point of this article is that not all of the mechanisms leading to unconscious or unrecognized emotion need involve motivated factors. Sometimes, a person may fail to recognize their own emotions because they have not yet (consciously or unconsciously) conceptualized the reason they are having those reactions. Further, they might fail to recognize their emotions because, based on past experience, such emotions are not expected within specific contexts. As discussed more below, some mechanisms for keeping emotions unconscious may involve bottom-up processing deficits (as opposed to top-down factors).

In contrast to the different unconscious/implicit processes discussed above, we have also described how conscious processing of emotion involves the selection of specific percept- and concept-level representations of one's emotional state for global broadcasting within the GNW frontal-parietal network. When this occurs, the relevant representations can be held in working memory, manipulated, combined with other information also held in working memory in novel ways, and used to guide deliberative decision-making processes (Dehaene and Sigman, 2012; Sackur and Dehaene, 2009; Zylberberg et al., 2011, 2010). We have also highlighted the fact that emotional experience includes multiple dissociable components, and described how representations of some of these components can be conscious while others are not in a given instance. For example, one might consciously experience a fear-related bodily reaction and associated action tendencies, even if one is not consciously aware that they are afraid. Much of what we have said in this paper has focused on situations in which one experiences an automatic emotional reaction, but fails to become aware of the concept-level emotional meaning of that reaction. Thus, our model helps clarify why, when thinking about unconscious emotion, it is important to distinguish between 1) conscious access to emotion-related bodily feelings and action tendencies, and 2) conscious access to the emotional meaning of those bodily feelings and action tendencies.

That being said, it should also be highlighted that recognizing the meaning of one's reaction can also indirectly cause changes in how that reaction feels. This can happen in at least two ways. First, to minimize error within a hierarchical internal model, the way that bodily feelings are represented can be adjusted so as to be more consistent with the winning emotion concept interpretation. Thus, once a person recognizes their reaction as one of "anger," for example, one's felt bodily reaction might be adjusted so as to feel closer to a prototypical anger feeling (e.g., having recognized that one is feeling angry, one may then be more likely to feel the most highly expected bodily expression of that feeling). Second, as described above, correctly recognizing one's emotional reaction can also decrease the intensity of that reaction. This is because, in the absence of correct recognition, it can trigger appraisals of danger and lack of control (e.g., it might be interpreted as signs of a heart attack), and promote a further more intense reaction. Thus, while conscious access to represented bodily reactions and their emotional meaning can be dissociated from one another, these recognition- and perception-related processes can also involve important bi-directional interactions.

7.2. Clinical implications

One of the stated aims of this paper was to explicate in cognitive neuroscientific terms how emotion can *remain* unconscious, as plausibly occurs in some clinical contexts. With respect to this aim, the present model appears to offer important potential insights regarding how clinicians understand the phenomenon of unconscious/unrecognized emotion in their patients/clients. One major insight relates to the fact that, in the present framework, there are actually several distinct mechanisms that can lead a patient/client to fail to consciously understand or experience specific aspects of their own emotional reactions. Some of these mechanisms bear some similarity to the psychoanalytic concept of repression, whereas others do not. Further, even those mechanisms that do appear similar to repression have a significant amount of added nuance and complexity.

To illustrate, unlike the more recent probabilistic conceptions of unconscious representation that we have appealed to, repression has often been thought to involve an emotion being represented unconsciously as a discretely recognized reaction (e.g., "I am angry"); in classic repression, this discrete representation would also only remain unconscious because of other unconscious motivations to avoid the pain of becoming aware of that fact. In contrast to this previous conception, in our framework the unconscious mind does not represent single, discrete conclusions. Instead, the unconscious represents multiple possible interpretations simultaneously, along with their probability of each being correct. These unconsciously represented probability distributions across interpretations are also constantly adjusted based on incoming sensory input, and the degree to which any stream of sensory input is able to update this internal model is weighted by attentional mechanisms. One important insight, therefore, is that the traditional question "was the undesired emotion recognized unconsciously or not?" actually assumes a false dichotomy. That is, as many interpretations of one's emotional state are unconsciously represented simultaneously, a better question would be "was the undesired emotion unconsciously represented as a high probability interpretation or not?" In most situations, at least some information from memory and/or sensory input will be consistent with many emotional interpretations, and hence it is unlikely that the undesired interpretation will be represented with a probability of zero. Therefore, the single, discrete thought or feeling experienced consciously does not provide a good model of unconscious representation.

The closest thing to the classic notion of repression within the framework we have appealed to would involve the undesired interpretation being unconsciously represented as having the highest probability (in comparison to other interpretations within the relevant hypothesis space), and yet still not winning the competition for conscious access. However, it is worth highlighting that even the top-down mechanisms we have detailed above may not be best described in this way. That is, while the undesired interpretation could start out having the highest probability, this does not need to be the case in order to trigger the resulting mechanisms. It only needs to be represented as sufficiently probable to drive appraisal mechanisms to generate detectable unpleasant arousal. Further, regardless of how probable it is originally estimated to be, the top-down mechanisms we have described can function to shift the unconsciously represented probabilities such that the undesired interpretation is represented as less and less probable over time. For example, the more one attends only to information consistent with a desired interpretation, and avoids attending to information consistent with the undesired interpretation, the lower the probability estimate will become that the undesired interpretation is correct. *Yet, if considerable evidence is present within a person's environment to support the undesired interpretation, they might end up vigilantly avoiding various aspects of their life in order to prevent the increases in unpleasant arousal that would result*. This perspective is highly consistent with the vigilance-avoidance theory of repressive coping (Derakshan et al., 2007). Thus, while some aspects of the top-down mechanisms we have described are superficially similar to repression, there are important differences. First, the undesired interpretation need not be the "winning" unconscious interpretation to motivate avoidance, and second, in addition to preventing conscious access, top-down mechanisms can also affect unconsciously represented interpretations and their estimated probabilities. Such considerations suggest that people whose symptoms are better described by top-down mechanisms – those that do unconsciously recognize the correct interpretation of their emotional reaction as a likely possibility – may need help recognizing their previously reinforced "avoidant" cognitive habits (such as the motivated thought substitution and attentional strategies described above), and they may also need assistance in learning to replace those habits with more emotionally adaptive strategies.

In contrast to a model of repression, the bottom-up mechanisms we have described may be more consistent with a type of dissociation, broadly construed, in which automatic emotional reactions (based on implicit statistical learning) can become dissociated from conscious understanding of what is causing them and why (Bucci, 2016). In such cases, the process of becoming aware of what one is feeling would consist of a transformative process in which a person learns to map their experienced bodily/cognitive reactions onto the appropriate emotion concept representations, often by learning to consciously identify/understand the aspects of their internal model that are causing such reactions. The ability to map one's own felt reactions to emotion concepts is also in part a function of the range of concepts one has in one's emotion repertoire. If someone's repertoire is very limited it may consist of simple good vs. bad distinctions, whereas if it is extensive it may consist of a wide range of concepts (with associated words, images and metaphors) that can be used to verbally communicate to oneself and others how one is feeling (Lane and Schwartz, 1987). As highlighted above, this ability to recognize one's emotions can also result in a feedback process that could plausibly change what is felt.

The neural mechanisms involved in these deficits are critical because they involve the rACC/MPFC, areas that are also involved in the regulation of physiological arousal (Thayer et al., 2012). As described above, the relationship with arousal is such that when arousal is high these structures become inactive, unable to execute their function of concept-level emotion representation. Thus, rather than an active mechanism at work keeping fully formed

mental contents (including emotion concepts) in the unconscious (repression), these mechanisms entail that consciously recognizing one's emotions will not occur if the level of arousal associated with engaging conceptual representations is too high (dissociation). Exactly how this evolves phenomenologically and neuroanatomically has not been studied, but one possibility is that individuals may try to express how they feel in words but simply find that they aren't able to do this, analogous to a weight-lifter who does not have the strength to lift a particular weight off the ground, and thus it appears that nothing has happened or been attempted. As described above, arousal may also increase as an undesired interpretation of one's emotional state gets closer and closer to becoming conscious (i.e., as its probability estimate continues to increase), and if arousal becomes too intense it could cause the rACC/MPFC to deactivate before the person gains conscious access. Relatedly, previous work has also discussed how this dissociation process can occur during memory encoding, in which the hippocampus is also deactivated by extremely high arousal (Bucci, 2016; Nadel and Jacobs, 1998). In such cases, implicit emotional learning will associate perceptual cues with automatic emotional responses, but one will fail to form a coherent episodic memory of the situation – and thus *fail to understand the reason for one's later emotional responses to those cues*. One major therapeutic implication for the therapist is the need to monitor and maintain intermediate levels of arousal in a patient/client during the process of trying to help them recognize and understand their own emotions, so as to avoid the arousal-induced inhibition of the MPFC and related brain regions (Lane et al., 2015a).

