Title

3.07 Emotion and consciousness

Authors

Winkielman, Piotr

Sher, Shlomi

Publication Date

2026

DOI

10.1016/b978-0-443-29258-3.00063-x

Emotion and Consciousness

Piotr Winkielman*

1. University of California, San Diego
2. SWPS University, Warsaw, Poland

&

Shlomi Sher

3. Pomona College

Running Head: Emotion and Consciousness

* To whom correspondence should be addressed.

Contact information: Department of Psychology, University of California, San Diego, 9500 Gilman Drive, Mailcode 0109, La Jolla, CA 92093-0109

pwinkielman@ucsd.edu

This is the final authors' version to appear in:

Consciousness: A Comprehensive Reference. (2026). Elsevier

Biographies

Piotr Winkielman is a Professor of Psychology at the University of California, San Diego. He attended the University of Warsaw in Poland and the University of Bielefeld in Germany for his undergraduate studies, completed his doctorate at the University of Michigan, and received postdoctoral training in social neuroscience at The Ohio State University. His research focuses on the relation between emotion, cognition, body, and consciousness using psychological and psychophysiological approaches.

Shlomi Sher is an Associate Professor of Psychological Science at Pomona College. He received a BA in mathematics from Harvard University and a PhD in psychology from Princeton University. He is interested in conceptual and experimental problems in the study of consciousness, knowledge, and rationality.

Glossary

1. **Affect**: The activation of valence-based representations across multiple systems. In dimensional models, valence is accompanied by arousal – the intensifying or energizing aspect of affect.

2. **Attention**: Selection of some stimuli over others for preferential processing.

4. **Attribution**: Determination of the source or cause of an observed event.

6. **Binding problem**: The problem of determining how features of any stimulus are integrated/combined together to form a unified and coherent percept of the stimulus.

7. **Blindsight**: An ability to guess, with above-chance accuracy, certain properties of visual inputs of which the subject is not consciously aware.

8. **Consciousness**: In experimental practice, "consciousness" is commonly taken to refer to representations of information with several linked properties, including: wide availability of information to cognitive systems and response modalities; flexibility of representation, learning, and response in novel circumstances; and expressed convictions about the subjective "reality" of the representation.

9. **Emotion**: Finely tuned variants of affect (e.g., anger, pride, disgust), motivating more specific reactions in more specific situations.

10. **Meta-consciousness**: Higher-order knowledge about one's own (past or present) conscious states.

11. **Subliminal stimuli**: Perceptual input that is not consciously accessible.

12. **Valence**: A single, basic encompassing dimension of evaluation, ranging from negative ("bad") to positive ("good").

Abstract

This contribution focuses on the relation between emotion and consciousness, drawing on key insights from psychology and neuroscience. We address current controversies about the neural basis of consciousness, and about the relationship between consciousness and attention, considering their potential implications for the scope of (un)conscious emotion. We review evidence that affective responses can be elicited by unconscious stimuli, and that affective responses can sometimes remain unconscious. Finally, we discuss studies of misattribution, in which people appear to misjudge the causes of their emotional states. Throughout we highlight the functional value of unconscious affective processing as well as the value of being conscious of one's emotion.

Keywords: Affect, attention, attribution, awareness, consciousness, emotion, meta-consciousness, unconscious emotion.

Key Points:

- Emotions are typically understood to extend beyond conscious feelings to include physiological and behavioral responses.
- Some theories of consciousness view emotions as necessarily conscious while other theories allow for unconscious emotion.
- Emotional responses can sometimes be elicited by stimuli that are not consciously perceived.
- Some components of emotional responses can sometimes remain unconscious.
- Attention can modulate affective processing with bidirectional interactions between allocation of attention and attention-grabbing emotionally salient stimuli.
- Individuals may not accurately identify or understand the causal relationship between stimuli and their emotional responses, leading to attribution errors.
- Both conscious and unconscious affective processes serve functional roles, supporting adaptive behavior in different contexts.

Introduction

This contribution discusses the relationship between emotions and consciousness. At first glance, this may seem like a puzzling topic. After all, in common-sense usage, emotions are feelings -- consciously experienced, reportable, affective states. In short, emotions are seen, almost by definition, as conscious. Relatedly, measuring emotions also seems straightforward -- all we need to do is ask. Indeed, in daily situations (at a doctor's office, in private conversations) as well as in many research studies, we ask, or are asked, "How do you feel right now?", "How strong is your pain?", "How intense is your anger?", or simply "Are you happy?" We answer these questions, usually quickly and competently – providing short summaries ("delighted!", "awful!", "hell no!"), using numbers on 1 to 10 scales, selecting the right "smile level", or richly expressing and elaborating on our sentiments in letters, songs, paintings, dance, and poetry. But do people always realize how they feel? What if they try but cannot access, properly articulate, or recognize their emotions? What if they make things up? A classic children's song goes: "If you're happy and you know it clap your hands". Can people be happy and not know it? Can you be angry and not realize it, until other people point out your own anger to you? Can you act out of fear, without ever subjectively feeling the fear itself?

In this chapter, we discuss some key concepts and illustrative research on the relationship between emotion and consciousness. The structure of our contribution is as follows. We start by providing a broad definition of affect and emotion, noting that most research sees emotions as more than "feelings," and includes other components in their

definition. Next, we offer a definition of consciousness, and briefly sketch some major theories of consciousness, highlighting their implications for emotion research. We then offer a short review of some key research topics. The review is organized around three major questions: (1) Are people always conscious of the stimuli that trigger their emotion states? (2) Are emotion states always conscious? And (3) even if people are conscious of both the eliciting stimuli and their emotion states, do they accurately understand their relationship? This structure allows us to highlight what we believe are some of the most interesting debates in this domain of research.

Definitions: Affect, Mood, and Emotion

"Affect" is usually taken to refer to an activation of valence representations (i.e., representations of goodness or badness). "Mood" refers to a valence representation that is relatively long lasting, less intense, and operates in the background of other mental activities as a broad modifying condition. "Emotion" refers to more finely tuned variants of affect, such as anger, sadness, or disgust on the negative side, and joy, pride, or awe on the positive side. These variants are more differentiated at the perceptual, cognitive, and physiological levels, motivating more intense, more specific reactions (screaming, weeping, rejecting, posturing) to more specific situations (injury, loss, contamination, triumph). We will sometimes use the terms "affect", "mood" and "emotion" interchangeably, but, as we note below, the consciousness requirements for specific forms of emotional processing may turn out to be more stringent than those for more general forms of affective response, and mood states. For example, whereas the onset of emotions is usually pretty sharp and often has identifiable external causes, the onset of of

moods might be more gradual, and due to a mix of multiple internal and external causes. We will return to this distinction later.