It is intriguing to consider the possibility that the more healthy, nurturing and empathic one's early childhood environment has been, the better equipped a person's medial prefrontal cortex may be for appropriately generating such concept-level representations of one's emotional states. Conversely, the more abusive and traumatic the experiences of childhood, the more limited these same mechanisms will be. Limitations of this sort, the extreme version of which is called affective agnosia, are associated with many types of maladaptive behaviors including but not limited to impulsivity, self-injury, addiction, and somatization (Lane et al., 2015b). Help from a psychotherapist in providing soothing support and assistance is essential in helping to keep arousal at a tolerable level so that experiencing, labeling and processing one's emotions can proceed. Consistent and accurate empathy from the therapist may extend the person's emotion concept repertoire and potentially enhance the integrative capacity of the prefrontal cortex. In addition to facilitating attunement and modeling good mentalization skills, such techniques can also be used to promote the mentalization capacities of the client/patient (Allen, 2013). We have previously argued that the mentalization function of the medial prefrontal cortex is related to arousal in an "inverted-U" function (Lane et al., 2015a,b), in which it is optimal in the moderate range of arousal. Thus, the therapist's role would include increasing the peak of the inverted-U and keeping arousal levels within appropriate bounds for recognizing one's emotions and their underlying causes.

Relatedly, one of us (RDL) and his colleagues (Lane et al., 2015a) has recently proposed that the essential ingredients of enduring change in psychotherapy are activation of old problematic memories (held within the internal model), activation of a new, corrective emotional experience that can be reconsolidated with the old memory, and practicing new ways of interpreting situations and responding to them until they become automatic. This process essentially involves converting a series of episodic memories into more enduring semantic structures that incorporate new emotional information. Put another way, the internal model can be updated and improved (in the sense of improved adaptability in the adult world) for therapeutic purposes by bringing into conscious awareness emotional experiences associated with the previous way of interpreting/responding, and modifying the internal model by having vivid conscious subjective emotional experiences that are corrective. By changing one's memories, and how they are conceptualized/interpreted (and thus changing what those memories predict within the internal model and the probability that they will be consciously accessed), future emotional reactions and voluntary behaviors in relevant contexts will be transformed. Repeatedly practicing news ways of interpreting and responding to the world will also serve the purpose of updating the implicit statistical probabilities that have been learned.

To illustrate, consider again cases like that of Walter. In some such cases, it is possible that no criteria are met that would justify the conclusion that unconscious anger is present (i.e., no anger-specific appraisal pattern, no anger-specific bodily/cognitive reaction, no unconscious anger recognition, etc.). Instead such individuals might simply experience a tense, uncomfortable bodily state resulting from conflicting automatic appraisals; the resulting bodily feeling might be conceptualized as a painful somatic state or recognized as "emotional" in nature, but in either case it will not feel consistent with any basic emotion category like anger or fear. In this situation, a therapist might instead be understood to help a patient contextualize or re-conceptualize the memory of the precipitating event – and this could lead to both a change in the pattern of automatic appraisals and a resultant change in their subsequent bodily/cognitive state (Liberzon and Sripada, 2007). If this were the case, it would imply that when a patient "gets in touch with" an emotion during therapy, that the emotion was *not actually present previously*. Instead the therapist could be understood to have facilitated a change in the conceptualization/appraisal of one's situation, and this would lead to a new and different emotional reaction, and that new reaction could be therapeutic for the patient. We do not suggest that this is always the case; however, some clinical cases of "getting in touch with emotion" may be better explained as the generation of a new emotional response, rather than gaining conscious access to a presently existent unconscious emotion.

If this were the case, a question arises regarding how one could account for the "aha" experiences that often occur in psychotherapy in which a person feels that they have indeed uncovered a previously concealed or defended-against emotion. Based on a classic psychodynamic model, which is implicitly used in psychotherapy modalities that involve "getting in touch with one's feelings," it is assumed that the latter involves becoming consciously aware of and labeling something that was already there. While this may occur at times, our model also allows an alternative that is different in two fundamental ways: 1) there isn't one thing there but rather multiple possible options available for selection about how one feels; 2) the feeling of rightness or correctness does not actually involve identifying something that was there, but rather finding the interpretation, at the conceptual level, that best "fits" the perceived situation. That is, it involves finding the concept-level representation that best predicts all aspects of the situation, and hence that best describes the significance of the interaction between the person and the situation immediately at hand – whether that be the immediate physical surroundings or the situation that one is considering in one's mind. Such a fit involves understanding the significance of the situation for one's needs, values and goals as they are described in the internal model. Thus, the "yes" experience – "yes, that is how I feel" – may not come from uncovering what was there but rather finding the conceptualization, emotion label, and/or description that best captures – for the first time – the implications of the current circumstances for meeting or not meeting one's needs, goals and values at that present moment.

This perspective has important implications for how therapists view their role. While their extensive training confers expertise in human psychology/behavior, there is the potential danger that

therapists can view themselves as knowing more about what a patient/client is feeling than they themselves do. One antidote to this potential "omniscient therapist" stance is to adopt the "Not Knowing" stance endorsed by mentalization-based therapists (e.g., Allen, 2013; Fonagy and Luyten, 2009). The latter stance is based on the realization that it is impossible for one person to know what another person is feeling without the active participation of that person (Summers, 2013). From the perspective of our model, one can treat both therapist and patient/client as each having unique vantage points – neither of which is superior to the other. The patient/client has privileged access to information within the domains of interoception and memory, but the therapist can also uniquely assess a patient/client's behavior exteroceptively from a more neutral third-party vantage point. Importantly, both vantage points can gather evidence consistent with different possible interpretations, and a therapist might readily attend to evidence that a patient/client has unintentionally ignored (and vice-versa). At this point, one can view a therapist, not as "telling a person how they really feel," but as suggesting hypotheses for the patient/client to attend to and assess. When an interpretation is found that feels right to the patient/client, and this increases their sense of understanding and wellbeing, this need not be seen as the result of the therapist's superior knowledge; instead, it can just be seen as a useful new interpretation that ultimately changed how the patient/client felt for the better.

7.3. Implications for research

Of most general importance to future research, the considerations that have contributed to our model provide many good neuroscientific- and psychological-level reasons for taking the idea of unconscious emotion seriously. Yet they also highlight the need for new studies that provide more direct support for the idea. Fortunately, the theoretical mechanisms discussed in this article appear to offer several interesting implications that would facilitate such studies.

As one example, our model suggests that appraisal mechanisms may operate on probabilistic internally represented descriptions of one's situation, and that automatic emotional responses may be generated proportional to the probabilities of accuracy associated with different interpretations. One way future studies might test this is by using a conditioning paradigm where participants learn to implicitly associate a perceptual cue with painful and non-painful outcomes at different probabilistic levels (e.g., where cue exposure predicts a 20% vs. 10% chance of painful electric shock). One could then measure peripheral physiological responses, self-reported beliefs about threat, and self-reported feelings of fear (and how correlated each of these measures are) in response to cue exposure under these different learned probabilistic relationships. Our model would be supported if 1) physiological responses were greater to cues associated with higher vs. lower probabilities of pain, and 2) this relationship remained even under conditions where no awareness of threat or fear was reported in relation to the cue. Relatedly, one might also test this aspect of our model by finding/creating stimuli that generate bi-stable percepts (such as a necker cube), but where one of the two competing perceptual interpretations is emotionally significant and the other is not. Our model would predict that brief exposures to such stimuli should trigger measurable emotional/physiological responses, even when the self-reported visual experience only includes the non-emotional perceptual interpretation.[3]

As another example, if unconscious/unrecognized emotion is in some cases the result of the top-down suppression and dynamic filtering mechanisms we have described, then this may imply several testable predictions. One prediction with regard to such patients is that semantic priming effects associated with the unconscious representation of emotion concepts ought to be detectable in both the neural and behavioral domain (e.g., Cacciamani et al., 2014; Sanguinetti et al., 2014). Since these semantic effects would not be observed if emotion concept interpretations were not represented unconsciously, this would offer an interesting test of the extent to which unconscious emotion recognition is present in individuals who show outward signs of a specific emotion but do not report feeling that emotion. Creative approaches may also be needed that allow intensive experimental studies in single subjects and the utilization of unique information specific to a given clinical case (Chassan, 1979).

Another interesting possibility is that factors known to promote the recovery of forgotten (including intentionally forgotten) memories might also increase the accessibility of unconscious emotions. For example, two factors known to promote spontaneous recovery of forgotten memories are cue reinstatement (Bäuml and Samenieh, 2012a,b; Goernert and Larson, 2010; Sahakyan and Kelley, 2002; Smith and Moynan, 2008) and repeated retrieval attempts (Erdelyi, 1996; Kazén and Solís-Macías, 2014). Cue reinstatement involves the finding that if one can be presented with enough cues that are sufficiently unique in their association to the forgotten information that one is attempting to access (e.g., related objects, contexts), this can increase accessibility. This suggests the possibility that if a therapist were capable of providing or eliciting enough sufficiently unique cues to the emotion concept in question, this might also increase their chances of gaining awareness of their emotional state. The free association method which is foundational in psychodynamic psychotherapies is highly consistent with this concept. Perhaps this could also help to explain how non-verbal therapies such as dance and art therapy work. The evidence regarding repeated retrieval attempts suggests that one's chances of successfully retrieving an inaccessible memory tend to increase with repeated retrieval attempts. Perhaps by encouraging a patient to pay attention to their own emotional state, and facilitating repeated attempts to understand it, such as with journaling (Pennebaker, 1993), therapists might also facilitate accessibility in a similar manner. Finally, the passage of time is also correlated with spontaneous recovery of memory (Wheeler, 1995), which suggests that, in some cases, inaccessible emotions might also eventually become accessible on their own. This may be a hidden benefit of longer term vs. brief therapies. While important differences likely exist between concept representations in the emotional and non-emotional domain of semantic memory, and also in the factors acting to maintain their relative accessibility/inaccessibility, testing these predictions offers one interesting means of examining whether mechanisms similar to retrieval-induced inhibition and retrieval suppression might also play an explanatory role in unconscious emotion as we have suggested.

The mechanisms we have described may also be relevant to understanding how emotion contributes to the etiology, onset and course of systemic medical disorders. Much has been learned over the past half century demonstrating that self-reported emotional states such as depression and anxiety are associated with increased morbidity and mortality (Frasure-Smith and Lesperance, 2005; Katon et al., 2005; Onitilo et al., 2006), and also that the processes that generate negative emotion can trigger autonomic/endocrine/immune changes (e.g., reduced vagal tone, cortisol dysregulation, increased inflammation) that promote systemic health risks (and also amplify somatic pain; reviewed in Slavich and Irwin, 2014). What has received less attention is the possible role of unconscious emotion in these processes, or whether

[3] Binocular rivalry paradigms (e.g., Zhang et al., 2011) might be adapted to test for emotional/physiological responses to unconscious perceptual interpretations in a similar manner (i.e., using competing emotional and non-emotional stimuli).

emotional feelings/appraisals must instead be conscious to have such effects. Attempts to understand unconscious emotional processes in health fell into disrepute by focusing exclusively on the hypothesized links between specific unconscious conflicts and specific diseases (Alexander, 1950) instead of the more general associations between unconscious emotion processes articulated here and adverse health outcomes that may vary depending upon the genetic predispositions and personal history of the individual. An important implication of the current perspective is that methods used to study unconscious emotion should be applied in clinical contexts to evaluate their possible role in contributing to poor physical health (Lane, 2008). This review has demonstrated that unconscious emotion can take many different forms and be maintained through many different mechanisms. If ways can be developed to identify such processes and differentially intervene for purposes of primary and secondary prevention, the benefits to personal health and the reduction of health care costs could be considerable.

With regard to intervention, it is noteworthy that some psychopharmacological medications plausibly act to alter unconscious emotional processes. Antidepressant medications, for example, have been shown to reduce the magnitude of amygdala responses to negative stimuli (Anand et al., 2007; Fales et al., 2009; Fu et al., 2004; Godlewska et al., 2012; Sheline et al., 2001), and these responses are plausibly associated with generating both the bodily and attentional aspects of an emotional response (LeDoux, 2012, 1996). This can be useful in reducing how intense and distracting such automatic emotional reactions are, potentially allowing psychotherapy-based interventions to work more effectively (e.g., March et al., 2004; Roiser et al., 2012). Further, in at least some cases, this psychotherapeutic process involves gaining conscious awareness of one's own emotions and their causes (Greenberg and Watson, 2006). Thus, in addition to modulating reportable emotional states like anxiety and depression, we suggest such medications can also be seen as potentially altering the unconscious emotion generation process; they may therefore indirectly facilitate one's ability to gain conscious understanding of their emotions in therapy as well, helping to explain why the combination of pharmacotherapy and psychotherapy is more effective than either alone (Cuijpers et al., 2014).

7.4. Conclusion

In conclusion, the proposed model offers multiple neuroscientifically plausible mechanisms whereby one's own emotional responses might not be consciously recognized. By applying broad lessons about neural function from other areas of cognitive neuroscience, we have highlighted multiple potential top-down and bottom-up processes that could maintain unconscious emotion, and highlighted means of testing these different proposed mechanisms. We have also highlighted how, if confirmed, these variants could have specific implications for psychotherapeutic treatment. More generally, this model appears to offer a clearer sense of the plausible nature of unconscious emotion from the perspective of cognitive neuroscience. It can include the unconscious appraisal of various aspects of a probabilistic internal model. It can include unrecognized biases in attention, memory retrieval, and action selection strategies. It can also include consciously felt, but misunderstood, bodily responses. Finally, it can occur both with and without unconscious emotion recognition. Future work should be designed to further examine each of these possibilities in greater detail.

References

Alexander, F., 1950. Psychosomatic Medicine. Norton, New York.

Allen, J., 2013. Mentalizing in the Development and Treatment of Attachment Trauma. Karnac.

Anand, A., Li, Y., Wang, Y., Gardner, K., Lowe, M., 2007. Reciprocal effects of antidepressant treatment on activity and connectivity of the mood regulating circuit: an FMRI study. J. Neuropsychiatry Clin. Neurosci. 19, 274–282, http://dx.doi.org/10.1176/jnp.2007.19.3.274.

Andersen, R., Cui, H., 2009. Intention, action planning, and decision making in parietal-frontal circuits. Neuron 63, 568–583, http://dx.doi.org/10.1016/j.neuron.2009.08.028.

Anderson, M., Hanslmayr, S., 2014. Neural mechanisms of motivated forgetting. Trends Cogn. Sci. 18, 279–282, http://dx.doi.org/10.1016/j.tics.2014.03.002.

Anderson, F., Sherman, E., 2013. Pathways to Pain Relief. Amazon Digital Platform.

Anderson, F., 1998. Psychic elaboration of musculoskeletal back pain: Ellen's story. In: Aron, L., Anderson, F. (Eds.), Relational Perspectives on the Body. Relational Perspectives Book Series, Vol. 12. Analytic Press, Mahwah, NJ, pp. 287–322.

Anderson, M., 2003. Rethinking interference theory: executive control and the mechanisms of forgetting. J. Mem. Lang. 49, 415–445, http://dx.doi.org/10.1016/j.jml.2003.08.006.

Anderson, F., 2017. It wasn't safe to feel angry: disrupted early attachment and the development of chronic pain in later life. In: Hamilton, L. (Ed.), Unlocking Pain: Disrupted Attachment and Chronic Physical Pain. Karnac Books.