Importantly, even though affect and emotion are commonly thought of as feelings, researchers conceptualize them as orchestrated reactions simultaneously represented across multiple components (Mauss et al., 2024). Furthermore, researchers usually assume, and empirically observe, a fairly loose coherence, or coordination between: (i) feeling——changes in subjective experience; (ii) cognition——changes in attentional, perceptual, and inferential processes (appraisals); (iii) action——changes in the predisposition for or execution of specific responses; (iv) expression——changes in facial, vocal, or postural appearance; and (v) physiology——changes in physiological and neural activity. The idea of a "loose coordination", or weak coherence between different components, will become central later in this chapter, when we discuss cases where conscious, self-reported feelings may not capture the entirety of the emotion process.

A Working Definition of Consciousness

Definitions of "consciousness" vary widely in their ambitions and limitations, ranging from the simple (e.g., "what the subject reports") to the tautological (e.g., "what the subject experiences"). For our purposes, it will suffice to observe that some instances of human information processing, but not others, are characterized by a complex syndrome of inter-related properties: (a) widespread and coordinated availability of information to a broad range of response systems, including those involved in behaviors such as verbal reports or button-presses; (b) flexibility in dealing with and learning from novel situations; and (c) the reported feeling that the current experience is in some sense

"real". Some writers point out that the consciousness is often connected with the feelings of the self – a sense of having private, unique experiences or "me-ness" (see Yoshimi & Ford, 2025 in Further Readings). In exploring conscious and unconscious emotional phenomena, we will be content to contrast cases in which these properties are jointly and robustly present (which we will call "conscious") from cases in which these properties appear to be jointly absent (which we will call "unconscious").

While the above properties are often assumed, implicitly or explicitly, in research on consciousness, there is much debate about its neural underpinnings. Next, we sketch the three most popular approaches, which in turn suggest different answers to the question of relation between emotion and consciousness (see Paul et al, 2020 for a comprehensive discussion).

Major Theories of Consciousness and Their Relation to Emotion

Global workspace (GW) theory. This theory emphasizes that the key function of consciousness is coherent and integrated processing of information, and flexible response selection (Baars, 2002). GW theory posits that the brain has multiple isolated modules that can operate automatically, unconsciously, and in parallel. These unconscious modules suffice in well-practiced, familiar situations, but in novel situations, a subset of information enters large-scale neural workspace and is broadcasted across a multimodal network – so that it is consistently available to many modules. The GW broadcast constitutes the contents of consciousness. Information sharing in the GW enables processing that is integrated, coordinated, and flexible, but the narrow information bandwidth of the GW entails accompanying costs in speed and efficiency.

When it comes to emotion, the logic of GW theory suggests that unconscious affective processing can occur in well-practiced, familiar situations, or when there is a module or circuit dedicated to processing a specific type of affect-relevant content (e.g., snakes, spiders, heights, sexual stimuli, etc). Thus, for example, stereotyped rage reactions are observed in some patients with a vegetative state diagnosis, who are presumed to lack consciousness (Plum et al., 1998 in Further Readings). The logic of this theory would also suggest that affect can become conscious when a situation requires flexible, context-specific, and integrated control. Indeed, one advantage of *conscious* affect is the ability to un-link emotional stimuli with an immediate response, thus allowing for emotional control, delay in responses, and other sophisticated strategies. The saying "revenge is a dish best served cold", reflects this insight about the benefits of having a conscious awareness of being wronged and the resulting response flexibility, including the option to delay it and choose its most appropriate and effective form (Ludwig, 2025).

Integrated information theory (IIT). IIT shares GW theory's emphasis on integration but differs radically in other ways. For one thing, rather than treating consciousness as an on/off state, it sees it as a continuous variable (Tononi et al., 2016). IIT develops an intricate, and controversial, mathematical apparatus for its arguments, proposing a formal measure of integrated information (phi; roughly, the quantity of information in the *whole* system that cannot be reduced to information in its *parts*), with higher values of phi associated with "more" consciousness. Thus, IIT reframes the emotion-consciousness question, asking not whether emotions are conscious (and

whether they can be unconscious), but how conscious they are (and what is the minimal level of consciousness that can be associated with them).

A key feature of IIT is that it does not tie consciousness to a specific neuro-cognitive *function* (as do both GW theory and the higher-order theories we describe below), but with a highly general measure of information. Thus IIT allows, in principle, many counter-intuitive dissociations between function and consciousness. In principle, a process could be high in phi, hence in consciousness, even if it was associated with no interesting cognitive function at all. And conversely, to argue, within the IIT framework, that a cognitive or affective function requires a high level of consciousness, one must explain why that function can only result (at least in the human brain) from information processing that is high in the particular measure of information integration that IIT favors – a tall order indeed. These and other counter-intuitive properties of IIT have made it a highly controversial theory among consciousness researchers.

Higher order theories of consciousness. Like GW theory, higher-order theories link consciousness to a neuro-cognitive function. These theories distinguish between two levels of mental representation. One level is a first-order representation, such as a perceptual process representing, say, a snake, or a bodily activity, representing, say, an elevated heart rate. The other level is a second-order representation representing that first-order representation (e.g., a thought about a perception or a sensation). On this theory, a mental process is conscious only if it is the object of an appropriate higher-order representation (Brown et al., 2019 in Further Readings).

From the perspective of higher-order theories, the signature of conscious emotion should be some form of emotional metacognition (or "cognition about emotion"). This

may range from developing an internal map linking one's bodily states to one's cognitive beliefs to the self-application of a sophisticated theory of mind. While several different higher-order theories have been proposed (involving different assumptions about the nature of the relevant higher-order representations), higher-order theorists are likely to believe that, to be in a true emotional state (e.g., fear, joy), a subject needs to be able to engage in cognitive differentiation and some representation of the self (LeDoux & Brown, 2017).

In this way, higher-order theories have a natural kinship with Barrett's (2017) theory of constructed emotion, which also emphasizes the role of higher-level concepts in the construction of emotional experience. More specifically, Barrett proposes that the construction of emotional experience can be understood by apply the "predictive coding" framework, according to which the brain actively constructs predictive models of the internal and external world, continuously updated to minimize prediction error. On this proposal, emotion concepts can be understood as high-level predictions, built from past experience, that anticipate probable sensory input, action-outcome contingencies, and internal states. To take a simple example, in the context of an athletic event, the sensation caused by multiple touches might be constructively interpreted and felt as excitement, whereas in the context of a crowded bus, the same sensory stimulation may be interpreted and felt as anxiety or annoyance. In this view, the present conscious experience of emotion arises from the brain's self-application of predictive emotion concepts drawn from past experiences in a particular context.

In ascribing a high level of conceptual richness and flexibility to emotional experience, the constructivist perspective stands in contrast with several classical emotion

theories, such as (i) James-Lange theory, which states that emotions can be reduced to the simple perception of one's physiological states and behavioral reactions, (ii) Basic Emotions theory, according to which feelings represent outputs of mostly innate emotion programs, and even its closest historical cousin (iii) Two-Factor theory, which says that what we feel is basically a cognitive interpretation of non-specific (ambiguous) arousal. Instead, the constructivist perspective is broadly compatible with LeDoux and Brown's (2017) higher-order theory of emotional consciousness, which assigns complex conceptual operations, including self-referential concepts, a central role in the constitution of emotional experience. Given those conceptual commitments, it is perhaps unsurprising that these theories have met with some skepticism, especially among researchers investigating rudimentary affective and motivational processes in simple animals (Adolphs, 2016). We will return to those issues later in this contribution.

Can behavioral or neurological evidence arbitrate between theories of consciousness? Given the importance of theoretical frameworks in guiding progress in science – from designing key research questions to designing studies and interpreting data – it is unsurprising that the consciousness research community is keen on testing competing consciousness theories. However, a recent controversy regarding adversarial testing of GW vs. IIT theory has suggested to some that the task is difficult and perhaps premature (Cogitate et al., 2025). At any rate, no theory of the neural basis of consciousness can currently be regarded as established. Nonetheless, a number of fascinating neural and behavioral phenomena have been documented, which allow us to better understand the relationship between emotion and consciousness.

Empirical phenomena linking emotion and consciousness

To explore these phenomena systematically, we structure our treatment in three sections, which address three levels of consciousness which may or may not be associated with an emotion: (i) the level of the stimulus, (ii) the level of emotional reaction, and (iii) the relation between the stimulus and the reaction. The specific questions we will ask are: (1) Is the subject conscious of the stimulus that triggers the emotion? (2) Is the emotion itself conscious? And (3) if the subject is conscious of both the eliciting stimulus and the elicited emotion, are they aware of the connection between stimulus and emotion – that is, does the subject accurately attribute the emotion as effect to the stimulus as cause?

Affective processing of unconscious and unattended stimuli

A large literature has investigated affective reactions to stimuli of which the subject is unaware, or at least minimally aware, using methods drawn from cognitive psychology, neuropsychology, and neuroimaging. In a typical experiment using paradigms often called *subliminal (or masked) affective presentation*, the subject is (i) exposed to an unconscious but affectively significant "prime" stimulus (explained below), and (ii) overtly responds to a second conscious "target" stimulus, which is often unfamiliar and affectively neutral or ambiguous. The target could be, for example, an ideograph from an unknown language, a neutral face, or some decision problem (e.g., to accept or reject a gamble that could result in a moderate gain or a moderate loss). The subject's task is to form an overt evaluation of the target. For example, subjects could be asked to rate, on a numerical scale, the liking of the ideograph or the attractiveness of the neutral face. Subjects could also be asked to make a simple accept/reject decision or make a simple joystick movement towards or away the target. Sometimes, in

experiments focusing on physiological reactions, researchers may also measure the underlying response of the central (CNS) and/or autonomic (ANS) nervous systems. CNS reactions can be measured with techniques like EEG (electroencephalography), fMRI (functional magnetic resonance imaging), or single-cell recording. ANS reactions can be measured with techniques such as facial electromyography (measurement of activity of muscles involving in smiling or frowning), pupil dilation, skin conductance, respiration, or cardiovascular reactivity. Such techniques, often employed in combination, allow researchers to examine changes associated with valence (e.g., more smiling) or with the level of arousal (increases in pupil dilation and/or skin conductance).

The key feature of these experiments is that researchers attempt to "hide" the prime stimulus from entering consciousness – to make it "subliminal", "unseen" or "invisible." Importantly, some researchers are reluctant to make strong claims about absolute imperceptibility and use terms like "sub-optimal" or "near threshold". The hiding processes may involve different techniques, such as tachistoscopic presentation (extremely brief stimulus duration), binocular rivalry (presenting a stronger competing stimulus to the dominant eye), continuous flash suppression (in which a rapidly flashing high-contrast mask in one eye causes prolonged suppression of the stimulus in the other eye), peripheral presentation (outside the central focus of vision and attention), or masking, where the briefly presented prime stimulus is preceded and/or followed by a stronger stimulus at the same location (e.g., a scrambled visual noise mask). Using these techniques, numerous studies have found that subjects evaluate the target image more favorably, on average, when a hidden face that preceded it had a happy rather than an angry or fearful expression (for review, see Zhang & Winkielman, 2026). This occurs

even though, as indexed by subjective reports or objective forced-choice tests, the subject appears to be unaware of the faces (but see Stockart et al., 2025 for methodological challenges). As we will discuss shortly, invisible facial expressions have also been found to have effects on behaviors other than overt evaluation, including risk-taking in gambling situations and consumption behavior, and on a host of physiological measures of CNS and ANS activity.

Some of the most interesting evidence for the notion that affective properties of a visual stimulus can be unconsciously computed comes from studies on *blindsight*. These studies rely on the fact that extensive damage to the primary visual area of the cerebral cortex may leave a patient phenomenally blind over part or all of their visual field. Nonetheless, when forced to guess about some property of a visual stimulus presented in their blind field – e.g., "Is the line segment vertical or horizontal?" – such patients may make inspired guesses, a phenomenon known as "blindsight" – e.g., correctly guessing the orientation of the unseen line segment far over 50% of the time. Blindsight is a striking consequence of the multiplicity of visual pathways in the brain, allowing visual information to affect behavior even when the primary pathway of conscious vision is blocked. Interestingly, there are some cases of *affective* blindsight (for review, see Tamietto & Gelder, 2010). When images of emotionally expressive faces are presented to these patients' blind fields, they are sometimes nonetheless able to guess the affective valence of the faces. Furthermore, these patients show facial reactions, recorded using electromyography as well as arousal responses, measured with pupil dilatation. Neuroimaging studies of these patients have associated their inspired guesses about the image's affective properties with changes in amygdala activity. Parallel findings have

emerged from neuroimaging studies of healthy subjects, with higher amygdala activation upon exposure to emotionally charged – and especially to fearful – faces, even when these faces are not consciously perceived (Wang et al., 2023). Thus, at least some affect-related processing takes place for stimuli that, because of brain damage or brief, suppressed or masked stimulus exposures, fail to reach awareness (Tamietto & Gelder, 2010).