Arnsten, A., Robbins, T., 2002. Neurochemical modulation of prefrontal cortical function in humans and animals. In: Principles of Frontal Lobe Function., pp. 51–84.

Arnsten, A., 1998. Catecholamine modulation of prefrontal cortical cognitive function. Trends Cogn. Sci. 2, 436–447.

Aron, A., Poldrack, R., 2006. Cortical and subcortical contributions to stop signal response inhibition: role of the subthalamic nucleus. J. Neurosci. 26, 2424–2433, http://dx.doi.org/10.1523/JNEUROSCI.4682-05.2006.

Aron, A., Durston, S., Eagle, D., Logan, G., Stinear, C., Stuphorn, V., 2007. Converging evidence for a fronto-basal-ganglia network for inhibitory control of action and cognition. J. Neurosci. 27, 11860–11864, http://dx.doi.org/10.1523/JNEUROSCI.3644-07.2007.

Bäuml, K., Samenieh, A., 2012a. Selective memory retrieval can impair and improve retrieval of other memories. J. Exp. Psychol. Learn. Mem. Cogn. 38, 488–494, http://dx.doi.org/10.1037/a0025683.

Bäuml, K., Samenieh, A., 2012b. Influences of part-list cuing on different forms of episodic forgetting. J. Exp. Psychol. Learn. Mem. Cogn. 38, 366–375, http://dx.doi.org/10.1037/a0025367.

Baars, B., 2005. Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. Prog. Brain Res. 150, 45–53.

Baddeley, A., Eysenck, M., Anderson, M., 2015. Memory, 2nd ed. Psychology Press.

Bar, M., 2004. Visual objects in context. Nat. Rev. Neurosci. 5, 617–629, http://dx.doi.org/10.1038/nrn1476.

Bargh, J., Morsella, E., 2008. The unconscious mind. Perspect. Psychol. Sci. 3, 73–79, http://dx.doi.org/10.1111/j.1745-6916.2008.00064.x.

Barrett, L., Satpute, A., 2013. Large-scale brain networks in affective and social neuroscience: towards an integrative functional architecture of the brain. Curr. Opin. Neurobiol. 23, 361–372, http://dx.doi.org/10.1016/j.conb.2012.12.012.

Barrett, L., Mesquita, B., Gendron, M., 2011. Context in emotion perception. Curr. Dir. Psychol. Sci. 20, 286–290, http://dx.doi.org/10.1177/0963721411422522.

Barrett, L., 2006. Are emotions natural kinds? Perspect. Psychol. Sci. 1, 28–58, http://dx.doi.org/10.1111/j.1745-6916.2006.00003.x.

Bastos, A., Usrey, W., Adams, R., Mangun, G., Fries, P., Friston, K., 2012. Canonical microcircuits for predictive coding. Neuron 76, 695–711, http://dx.doi.org/10.1016/j.neuron.2012.10.038.

Bechara, A., Damasio, H., Tranel, D., Damasio, A., 1997. Deciding advantageously before knowing the advantageous strategy. Science 275 (80), 1293–1295.

Benoit, R., Anderson, M., 2012. Opposing mechanisms support the voluntary forgetting of unwanted memories. Neuron 76, 450–460, http://dx.doi.org/10.1016/j.neuron.2012.07.025.

Braver, T., 2012. The variable nature of cognitive control: a dual mechanisms framework. Trends Cogn. Sci. 16, 106–113, http://dx.doi.org/10.1016/j.tics.2011.12.010.

Brenner, C., 1973. An Elementary Textbook of Psychoanalysis. International Universities Press.

Brosch, T., Sander, D., 2013. The appraising brain: towards a neuro-cognitive model of appraisal processes in emotion. Emot. Rev. 5, 163–168, http://dx.doi.org/10.1177/1754073912468298.

Brosch, T., Coppin, G., Schwartz, S., Sander, D., 2012. The importance of actions and the worth of an object: dissociable neural systems representing core value and economic value. Soc. Cogn. Affect. Neurosci. 7, 497–505, http://dx.doi.org/10.1093/scan/nsr036.

Brosschot, J., Verkuil, B., Thayer, J., 2010. Conscious and unconscious perseverative cognition: is a large part of prolonged physiological activity due to unconscious stress? J. Psychosom. Res. 69, 407–416.

Brosschot, J., 2010. Markers of chronic stress: prolonged physiological activation and (un) conscious perseverative cognition. Neurosci. Biobehav. Rev. 35, 46–50.

Bucci, W., 2016. Divide and multiply; a multi-dimensional view of dissociative processes. In: Howell, E., Itzkowitz, S. (Eds.), The Dissociative Mind in Psychoanalysis: Understanding and Working with Trauma and Dissociation. Routledge, pp. 187–199.

Buelow, M., Suhr, J., 2009. Construct validity of the iowa gambling task. Neuropsychol. Rev. 19, 102–114, http://dx.doi.org/10.1007/s11065-009-9083-4.

Buhle, J., Silvers, J., Wager, T., Lopez, R., Onyemekwu, C., Kober, H., Weber, J., Ochsner, K., 2014. Cognitive reappraisal of emotion: a meta-analysis of human neuroimaging studies. Cereb. Cortex 24, 2981–2990, http://dx.doi.org/10.1093/cercor/bht154.
Burns, J., Quartana, P., Bruehl, S., 2008. Anger inhibition and pain: conceptualizations, evidence and new directions. J. Behav. Med. 31, 259–279, http://dx.doi.org/10.1007/s10865-008-9154-7.
Cacciamani, L., Mojica, A., Sanguinetti, J., Peterson, M., 2014. Semantic access occurs outside of awareness for the ground side of a figure. Atten. Percept. Psychophys., http://dx.doi.org/10.3758/s13414-014-0743-y.
Chan, J., 2009. When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. J. Mem. Lang. 61, 153–170, http://dx.doi.org/10.1016/j.jml.2009.04.004.
Chassan, J., 1979. Research Design in Clinical Psychology and Psychiatry, 2nd edition, revised and enlarged. Irvington Publishers, New York.
Conenna, S., 2013. Use Your Mind to Heal Your Body. CreateSpace Publishers, Las Vegas.
Cuijpers, P., Sijbrandij, M., Koole, S., Andersson, G., Beekman, A., Reynolds, C., 2014. Adding psychotherapy to antidepressant medication in depression and anxiety disorders: a meta-analysis. World Psychiatry 13, 56–67, http://dx.doi.org/10.1002/wps.20089.
Cushman, F., 2013. Action, outcome, and value: a dual-system framework for morality. Personal. Soc. Psychol. Rev. 17, 273–292, http://dx.doi.org/10.1177/1088868313495594.
Danker, J., Anderson, J., 2010. The ghosts of brain states past: remembering reactivates the brain regions engaged during encoding. Psychol. Bull. 136, 87–102, http://dx.doi.org/10.1037/a0017937.
Daw, N., Shohamy, D., 2008. The cognitive neuroscience of motivation and learning. Soc. Cogn. 26, 593–620, http://dx.doi.org/10.1521/soco.2008.26.5.593.
Dayan, P., Daw, N., 2008. Decision theory, reinforcement learning, and the brain. Cogn. Affect. Behav. Neurosci. 8, 429–453, http://dx.doi.org/10.3758/CABN.8.4.429.
Dehaene, S., Naccache, L., 2001. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. Cognition 79, 1–37.
Dehaene, S., Sigman, M., 2012. From a single decision to a multi-step algorithm. Curr. Opin. Neurobiol. 22, 937–945, http://dx.doi.org/10.1016/j.conb.2012.05.006.
Dehaene, S., Sergent, C., Changeux, J., 2003. A neuronal network model linking subjective reports and objective physiological data during conscious perception. Proc. Natl. Acad. Sci. 100, 8520–8525.
Dehaene, S., Changeux, J., Naccache, L., Sackur, J., Sergent, C., 2006. Conscious, preconscious, and subliminal processing: a testable taxonomy. Trends Cogn. Sci. 10, 204–211, http://dx.doi.org/10.1016/j.tics.2006.03.007.
Dehaene, S., 2014. Consciousness and the Brain. Viking Press.
Del Cul, A., Dehaene, S., Reyes, P., Bravo, E., Slachevsky, A., 2009. Causal role of prefrontal cortex in the threshold for access to consciousness. Brain 132, 2531–2540, http://dx.doi.org/10.1093/brain/awp111.
Depue, B., 2012. A neuroanatomical model of prefrontal inhibitory modulation of memory retrieval. Neurosci. Biobehav. Rev. 36, 1382–1399, http://dx.doi.org/10.1016/j.neubiorev.2012.02.012.
Derakshan, N., Eysenck, M., Myers, L., 2007. Emotional information processing in repressors: the vigilance-avoidance theory. Cogn. Emot. 21, 1585–1614, http://dx.doi.org/10.1080/02699930701499857.
Druzgal, T., D'Esposito, M., 2003. Dissecting contributions of prefrontal cortex and fusiform face area to face working memory. J. Cogn. Neurosci. 15, 771–784, http://dx.doi.org/10.1162/089892903322370708.
Dwyer, S., Huebner, B., Hauser, M.D., 2010. The linguistic analogy: motivations, results, and speculations. Top. Cogn. Sci. 2, 486–510, http://dx.doi.org/10.1111/j.1756-8765.2009.01064.x.
Erdelyi, M., 1996. The Recovery of Unconscious Memories: Hypermnesia and Reminiscence. University of Chicago Press, Chicago, IL.
Fales, C., Barch, D., Rundle, M., Mintun, M., Mathews, J., Snyder, A., Sheline, Y., 2009. Antidepressant treatment normalizes hypoactivity in dorsolateral prefrontal cortex during emotional interference processing in major depression. J. Affect. Disord. 112, 206–211.
Feldman, H., Friston, K., 2010. Attention, uncertainty, and free-energy. Front. Hum. Neurosci. 4, 215, http://dx.doi.org/10.3389/fnhum.2010.00215.
Fonagy, P., Luyten, P., 2009. A developmental, mentalization-based approach to the understanding and treatment of borderline personality disorder. Dev. Psychopathol. 21, 1355–1381, http://dx.doi.org/10.1017/S0954579409990198.
Frasure-Smith, N., Lesperance, F., 2005. Reflections on depression as a cardiac risk factor. Psychosom. Med. 67 (Suppl. 1), S19–S25, http://dx.doi.org/10.1097/01.psy.0000162253.07959.db, 67/Supplement_1/S19 [pii].
Friedman, B., Kreibig, S., 2010. The biopsychology of emotion: current theoretical, empirical, and methodological perspectives. Biol. Psychol. 84, 381–382.
Friedman, B., Thayer, J., 1998. Autonomic balance revisited: panic anxiety and heart rate variability. J. Psychosom. Res. 44, 133–151, S002239999700202X [pii].
Frijda, N., 1986. The Emotions. Cambridge University Press, Cambridge, UK.
Friston, K., Frith, C., 2015. A duet for one. Conscious. Cogn. 36, 390–405, http://dx.doi.org/10.1016/j.concog.2014.12.003.
Friston, K., 2005. A theory of cortical responses. Philos. Trans. R. Soc. Lond. B: Biol. Sci. 360, 815–836, http://dx.doi.org/10.1098/rstb.2005.1622.
Friston, K., 2010. The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. 11, 127–138, http://dx.doi.org/10.1038/nrn2787.
Fu, C., Williams, S., Cleare, A., Brammer, M., Walsh, N., Kim, J., Andrew, C., Pich, E., Williams, P., Reed, L., Mitterschiffthaler, M., Suckling, J., Bullmore, E., 2004. Attenuation of the neural response to sad faces in major depression by antidepressant treatment: a prospective, event-related functional magnetic resonance imaging study. Arch. Gen. Psychiatry 61, 877–889, http://dx.doi.org/10.1001/archpsyc.61.9.877.
Gagnepain, P., Henson, R., Anderson, M., 2014. Suppressing unwanted memories reduces their unconscious influence via targeted cortical inhibition. Proc. Natl. Acad. Sci. U. S. A. 111, E1310–E1319, http://dx.doi.org/10.1073/pnas.1311468111.
Gazzaniga, M., Ivry, R., Mangun, G., 2014. Cognitive Neuroscience: The Biology of the Mind, 4rd ed. W.W. Norton, New York.
Gendlin, E., 1982. Focusing. Bantam Books.
Ginot, E., 2015. The Neuropsychology of the Unconscious—Integrating Brain and Mind in Psychotherapy. Norton, New York.
Godlewska, B., Norbury, R., Selvaraj, S., Cowen, P., Harmer, C., 2012. Short-term SSRI treatment normalises amygdala hyperactivity in depressed patients. Psychol. Med. 42, 2609–2617, http://dx.doi.org/10.1017/S0033291712000591.
Goernert, P., Larson, M., 2010. The initiation and release of retrieval inhibition. J. Gen. Psychol. 121, 61–66.
Greenberg, L., Watson, J.C., 2006. Emotion-focused Therapy for Depression. American Psychological Association, Washington, DC, USA.
Greenberg, L., 2010. Emotion-Focused Therapy: Theory and Practice. APA Press, Washington, DC.
Greene, J., Haidt, J., 2002. How (and where) does moral judgment work? Trends Cogn. Sci. 6, 517–523.
Griffith, J., Griffith, M., 1994. The Body Speaks: Therapeutic Dialogues for Mind-Body Problems, 1st ed. Basic Books, New York.