Affect, Attention, and Binding

One debated question concerns the role of attention in affective processing. Clearly, human perception is capacity-limited, such that allocating attention to one property, object, or spatial region impairs attention to and the processing of other properties, objects, or parts of the field. This limitation underlies several well-known phenomena in cognitive psychology. One classic phenomenon is *inattentional blindness*, illustrated by demonstrations in which subjects, whose main task is to track multiple people throwing balls, may fail to notice a gorilla appearing in the middle of the display. Another classic phenomenon is *change blindness*, illustrated by demonstrations in which subjects fail to detect large differences between one visual display and another. Although the interpretation of these experiments is controversial (e.g., Haun & Tononi, 2025 in Further Readings), some researchers take them to show that we have little, if any, conscious awareness of what we do not attend to (Cohen et al., 2012 in Further Readings).

Affective processing of perceptual input has often been conceived as *automatic* – i.e., proceeding independently of attention and cognitive strategies, perhaps along neural pathways distinct from those which culminate in consciousness of the visual and other

perceptual features of the input. Some evidence for the relative independence of affective processing from attention comes from studies in which affect-laden (especially unpleasant) stimuli appear to grab attention to themselves – or to resist experimental manipulations that tend to diminish attention – suggesting that their affective properties may, to some extent, be "preattentively" computed (though the evidence along these lines is somewhat mixed).

This research, arguing for affective processing without attention, generally adapts attention-related paradigms widely used in cognitive psychology, including visual search, "attentional blink", and attentional cueing tasks. However, there is clear evidence that affective processing can be modulated by attention. In fact, the relationships between affect and attention, and between affect and awareness, are two-way streets: Just as attention and conscious awareness may influence affective processing, so affective processing may change the probability that a stimulus will reach the threshold for awareness and whether our attention is drawn to it (Vuilleumier, 2005).

The relationship between emotion and attention has been extensively investigated using functional magnetic resonance imaging (fMRI). As noted above, increased amygdala activation to emotional (especially fearful or angry) faces has been found in a number of fMRI studies, using both supraliminal (above-threshold) and subliminal (below-threshold) faces. A key question addressed by studies concerns the degree to which this differential amygdala activation for emotional faces persists when the faces are unattended. In these studies, subjects see displays containing faces and other objects; in different conditions, the task requires subjects either to attend to and make a judgment about the faces, or to attend to and make a judgment about the other objects in the same

displays. The question is whether the observed amygdala response to the faces is reduced, eliminated, or otherwise altered when the eyes and/or attention are directed away from the faces. Though the debate is still ongoing, numerous studies have observed that allocation of attention can substantially modulate the amygdala's response to emotional faces – with some studies reporting that emotion-related amygdala activations disappear when attention is allocated to other items in the same displays (Pessoa, 2005). Still, it is not entirely clear how general these effects of attention are. The potentially relevant factors include: how demanding the distracting task is, with evidence that more demanding distracting tasks may more severely limit activation from unattended faces; the location of the face images in the visual field, which may influence the visual pathways through which the face information flows; the extent to which the tasks involve active suppression of irrelevant stimuli, as opposed to mere inattention; and the role of individual differences in different subject populations. It is important to remember that the amygdala activations revealed by fMRI offer only a crude measure of what the amygdala may or may not be computing. In addition, while these studies have focused on affective processing of face stimuli, different kinds of affective stimuli may show different kinds and degrees of attentional modulation. For example, faces and words differ in evolutionary significance, in visual properties such as spatial frequency, in how they are learned, and in their neural coding. Accordingly, attentional constraints on affective processing in the two cases may (or may not) be correspondingly different. Nevertheless, there is some evidence that potent emotional words can "grab attention." In one study, subjects saw a rapid stream of words appearing sequentially in the same location (Anderson & Phelps, 2001). Participants' task was to detect target words

presented in colored ink. The first word (T1) was always neutral (e.g., tree). The second word (T2) was either an emotional word (e.g., rape) or another neutral word (e.g., table). When the second neutral word (T2) appeared shortly after T1, its detection was heavily impaired, confirming an attentional "blink." However, when T2 was emotionally arousing, it was detected more often than a neutral T2. This suggests that emotional words are less susceptible to the *attentional blink*—they can partially overcome the limited capacity of conscious attention. Interestingly, this capacity for emotional words to break through attentional limits was reduced in a patient with damage to the amygdala, which plays a role in the early processing of salient stimuli.

A fascinating problem in the relationship between attention and emotion concerns binding. According to the feature integration theory, focused spatial attention helps solve the visual system's *binding problem* (Treisman & Gelade, 1980). In distributing your attention across a visual display, you may perceive the presence of different features of objects – redness, greenness, a circle, and a square. However, according to the theory, the formation of a coherent integrated percept – knowing that the circle is red and the square is green, rather than the other way around – requires focused attention. That is, attention focused at a location may be needed to bind the colors, shapes, and other visual features at that location together.

Insofar as affective properties can be processed in the absence of focused attention, feature integration theory raises two questions about the relationship between emotion and attention. First, when the affective value of a stimulus depends on the precise combination of the simple visual features in a display (e.g., a brown face next to a green leaf may elicit different affective reactions than would be sparked by a green face

next to a brown leaf), must the items be individually attended and their features globally integrated in order for affective responses to be formed? Or might affective processing proceed in part along a different pathway in which focused attention is not required for crude feature integration, at least for familiar stimuli? Second, how do we bind affective evaluations to the objects that elicit them? This corresponds to a more complex case of the problem of affective attribution, described below. In this case, multiple stimuli are present in the visual array, potentially triggering multiple distinct affective evaluations. The problem now is to decide which evaluation goes with which stimulus. Even if attention is not absolutely necessary in the formation of evaluations, it may (or may not) be necessary for their assignment among the different objects we consciously perceive.