Grippo, A.J., Johnson, A.K., 2009. Stress, depression and cardiovascular dysregulation: a review of neurobiological mechanisms and the integration of research from preclinical disease models. Stress Int. J. Biol. Stress 12, 1–21.
Gross, J., Levenson, R., 1997. Hiding feelings: the acute effects of inhibiting negative and positive emotion. J. Abnorm. Psychol. 106, 95–103, http://dx.doi.org/10.1037//0021-843X.106.1.95.
Gu, X., Hof, P., Friston, K., Fan, J., 2013. Anterior insular cortex and emotional awareness. J. Comp. Neurol. 521, 3371–3388, http://dx.doi.org/10.1002/cne.23368.
Gupta, R., Koscik, T., Bechara, A., Tranel, D., 2011. The amygdala and decision-making. Neuropsychologia 49, 760–766, http://dx.doi.org/10.1016/j.neuropsychologia.2010.09.029.
Haidt, J., 2001. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. Psychol. Rev. 108, 814–834.
Hodgson, R., Rachman, S., 1974. II. Desynchrony in measures of fear. Behav. Res. Ther. 12, 319–326, http://dx.doi.org/10.1016/0005-7967(74)90006-0.
Hohwy, J., 2014. The Predictive Mind. Oxford University Press.
Hughes, G., Velmans, M., De Fockert, J., 2009. Unconscious priming of a no-go response. Psychophysiology 46, 1258–1269, http://dx.doi.org/10.1111/j.1469-8986.2009.00873.x, PSYP873 [pii].
Huntsinger, J.R., 2013. Does emotion directly tune the scope of attention? Curr. Dir. Psychol. Sci. 22, 265–270, http://dx.doi.org/10.1177/0963721413480364.
Irwin, M., Cole, S., 2011. Reciprocal regulation of the neural and innate immune systems. Nat. Rev. Immunol. 11, 625–632, http://dx.doi.org/10.1038/nri3042.
Katon, W.J., Rutter, C., Simon, G., Lin, E.H., Ludman, E., Ciechanowski, P., Kinder, L., Young, B., Von Korff, M., 2005. The association of comorbid depression with mortality in patients with type 2 diabetes. Diabetes Care 28, 2668–2672, 28/11/2668 [pii].
Kazén, M., Solís-Macías, V., 2014. Recall and recognition hypermnesia for Socratic stimuli. Memory, 1–18, http://dx.doi.org/10.1080/09658211.2014.990981.
Kemp, A., Quintana, D., Gray, M., Felmingham, K., Brown, K., Gatt, J., 2010. Impact of depression and antidepressant treatment on heart rate variability: a review and meta-analysis. Biol. Psychiatry 67, 1067–1074.
Khalsa, S., Rudrauf, D., Feinstein, J., Tranel, D., 2009. The pathways of interoceptive awareness. Nat. Neurosci. 12, 1494–1496, http://dx.doi.org/10.1038/nn.2411.
Khalsa, S., Feinstein, J., Li, W., Feusner, J.D., Adolphs, R., Hurlemann, R., 2016. Panic anxiety in humans with bilateral amygdala lesions: pharmacological induction via cardiorespiratory interoceptive pathways. J. Neurosci. 36, 3559–3566, http://dx.doi.org/10.1523/JNEUROSCI.4109-15.2016.
Kiefer, M., Barsalou, L., 2013. Grounding the human conceptual system in perception, action, and internal states. In: Prinz, W., Beisert, M., Herwig, A. (Eds.), Action Science: Foundations of an Emerging Discipline. MIT Press, Cambridge, MA, pp. 381–407.
Kihlstrom, J., Mulvaney, S., Tobias, B., Tobis, I., 2000. The emotional unconscious. In: Eich, E., Kihlstrom, J., Bower, G., Forgas, J., Niedenthal, P. (Eds.), Cognition and Emotion. Oxford University Press, New York, pp. 30–86.
Kihlstrom, J., 2002. No need for repression. Trends Cogn. Sci. 6, 502.
Kihlstrom, J., 2006. Trauma and memory revisited. In: Memory and Emotion. Blackwell Publishing Ltd, Oxford, UK, pp. 259–291, http://dx.doi.org/10.1002/9780470756232.ch12.
Kikuchi, H., Fujii, T., Abe, N., Suzuki, M., Takagi, M., Mugikura, S., Takahashi, S., Mori, E., 2010. Memory repression: brain mechanisms underlying dissociative amnesia. J. Cogn. Neurosci. 22, 602–613, http://dx.doi.org/10.1162/jocn.2009.21212.
Knight, R., Grabowecky, M., 1995. Escape from linear time: prefrontal cortex and conscious experience. In: Gazzaniga, M. (Ed.), The Cognitive Neurosciences. MIT Press, Cambridge, MA, pp. 1357–1371.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., Fehr, E., 2006. Diminishing reciprocal fairness by disrupting the right prefrontal cortex. Science 314 (80), 829–832, http://dx.doi.org/10.1126/science.1129156.
Konnopka, A., Schaefert, R., Heinrich, S., Kaufmann, C., Luppa, M., Herzog, W., König, H.-H., 2012. Economics of medically unexplained symptoms: a systematic review of the literature. Psychother. Psychosom. 81, 265–275, http://dx.doi.org/10.1159/000337349.
Kouider, S., Dehaene, S., Jobert, A., Le Bihan, D., 2007. Cerebral bases of subliminal and supraliminal priming during reading. Cereb. Cortex 17, 2019–2029.
Kragel, P., Labar, K., 2013. Multivariate pattern classification reveals autonomic and experiential representations of discrete emotions. Emotion 13, 681–690, http://dx.doi.org/10.1037/a0031820.
Kreibig, S., 2010. Autonomic nervous system activity in emotion: a review. Biol. Psychol. 84, 394–421, http://dx.doi.org/10.1016/j.biopsycho.2010.03.010.
Kreiman, G., Koch, C., Fried, I., 2000. Imagery neurons in the human brain. Nature 408, 357–361, http://dx.doi.org/10.1038/35042575.
Kroenke, K., 2003. Patients presenting with somatic complaints: epidemiology, psychiatric comorbidity and management. Int. J. Methods Psychiatr. Res. 12, 34–43.
Kuhl, B., Dudukovic, N., Kahn, I., Wagner, A., 2007. Decreased demands on cognitive control reveal the neural processing benefits of forgetting. Nat. Neurosci. 10, 908–914, http://dx.doi.org/10.1038/nn1918.
Lambie, J., Marcel, A., 2002. Consciousness and the varieties of emotion experience: a theoretical framework. Psychol. Rev. 109, 219–259.
Landa, A., Peterson, B., Fallon, B., 2012. Somatoform pain: a developmental theory and translational research review. Psychosom. Med. 74, 717–727.
Lane, R., Schwartz, G., 1987. Levels of emotional awareness: a cognitive-developmental theory and its application to psychopathology. Am. J. Psychiatry 144, 133–143.
Lane, R., Nadel, L., Allen, J., Kaszniak, A., 2000. The study of emotion from the perspective of cognitive neuroscience. In: Lane, R., Nadel, L. (Eds.), Cognitive Neuroscience of Emotion. Oxford University Press.
Lane, R., Ryan, L., Nadel, L., Greenberg, L., 2015a. Memory reconsolidation, emotional arousal and the process of change in psychotherapy: new insights from brain science. Behav. Brain Sci. (in press).
Lane, R., Weihs, K., Herring, A., Hishaw, A., Smith, R., 2015b. Affective agnosia: expansion of the alexithymia construct and a new opportunity to integrate and extend Freud's legacy. Neurosci. Biobehav. Rev., http://dx.doi.org/10.1016/j.neubiorev.2015.06.007 (in press).
Lane, R., 2008. Neural substrates of implicit and explicit emotional processes: a unifying framework for psychosomatic medicine. Psychosom. Med. 70, 214–231.
Lang, P., 1968. Fear reduction and fear behavior: problems in treating a construct. In: Research in Psychotherapy. American Psychological Association, Washington, pp. 90–102, http://dx.doi.org/10.1037/10546-004.
Lang, P., 1988. What are the data of emotion? In: Cognitive Perspectives on Emotion and Motivation. Springer, Netherlands, Dordrecht, pp. 173–191, http://dx.doi.org/10.1007/978-94-009-2792-6_7.
Lazarus, R., 1991. Emotion and Adaptation. Oxford University Press, New York.
LeDoux, J., 1996. The Emotional Brain: The Mysterious Underpinnings of Emotional Life. Simon & Schuster, New York.
LeDoux, J., 2012. Rethinking the emotional brain. Neuron 73, 653–676, http://dx.doi.org/10.1016/j.neuron.2012.02.004.
LeDoux, J., 2013. The slippery slope of fear. Trends Cogn. Sci. 17, 155–156, http://dx.doi.org/10.1016/j.tics.2013.02.004.
Levenson, R., 1994. Human emotion: a functional view. In: Ekman, P., Davidson, R. (Eds.), The Nature of Emotion—Fundamental Questions. Oxford University Press, pp. 123–126.
Lewis, P.A., Critchley, H.D., Smith, A.P., Dolan, R.J., 2005. Brain mechanisms for mood congruent memory facilitation. Neuroimage 25, 1214–1223, http://dx.doi.org/10.1016/j.neuroimage.2004.11.053.
Liberzon, I., Sripada, C., 2007. The functional neuroanatomy of PTSD: a critical review. Prog. Brain Res. 167, 151–169, http://dx.doi.org/10.1016/S0079-6123(07)67011-3.
Lindquist, K., Barrett, L., 2008. Constructing emotion: the experience of fear as a conceptual act. Psychol. Sci. 19, 898–903, http://dx.doi.org/10.1111/j.1467-9280.2008.02174.x.
Lynch, M.A., 2004. Long-term potentiation and memory. Physiol. Rev. 84, 87–136.
Mahler, M., Pine, F., Bergman, A., 2008. The Psychological Birth of the Human Infant: Symbiosis and Individuation. Basic Books.
March, J., Silva, S., Petrycki, S., Curry, J., Wells, K., Fairbank, J., Burns, B., Domino, M., McNulty, S., Vitiello, B., Severe, J., 2004. Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: treatment for Adolescents With Depression Study (TADS) randomized controlled trial. JAMA 292, 807–820, http://dx.doi.org/10.1001/jama.292.7.807.
McConnell, A., Leibold, J., 2001. Relations among the implicit association test discriminatory behavior, and explicit measures of racial attitudes. J. Exp. Soc. Psychol. 37, 435–442, http://dx.doi.org/10.1006/jesp.2000.1470.
Medford, N., Critchley, H., 2010. Conjoint activity of anterior insular and anterior cingulate cortex: awareness and response. Brain Struct. Funct. 214, 535–549, http://dx.doi.org/10.1007/s00429-010-0265-x.
Metzinger, T., 2003. Being No One: The Self-model Theory of Subjectivity. MIT Press, Cambridge, Mass.
Mitchell, D., 2011. The nexus between decision making and emotion regulation: a review of convergent neurocognitive substrates. Behav. Brain Res. 217, 215–231, http://dx.doi.org/10.1016/j.bbr.2010.10.030.
Modell, A., 2008. Horse and rider revisited. Contemp. Psychoanal. 44, 351–366, http://dx.doi.org/10.1080/00107530.2008.10745962.
Modell, A., 2010. The unconscious as a knowledge processing center. In: Petrucelli, J. (Ed.), Knowing, Not-Knowing and Sort-of-Knowing: Psychoanalysis and the Experience of Uncertainty. Karnac, pp. 45–61.
Moors, A., Ellsworth, P.C., Scherer, K., Frijda, N., 2013. Appraisal theories of emotion: state of the art and future development. Emot. Rev. 5, 119–124, http://dx.doi.org/10.1177/1754073912468165.
Moreno-Bote, R., Knill, D., Pouget, A., 2011. Bayesian sampling in visual perception. Proc. Natl. Acad. Sci. U. S. A. 108, 12491–12496, http://dx.doi.org/10.1073/pnas.1101430108.
Nadel, L., Jacobs, W., 1998. Traumatic memory is special. Curr. Dir. Psychol. Sci. 7, 154–157.
Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., Panksepp, J., 2006. Self-referential processing in our brain-a meta-analysis of imaging studies on the self. Neuroimage 31, 440–457, http://dx.doi.org/10.1016/j.neuroimage.2005.12.002.
Nummenmaa, L., Glerean, E., Hari, R., Hietanen, J.K., 2014. Bodily maps of emotions. Proc. Natl. Acad. Sci. U. S. A. 111, 646–651, http://dx.doi.org/10.1073/pnas.1321664111.
O'Craven, K.M., Kanwisher, N., 2000. Mental imagery of faces and places activates corresponding stimulus-specific brain regions. J. Cogn. Neurosci. 12, 1013–1023, http://dx.doi.org/10.1162/08989290051137549.
Onitilo, A.A., Nietert, P.J., Egede, L.E., 2006. Effect of depression on all-cause mortality in adults with cancer and differential effects by cancer site. Gen. Hosp. Psychiatry 28, 396–402, http://dx.doi.org/10.1016/j.genhosppsych.2006.05.006, S0163-8343(06)00091-0 [pii].
Paivio, S., Laurent, C., 2001. Empathy and emotion regulation: reprocessing memories of childhood abuse. J. Clin. Psychol. 57, 213–226.
Pennebaker, J., 1993. Putting stress into words: health, linguistic, and therapeutic implications. Behav. Res. Ther. 31, 539–548, http://dx.doi.org/10.1016/0005-7967(93)90105-4.
Pessoa, L., Adolphs, R., 2010. Emotion processing and the amygdala: from a low road to many roads of evaluating biological significance. Nat. Rev. Neurosci. 11, 773–783, http://dx.doi.org/10.1038/nrn2920.
Pezzulo, G., Rigoli, F., Friston, K., 2015. Active inference, homeostatic regulation and adaptive behavioural control. Prog. Neurobiol. 134, 17–35, http://dx.doi.org/10.1016/j.pneurobio.2015.09.001.
Pobric, G., Jefferies, E., Lambon Ralph, M., 2010. Category-specific versus category-general semantic impairment induced by transcranial magnetic stimulation. Curr. Biol. 20, 964–968, http://dx.doi.org/10.1016/j.cub.2010.03.070.
Pothos, E., 2007. Theories of artificial grammar learning. Psychol. Bull. 133, 227–244, http://dx.doi.org/10.1037/0033-2909.133.2.227.
Pouget, A., Dayan, P., Zemel, R., 2000. Information processing with population codes. Nat. Rev. Neurosci. 1, 125–132, http://dx.doi.org/10.1038/35039062.
Rachman, S., Hodgson, R., 1974. I. Synchrony and desynchrony in fear and avoidance. Behav. Res. Ther. 12, 311–318, http://dx.doi.org/10.1016/0005-7967(74)90005-9.
Rachman, S., 1978. Human fears: a three systems analysis. Scand. J. Behav. Ther. 7, 237–245, http://dx.doi.org/10.1080/16506077809456104.
Railton, P., 2014. The affective dog and its rational tale: intuition and attunement. Ethics 124, 813–859, http://dx.doi.org/10.1086/675876.
Reisenzein, R., 2006. Arnold's theory of emotion in historical perspective. Cogn. Emot. 20, 920–951, http://dx.doi.org/10.1080/02699930600616445.
Robbins, T., Arnsten, A., 2009. The neuropsychopharmacology of fronto-executive function: monoaminergic modulation. Annu. Rev. Neurosci. 32, 267–287, http://dx.doi.org/10.1146/annurev.neuro.051508.135535.
Rofé, Y., 2008. Does repression exist? Memory, pathogenic, unconscious and clinical evidence. Rev. Gen. Psychol. 12, 63–85.
Roiser, J., Elliott, R., Sahakian, B., 2012. Cognitive mechanisms of treatment in depression. Neuropsychopharmacology 37, 117–136, http://dx.doi.org/10.1038/npp.2011.183.
Roseman, I., Spindel, M., Jose, P., 1990. Appraisals of emotion-eliciting events: testing a theory of discrete emotions. J. Pers. Soc. Psychol. 59, 899–915.
Roy, M., Shohamy, D., Wager, T.D., 2012. Ventromedial prefrontal-subcortical systems and the generation of affective meaning. Trends Cogn. Sci. 16, 147–156, http://dx.doi.org/10.1016/j.tics.2012.01.005, S1364-6613(12)00027-7 [pii].
Sackur, J., Dehaene, S., 2009. The cognitive architecture for chaining of two mental operations. Cognition 111, 187–211, http://dx.doi.org/10.1016/j.cognition.2009.01.010.
Sahakyan, L., Kelley, C., 2002. A contextual change account of the directed forgetting effect. J. Exp. Psychol. Learn. Mem. Cogn. 28, 1064–1072.
Sanguinetti, J., Allen, J., Peterson, M., 2014. The ground side of an object: perceived as shapeless yet processed for semantics. Psychol. Sci. 25, 256–264, http://dx.doi.org/10.1177/0956797613502814.
Schacter, D., 2001. Suppression of unwanted memories: repression revisited? Lancet 357, 1724–1725, http://dx.doi.org/10.1016/S0140-6736(00)04931-X.
Scherer, K., 1984. On the nature and function of emotion: a component process approach. In: Scherer, K., Ekman, P. (Eds.), Approaches to Emotion. Erlbaum, Hillsdale, NJ, pp. 293–318.
Scherer, K., 1997. The role of culture in emotion-antecedent appraisal. J. Pers. Soc. Psychol. 73, 902–922.