One demonstration suggestive of failures of binding comes from studies using Continuous Flash Suppression (CFS) to eliminate conscious perception of emotional faces. These studies found that participants' ratings of how smiling vs. scowling a neutral face looked depended on whether subjects were shown suppressed happy or scowling facial expressions (Siegel et al., 2018). Such effects are sometimes explained as "affect misattribution" – an induction of an affective reaction by the prime and its transfer to the target. That is, participants could have conscious affective reactions to suppressed visual stimuli, despite not being conscious of any of their visual properties (including features related to identity, gender, age, expression, etc.). However, one alternative explanation of these results is that a visual feature "smile" from the suppressed priming face may have been conscious, but incorrectly bound to the visible neutral target face.

The just discussed findings are generally consistent with suggestions that conscious perception is heavily imbued with valence, even though the sources of valence

are not always known. This position is most associated with the work of experimental social psychologist Robert Zajonc (see Zajonc, 2000 in Further Readings). Intuitively, this view fits with a venerable tradition of thinking about emotions as modifiers of consciously perceived intrinsic quality of objects – making us perceive our own romantic partners as "objectively" attractive or food as "objectively" tasty when hungry. This affective guidance of conscious perception allows people to prioritize certain options, form quick likes and dislikes, in adaptive decision making (Cleeremans & Tallon-Baudry, 2022).

Conscious and unconscious emotion

The previous section argued that the external triggers of an emotion may not always be consciously represented. But what about the emotion state itself? Are emotions necessarily conscious? If not, how do conscious and unconscious emotions differ? To answer these questions, it is worth recalling that "emotion" is typically defined as a coordinated response to a significant event across several systems – perceptual, cognitive, motivational, expressive, bodily, and experiential (Grandjean, Sander, & Scherer, 2008). Therefore, the key empirical question in this area requires consideration of which components of the response are consciously accessible, and whether the consciously inaccessible components of a large-scale affective reaction can be activated in the absence of the consciously accessible components.

The importance of conscious, conceptual aspects of emotions is reflected in theories such as the Theory of Constructed Emotion (Barrett, 2017) and the Higher-Order Theory of Emotional Consciousness (LeDoux & Brown, 2017). These frameworks see complex emotion states, such as jealousy, envy, contempt, or pride, emerging via a

process of actively predicting, interpreting, and categorizing internal sensations, in the context of the current environment and the implications of this situation for the self. On these accounts, not every element of the process needs to be conscious -- explaining why, for example, a fit of anger, fear, or even jealousy may feel unbidden. However, the final outcomes of the process usually are, as they involve sophisticated cognition and some application of the model of the self.

These considerations may seem to argue for a necessary role for consciousness in constructively interpreted emotion. If emotions require sophisticated differentiations, how can they emerge from unconscious processing mechanisms? More critically, what about feelings? After all, what is anxiety without the subjective feeling of apprehension, worry, or impending doom, along with the bodily perception of trembling, sweating, and loss of control? What is loneliness if not the painful feeling of isolation and longing for social company? Similarly, it seems that the essence of love is the subjective feeling of care, concern, attachment, longing, and warmth towards another. Subjective feelings are also powerful motivators of behavior. It is hard to imagine Othello's jealousy or Iago's envy driving their intricate thoughts and elaborate actions without any consciousness of their emotions. Few clinical patients complain of unconscious anxiety or loneliness, and few people would take antidepressants, or recreational drugs if they only made them unconsciously happy. Unsurprisingly, attempts to map the structure and sources of conscious feelings play a central role in philosophy (phenomenology of emotion), experimental research on human emotion, clinical practice, and even biological psychiatry. In fact, pharmacological interventions into affective neurochemistry

primarily aim to address conscious emotional experience (e.g., benzodiazepines, or SSRIs for anxiety and depression) and their effectiveness is primarily measured via self-reports.

However, the idea that emotion can also be unconscious has a long history in emotion research, especially among scientists interested in biological foundations. Accordingly, basic behavioral and biochemical profiles associated with human emotion have significant overlap with those of non-human animals, including reptiles, fish, insects, and shrimp -- which presumably differ widely in the quality, and perhaps the presence, of consciousness. Critically, emerging research suggests that fine differentiation of qualitatively unique emotional behaviors (attachment, aggression, nurturing, mating) is controlled by subcortical circuits (Adolphs, 2016). Indeed, the most direct and effective neural manipulations of basic emotional reactions involve electrical or chemical intervention into subcortical structures. For example, in animal studies, brain microinjections of drug droplets that activate opioid receptors in subcortical nucleus accumbens cause increased “liking” responses for sweetness (Winkielman & Berridge, 2004). In bees, administration of dopamine enhances behaviors suggestive of an optimism bias (Solvi, Baciadonna, & Chittka, 2016). Importantly, even in humans, some basic affective processes do not depend on “higher-order” neural machinery. For example, in anencephalic infants, whose brains lack nearly all of the forebrain including the entire neocortex, sweet tastes still elicit positive facial expressions whereas bitter tastes elicit negative facial expressions. The causal dissociation of low-level affective machinery from conscious experience is highlighted by a recent study where direct stimulation of the amygdala led to physiological changes associated with emotions, such as increased heart rate, respiration, and skin conductance. In fact, among the eight

participants showing those changes, only one reported experiencing conscious feeling of fear (Inman et al., 2020).

In short, affective neuroscience highlights the role of subcortical structures in basic emotional reactions. For theorists who believe that consciousness has a cortical substrate, this raises the possibility that some elicitors of human emotion, and perhaps even some emotional reactions themselves, might not be accessible to full-blown conscious awareness.

Data from psychological studies with neurologically healthy subjects support this possibility. As noted earlier, affective responses can be triggered by stimuli of which the subject is unaware. But can the affective reaction itself be unconscious? Experimental evidence suggests that, at least under some circumstances, people are unable to report any shift in conscious feelings even as physiological indicators and behavior appear to reveal the presence of a covert affective reaction. For example, several researchers examined behavioral and physiological consequences of unobtrusive exposure to happy or angry facial expressions (for a review, see Zhang & Winkielman, 2026). Some studies report that such unobtrusive facial expressions can elicit physiological reactions suggestive of affect induction, such as congruent facial muscle responses, startle responses (indicating valence), or pupillary responses (indicating arousal). Intriguingly, these physiological changes are not aways accompanied by subjective feelings. This conclusion is suggested by findings that people do not report or rely on their subjective feelings even when they are told that using their subjective feelings might help them with a difficult perceptual task, such as detecting the valence of briefly flashed stimuli (Borenemann et al., 2012). This separation of physiological affective responses from subjective experience is

consistent with the studies described earlier, in which stimulation of subcortical brain regions, such as the amygdala, resulted in detectable physiological reactions but no changes in reported subjective experience (Inman et al., 2020).

But what about behavior? Several studies reported that unobtrusive faces can influence simple ratings, such as pleasantness ratings of images following presentation of covert faces, as well as simple decisions, such as risk taking (Winkielman et al., 2022). There is also some evidence that implies more active and elaborate behavior. Specifically, in studies by Winkielman, Berridge, & Wilbarger (2005), immediately after the subliminal elicitation of affect, participants reported their conscious feelings (valence and arousal) and also consumed and rated a novel beverage. The ratings of conscious feelings were unaffected by subliminal faces. However, thirsty participants consumed more beverage after happy rather than after angry faces, and rated the beverage more favorably. In short, these results suggest the possibility of genuinely unconscious affect, in the sense of a valenced (positive/negative) reaction that is strong enough to alter behavior and physiology, but of which people are not subjectively aware.

To be sure, many questions in the methodology, conceptualization, and mechanisms of unconscious emotion remain open. One big challenge is ensuring that participants are given a fair chance to report subtle and transient changes in feelings states. In addition to probing for neural substrates of conscious and unconscious emotional reactions, ongoing research in several labs is asking whether the critical property of unconscious affective states is simply positive-negative valence (unconscious affect), or whether there are unconscious states that drive behavior in differentiated fashion associated with specific emotions. There are several critical issues here. First,

the stimuli used in most "unconscious emotion" research typically represent basic emotions (fear, anger, disgust, happiness). Second, these stimuli are often restricted to a single category - expressive faces – a very common and thus highly practiced stimulus. This limitation is partly due to the ability to successfully standardize and mask face stimuli, as opposed to less common, more complex, but also more richly meaningful and potent pictures (e.g., scenes representing joy or social threats). Third, the observed behavioral and physiological reactions are often rather undifferentiated and can be characterized by a simple axis of approach-avoidance. Fourth, a particular empirical and conceptual challenge concerns the potential existence of "unconscious" guilt, compassion, pride, empathy, or, as mentioned earlier, jealousy and envy. These emotions almost "by definition" involve some high-level representation of the self and its complex relation to the physical or social environment. On the other hand, as shown by the work of John Cacioppo and colleagues, some seemingly complex social emotions, like loneliness, are grounded in basic biological, evolutionarily old mechanisms (see Cacioppo et al., 2014 in Further Readings).

Another open question concerns the exact mechanisms by which emotional stimuli change perception, behavior, and decisions. Returning to the earlier discussions on affect and attention, unconscious affect could change allocation of attention to congruent options (see Zhang & Winkielman, 2026). For example, unconscious positive affect could draw attention to previously rewarded, positive features. This could lead to changes in decision making, since attention to a given option can also increase its relative decision weight. This suggests a concrete mechanism by which unconscious affective stimuli can bias decisions without ever eliciting conscious feelings.

Thinking about feelings and their causes

The preceding sections of our contribution discussed some situations where either the affective stimulus or the affective reaction may not be represented in consciousness. However, one could reasonably argue that such situations represent relatively rare cases. After all, much of our emotional life reflects reactions to images that we consciously perceive, sentences that we clearly read or hear, and feelings that we very much consciously feel. This can range from the warm joy we feel upon reading a thank you card, to horror or outrage after news about a war or injustice. Still, even when the eliciting stimulus and the resulting emotion are both consciously accessible, it is an open question whether people properly understand the causal connection that links them. More generally, we daily keep ourselves busy diagnosing the causes of our emotional welfare and – especially – our emotional ills. How competent are we at such diagnosis? Do emotions emerge into consciousness tagged with their sources of origin? Or may affective causal attribution best be conceived as a complex matching problem in which we often guess and sometimes err?

A large body of research has repeatedly demonstrated that the task of affective attribution is not easy and can result in errors (for review, see Schwarz, 2012). The classic illustrations come from studies suggesting that prior induction of arousal (e.g., from crossing a suspension bridge, physical exercise, or a scary movie) may bias subsequent evaluative judgments via misattribution (e.g., making a potential romantic partner seem more enticing). Research on misattribution of anger suggest that sometimes people may lash out at "innocent" targets, which are mistakenly seen as the cause of the irritation. These and related findings are often interpreted as suggesting that people take

their current feelings as a valid source of information about the stimulus – unless an alternative, and more plausible, attribution for the arousal or affect is brought to the subject's attention. As such, under suitable experimental conditions, the attribution is subject to systematic misconstrual.

Here is it is also worth returning to the earlier distinction between affect, mood, and emotion. Recall that, compared to mood and general affect, emotion states are usually more intense, shorter, and more specific, often with identifiable causes. Some of those features may make emotion states less prone to misattribution. For example, it may be easier to mistake the causes of one's own broad negative mood (e.g., grumpiness or cheerfulness) than one's specific emotion of disgust or pride. On the other hand, it is often the intense, sharp, punctuated nature of emotion that gets people in trouble – a flash of anger, or joy, that spills over unsuspecting and innocent strangers. Another interesting distinction is that well-differentiated emotions may be more easily noticed, whereas moods may lurk in the background, coming in and out of metaconsciousness, as we discuss next.

Noticing, thinking, and talking about one's affect: Metaconsciousness, interoception, and emotion concepts

One framework in the literature argues for a tripartite distinction between the unconscious, the conscious, and the *meta-conscious* (Schooler, 2002). This distinction proposes that one can have a subjective experience without fully knowing that one is having it. Awareness that one is in an experiential state is then said to render that state (transiently) "meta-conscious". Importantly, "meta-consciousness" is not an all-or-nothing proposition, and can vary in articulateness, sophistication, and veridicality. For

example, while watching a movie in an old theatre, we may occasionally realize that we are feeling uncomfortable from sitting in a cramped seat. Yet, the feeling of discomfort may not fully appear in the center of attention, and thus never create a fully articulate, meta-conscious thought of "I am feeling uncomfortable". But if someone asked and altered us to the background feeling, we would probably then produce a verbal report of discomfort, suggesting the uncomfortable experience was there to begin with.