Schnall, S., Haidt, J., Clore, G., Jordan, A., 2008. Disgust as embodied moral judgment. Pers. Soc. Psychol. Bull. 34, 1096–1109, http://dx.doi.org/10.1177/0146167208317771.
Schultze-Florey, C., Martínez-Maza, O., Magpantay, L., Breen, E., Irwin, M., Gündel, H., O'Connor, M.-F., 2012. When grief makes you sick: bereavement induced systemic inflammation is a question of genotype. Brain Behav. Immun. 26, 1066–1071, http://dx.doi.org/10.1016/j.bbi.2012.06.009.
Seth, A.K., Critchley, H.D., 2013. Extending predictive processing to the body: emotion as interoceptive inference. Behav. Brain Sci. 36, 47–48, http://dx.doi.org/10.1017/S0140525x12002270.
Seth, A.K., 2013. Interoceptive inference, emotion, and the embodied self. Trends Cogn. Sci. 17, 565–573, http://dx.doi.org/10.1016/j.tics.2013.09.007.
Sharpe, M., Carson, A., 2001. Unexplained somatic symptoms, functional syndromes, and somatization: do we need a paradigm shift? Ann. Intern. Med. 134, 926–930.
Sheline, Y., Barch, D., Donnelly, J., Ollinger, J., Snyder, A., Mintun, M., 2001. Increased amygdala response to masked emotional faces in depressed subjects resolves with antidepressant treatment: an fMRI study. Biol. Psychiatry 50, 651–658.
Shimamura, A., 2000. The role of the prefrontal cortex in dynamic filtering. Psychobiology 28, 207–218, http://dx.doi.org/10.3758/BF03331979.
Shiota, M., Kalat, J., 2012. Emotion, 2nd ed. Cengage Learning.
Simons, D.J., Chabris, C.F., 1999. Gorillas in our midst: sustained inattentional blindness for dynamic events. Perception 28, 1059–1074.
Slavich, G., Irwin, M., 2014. From stress to inflammation and major depressive disorder: a social signal transduction theory of depression. Psychol. Bull. 140, 774–815, http://dx.doi.org/10.1037/a0035302.
Smith, R., Lane, R., 2015. The neural basis of one's own conscious and unconscious emotional states. Neurosci. Biobehav. Rev. (in press).
Smith, S., Moynan, S., 2008. Forgetting and recovering the unforgettable. Psychol. Sci. 19, 462–468, http://dx.doi.org/10.1111/j.1467-9280.2008.02110.x.
Smith, R., Alkozei, A., Lane, R.D., Killgore, W.D., 2016. Unwanted reminders: the effects of emotional memory suppression on subsequent neuro-cognitive processing. Conscious. Cogn. 44, 103–113.
Solms, M., Panksepp, J., 2012. The Id knows more than the ego admits: neuropsychoanalytic and primal consciousness perspectives on the interface between affective and cognitive neuroscience. Brain Sci. 2, 147–175, http://dx.doi.org/10.3390/brainsci2020147.
Stonnington, C., Locke, D., Hsu, C.-H., Ritenbaugh, C., Lane, R., 2013. Somatization is associated with deficits in affective Theory of Mind. J. Psychosom. Res. 74, 479–485, http://dx.doi.org/10.1016/j.jpsychores.2013.04.004.
Subic-Wrana, C., Beutel, M., Knebel, A., Lane, R., 2010. Theory of mind and emotional awareness deficits in patients with somatoform disorders. Psychosom. Med. 72, 404–411, http://dx.doi.org/10.1097/PSY.0b013e3181d35e83.
Summers, F., 2013. The Psychoanalytic Vision—The Experiencing Subject, Transcendence and the Therapeutic Process. Routledge, New York.
Tamietto, M., de Gelder, B., 2010. Neural bases of the non-conscious perception of emotional signals. Nat. Rev. Neurosci. 11, 697–709, http://dx.doi.org/10.1038/nrn2889.
Taylor, G., 2000. Recent developments in alexithymia theory and research. Can. J. Psychiatry 45, 134–142.
Thayer, J., Faith, M., 1994. Idiographic nonlinear pattern-classification of autonomic and self-report measures of emotion. Psychosom. Med. 56, 178.
Thayer, J., Lane, R., 2007. The role of vagal function in the risk for cardiovascular disease and mortality. Biol. Psychol. 74, 224–242.
Thayer, J., Yamamoto, S.S., Brosschot, J., 2010. The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors. Int. J. Cardiol. 141, 122–131, http://dx.doi.org/10.1016/j.ijcard.2009.09.543, S0167-5273(09)01487-9 [pii].
Thayer, J., Ahs, F., Fredrikson, M., Sollers, J.J., Wager, T.D., 2012. A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. Neurosci. Biobehav. Rev. 36, 747–756.
Thompson-Schill, S., D'Esposito, M., Kan, I., 1999. Effects of repetition and competition on activity in left prefrontal cortex during word generation. Neuron 23, 513–522, http://dx.doi.org/10.1016/S0896-6273(00)80804-1.
Thompson-Schill, S., Bedny, M., Goldberg, R., 2005. The frontal lobes and the regulation of mental activity. Curr. Opin. Neurobiol. 15, 219–224, http://dx.doi.org/10.1016/j.conb.2005.03.006.
Tomkins, S., 1995. The quest for primary motives: biography and autobiography of an idea. In: Demos, E. (Ed.), In Exploring Affect—The Selected Writings of Silvan S. Tomkins. Cambridge University Press.
Tramoni, E., Aubert-Khalfa, S., Guye, M., Ranjeva, J., Felician, O., Ceccaldi, M., 2009. Hypo-retrieval and hyper-suppression mechanisms in functional amnesia. Neuropsychologia 47, 611–624, http://dx.doi.org/10.1016/j.neuropsychologia.2008.11.012.
Verkuil, B., Brosschot, J., Tollenaar, M., Lane, R., Thayer, J., 2016. Prolonged non-metabolic heart rate variability reduction as a physiological marker of psychological stress in daily life. Ann. Behav. Med., 1–11, http://dx.doi.org/10.1007/s12160-016-9795-7.
Vul, E., Pashler, H., 2008. Measuring the crowd within: probabilistic representations within individuals. Psychol. Sci. 19, 645–647, http://dx.doi.org/10.1111/j.1467-9280.2008.02136.x.
Vul, E., Hanus, D., Kanwisher, N., 2009. Attention as inference: selection is probabilistic; responses are all-or-none samples. J. Exp. Psychol. Gen. 138, 546–560, http://dx.doi.org/10.1037/a0017352.
Wheatley, T., Haidt, J., 2005. Hypnotic disgust makes moral judgments more severe. Psychol. Sci. 16, 780–784, http://dx.doi.org/10.1111/j.1467-9280.2005.01614.x.
Wheeler, M., 1995. Improvement in recall over time without repeated testing: spontaneous recovery revisited. J. Exp. Psychol. Learn. Mem. Cogn. 21, 173–184.
Wiens, S., 2005. Interoception in emotional experience. Curr. Opin. Neurol. 18, 442–447.
Wilson-Mendenhall, C., Barrett, L., Simmons, W., Barsalou, L., 2011. Grounding emotion in situated conceptualization. Neuropsychologia 49, 1105–1127, http://dx.doi.org/10.1016/j.neuropsychologia.2010.12.032.
Winkielman, P., Berridge, K., 2004. Unconscious emotion. Curr. Dir. Psychol. Sci. 13, 120–123, http://dx.doi.org/10.1111/j.0963-7214.2004.00288.x.
Woolf, C., Allchorne, A., Safieh-Garabedian, B., Poole, S., 1997. Cytokines, nerve growth factor and inflammatory hyperalgesia: the contribution of tumour necrosis factor α. Br. J. Pharmacol. 121, 417–424, http://dx.doi.org/10.1038/sj.bjp.0701148.
Zahn, R., Moll, J., Krueger, F., Huey, E.D., Garrido, G., Grafman, J., 2007. Social concepts are represented in the superior anterior temporal cortex. Proc. Natl. Acad. Sci. U. S. A. 104, 6430–6435, http://dx.doi.org/10.1073/pnas.0607061104.
Zajonc, R., 1980. Feeling and thinking: preferences need no inferences. Am. Psychol. 35, 151–175.
Zhang, P., Jamison, K., Engel, S., He, B., He, S., 2011. Binocular rivalry requires visual attention. Neuron 71, 362–369, http://dx.doi.org/10.1016/j.neuron.2011.05.035.
Zylberberg, A., Fernández Slezak, D., Roelfsema, P., Dehaene, S., Sigman, M., 2010. The brain's router: a cortical network model of serial processing in the primate brain. PLoS Comput. Biol. 6, e1000765, http://dx.doi.org/10.1371/journal.pcbi.1000765.
Zylberberg, A., Dehaene, S., Roelfsema, P., Sigman, M., 2011. The human Turing machine: a neural framework for mental programs. Trends Cogn. Sci. 15, 293–300, http://dx.doi.org/10.1016/j.tics.2011.05.007.
van Gaal, S., Ridderinkhof, K., Scholte, H., Lamme, V., 2010. Unconscious activation of the prefrontal no-go network. J. Neurosci. 30, 4143–4150, http://dx.doi.org/10.1523/JNEUROSCI.2992-09.2010.