Another challenge of meta-consciousness is the related to the process of determining that one has a "feeling". As we discussed, according to several frameworks, a subjective experience (feeling) is to some extent "built" from the underlying components (Barrett, 2017; Grandjean et al., 2008; LeDoux & Brown, 2017). There might also be a decision process involved in determining if one has a feeling, and if so what kind of feeling it is. Not unlike other metacognitive decisions about one's own beliefs, this process may involve evaluation of evidence (for and against having a feeling) and some threshold for reporting a feeling (is the evidence sufficient to claim a feeling?). This perspective naturally raises the question of how one could assess an individual's correctness in assessing the presence of a feeling. Recently, some researchers have developed paradigms in which, they argue, non-normative emotional ratings of stimuli (i.e., ratings that diverge from those of most participants) show subtle but distinctive behavioral signatures of response error, rather than merely representing non-standard feelings (Basu & Srinivasan, 2026; Givon et al, 2022). This perspective suggests that, especially when the signals are subtle or weak, people could be wrong about how they feel – missing subtle existing feelings, and reporting non-existing feelings.

What makes "meta-consciousness" particularly interesting in the context of emotion is that it can modify the underlying affective state. This conclusion is suggested by an intriguing study where participants listened to Stravinsky's *The Rite of Spring* – an affectively ambiguous piece of music. Subjects who were asked to continuously rate their happiness while listening to the piece subsequently reported being less happy than other subjects. The authors speculated that focused hedonic monitoring may limit the attention devoted to the experience itself, and may diminish sensitivity to subtle and nuanced aspects of the experience which resist clear articulation. While further investigation is needed in this area, these results suggest that the intensity and direction of affective meta-consciousness can alter conscious affective states (see Schooler, 2002 for discussion).

More broadly, there is now intense research interest in the general phenomenon of interoception – awareness and articulation of internal bodily states (Critchley & Garfinkel, 2017). Examples of interoception include awareness of one's cardiovascular, digestive, respiratory, muscular, or glucose state. Research suggests large individual differences in people's interoceptive skills, though there is also large within-person variability in awareness between specific tasks. While there might be no one general interoceptive ability, some researchers have noted that low interoceptive skills are generally associated with poor psychological functioning. One fascinating question for future research is whether people can improve in their interoceptive skills, and whether this improvement will translate into better individual and social functioning. This is especially important given evidence that emotion awareness and interoceptive accuracy play key roles in social relations (Arnold, Winkielman, & Dobkins, 2019).

This relatively new focus on interoception resonates with a long tradition of research on emotion regulation—the ability to modulate (inhibit, amplify, delay, or transform) various components of emotion. Such regulation can target subjective feelings, cognitive appraisals, physiology, and behavior (verbal and bodily actions or expressions). While some of these regulatory processes may be spontaneous and automatized, others rely on conscious awareness. After all, to stop oneself from lashing out at an irritating driver, stop blaming an innocent stranger, or to amplify a positive feeling toward a friend, one needs to be able to access those feelings and reactions. This topic of emotion regulation also connects closely with the problem of emotional (mis)attribution, discussed earlier, where the task is to properly link our emotional experiences with their causes.

Finally, a related area of inquiry concerns the relationship between emotion, concepts, and language. A distinctive feature of emotion concepts is that they simultaneously capture bodily, sensorimotor states as well as higher-order abstract relations (Winkielman, Davis, & Coulson, 2023). Take, for example, the concept of "anger." It relates to bodily states—clenched fists, reddened faces, and a sense of "boiling inside." There is a real sense in which the concept of anger fits the classic idea of emotion as a perception of one's own bodily state. At the same time, the concept of anger includes thoughts that are usually directed at a specific target and reflect the conviction that the target has acted unfairly, thwarted an important goal, and is to blame for it. Interestingly, anger also often (though not always) involves the concept of the self—a belief that one has been unjustly thwarted and has the right to protest and retaliate. Without those specific beliefs, the same bodily states could mean something else (a

different emotion or no emotion at all). In short, the concept of "anger" is a hybrid concept—mixing concrete and abstract dimensions. Current research explores how underlying physiological states are shaped by emotion concepts, how these concepts develop, and how they differ across cultures. The hybrid nature of emotion concepts also highlights that when people misunderstand their feelings, they may have difficulty either perceiving and integrating physiological signals (bodily states) or interpreting them by applying the appropriate conceptual terms.

As such, one particularly important domain of research is psychopathology, including the study of individuals with alexithymia who have difficulty identifying, labeling, and discussing their feelings (see Smith & Lane, 2016, in Further Readings). Of course, the relationship between consciousness, concepts, and language is complex and contested, but these lines of inquiry help us better understand how people become aware of, conceptualize, report, and regulate their own emotions.

Summary

This chapter covered the relation between consciousness and emotion. We have highlighted the deep interrelations between these two phenomena – relations that become especially clear when considering our ability to attend, articulate, and regulate our affect. However, we have also reviewed some dissociations. Specifically, evidence suggests that affective states can be elicited by unconscious stimuli. Furthermore, affective states themselves may sometimes fail to reach full awareness. Finally, people may not always be fully aware of the relationship between the stimulus and the emotion – much as we rely on fallible inference in explaining the emotions of others, understanding our own

feelings may require inference from ambiguous introspective, biographical, and contextual data.

So, returning to our original question: Do people know how they feel? The research described here suggests that emotional self-knowledge involves a host of complex processes that may sometimes get things right, but are also subject to failures of access, articulation, and attribution. For students of emotion and consciousness, the seemingly simple questions of whether you know how you feel, and if so how you know, remain both fascinating and challenging.

References

1. Adolphs, R. (2016). How should neuroscience study emotions? By distinguishing emotion states, concepts, and experiences. *Social Cognitive and Affective Neuroscience, 12*(1), 24–31. https://doi.org/10.1093/scan/nsw153
2. Anderson A. K., & Phelps E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature, 411*, 305–309 10.1038/35077083
3. Arnold, A. J., Winkielman, P., & Dobkins, K. (2019). Interoception and social connection. *Frontiers in Psychology, 10*, Article 2589. https://doi.org/10.3389/fpsyg.2019.02589
4. Baars, B. J. (2002). The conscious access hypothesis: Origins and recent evidence. *Trends in Cognitive Sciences, 6*(1), 47–52. https://doi.org/10.1016/S1364-6613(00)01819-2
5. Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience, 12*(11), 1833–1833. https://doi.org/10.1093/scan/nsx060
6. Bornemann, B., Winkielman, P., & van der Meer, E. (2012). Can you feel what you do not see? Using internal feedback to detect briefly presented emotional stimuli. *International Journal of Psychophysiology, 85*(1), 116–124. https://doi.org/10.1016/j.ijpsycho.2011.04.007

7. Cleeremans, A., & Tallon-Baudry, C. (2022). Consciousness matters: Phenomenal experience has functional value. *Neuroscience of Consciousness, 2022*(1), niac007. https://doi.org/10.1093/nc/niac007
8. Cogitate Consortium, Ferrante, O., Górska-Klimowska, U., Henin, S., Hirschhorn, R., Khalaf, A., Lepauvre, A., Liu, L., Richter, D., Vidal, Y., Bonacchi, N., Brown, T., Sripad, P., Armendariz, M., Bendtz, K., Ghafari, T., Hetenyi, D., Jeschke, J., Kozma, C., … Melloni, L. (2025). Adversarial testing of global neuronal workspace and integrated information theories of consciousness. *Nature, 642*(8066), 133–142. https://doi.org/10.1038/s41586-025-08888-1
9. Critchley, H. D., & Garfinkel, S. N. (2017). Interoception and emotion. *Current Opinion in Psychology, 17*, 7–14. https://doi.org/10.1016/j.copsyc.2017.04.020
10. Grandjean, D., Sander, D., & Scherer, K. R. (2008). Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization. *Consciousness and Cognition, 17*(2), 484–495. https://doi.org/10.1016/j.concog.2008.03.019
11. Inman, C. S., Bijanki, K. R., Bass, D. I., Gross, R. E., Hamann, S., & Willie, J. T. (2020). Human amygdala stimulation effects on emotion physiology and emotional experience. *Neuropsychologia, 145*, Article 106722. https://doi.org/10.1016/j.neuropsychologia.2018.03.019
12. LeDoux, J. E., & Brown, R. (2017). A higher-order theory of emotional consciousness. *Proceedings of the National Academy of Sciences of the United States of America, 114*(10), E2016–E2025. https://doi.org/10.1073/pnas.1619316114

13. Ludwig, D. (2025). Functions of consciousness in emotional processing. *Consciousness and Cognition, 127*, Article 103801. https://doi.org/10.1016/j.concog.2025.103801

14. Mauss, I. B., Zerwas, F. K., Wilhelm, F. H., & John, O. P. (2024). Coherence of emotional response systems: Theory, measurement, and benefits. *Advances in Experimental Social Psychology, 69*, 59–149. https://doi.org/10.1016/bs.aesp.2023.11.002

15. Paul, E. S., Sher, S., Tamietto, M., Winkielman, P., & Mendl, M. T. (2020). Towards a comparative science of emotion: Affect and consciousness in humans and animals. *Neuroscience & Biobehavioral Reviews, 108*, 749–770. https://doi.org/10.1016/j.neubiorev.2019.11.014

16. Schooler, J. W., Mrazek, M. D., Baird, B., & Winkielman, P. (2015). Minding the mind: The value of distinguishing among unconscious, conscious, and metaconscious processes. In M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (Eds.), *APA handbook of personality and social psychology: Vol. 1. Attitudes and social cognition* (pp. 179–202). American Psychological Association.

17. Schwarz, N. (2012). Feelings-as-information theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 289–308). Sage. https://doi.org/10.4135/9781446249215.n15

18. Siegel, E. H., Wormwood, J. B., Quigley, K. S., & Barrett, L. F. (2018). Seeing what you feel: Affect drives visual perception of structurally neutral faces.

Psychological Science, 29(4), 496–503. https://doi.org/10.1177/0956797617741718

19. Solvi, C., Baciadonna, L., & Chittka, L. (2016). Unexpected rewards induce dopamine-dependent positive emotion-like state changes in bumblebees. *Science, 353*(6307), 1529–1531. https://doi.org/10.1126/science.aaf4454
20. Stockart, F., Schreiber, M., Amerio, P., Carmel, D., Cleeremans, A., Deouell, L. Y., … Mudrik, L. (2025). Studying unconscious processing: Contention and consensus. *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X25101489
21. Tamietto, M., & de Gelder, B. (2010). Neural bases of the non-conscious perception of emotional signals. *Nature Reviews Neuroscience, 11*, 697–709. https://doi.org/10.1038/nrn2889
22. Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience, 17*(7), 450–461. https://doi.org/10.1038/nrn.2016.44
23. Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5
24. Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences, 9*(12), 585–594. https://doi.org/10.1016/j.tics.2005.10.011
25. Wang, Y., Luo, L., Chen, G., Luan, G., Wang, X., Wang, Q., & Fang, F. (2023). Rapid processing of invisible fearful faces in the human amygdala. *Journal of*

Neuroscience, 43(8), 1405–1413. https://doi.org/10.1523/JNEUROSCI.1294-22.2022

26. Winkielman, P., & Berridge, K. C. (2004). Unconscious emotion. *Current Directions in Psychological Science, 13*, 120–123. https://doi.org/10.1111/j.0963-7214.2004.00288.x
27. Winkielman, P., Berridge, K. C., & Wilbarger, J. L. (2005). Unconscious affective reactions to masked happy versus angry faces influence consumption behavior and judgments of value. *Personality and Social Psychology Bulletin, 31*(1), 121–135. https://doi.org/10.1177/0146167204271309
28. Winkielman, P., Davis, J. D., & Coulson, S. (2023). Moving thoughts: Emotion concepts from the perspective of context-dependent embodied simulation. *Language, Cognition and Neuroscience, 38*, 1531–1553. https://doi.org/10.1080/23273798.2023.2236731
29. Winkielman, P., Trujillo, J. L., Bornemann, B., Knutson, B., & Paulus, M. P. (2022). Taking gambles at face value: Effects of emotional expressions on risky decisions. *Frontiers in Psychology, 13*, Article 958918. https://doi.org/10.3389/fpsyg.2022.958918
30. Zhang, Z., & Winkielman, P. (2026). How do faces influence behavior? A proposal for distinguishing between mechanisms that involve cognitive inferences, emotional feelings, and unconscious affective reactions. *Affective Science, 7,* 74–87. https://doi.org/10.1007/s42761-025-00350-9

Further Readings

1. Barrett, L. F., Niedenthal, P., & Winkielman, P. (Eds.). (2005). *Emotion and consciousness*. New York: Guilford.
2. Basu, S., & Srinivasan, N. (2026). Measurement of metacognition of emotional dimensions: A ROC based measurement method for metacognition of valence and arousal. *Frontiers in Psychology, 17*, 1761622.
3. Brown, R., Lau, H., & LeDoux, J. E. (2019). Understanding the higher-order approach to consciousness. *Trends in Cognitive Sciences, 23*(9), 754-768.
4. Cacioppo, J. T., Cacioppo, S., & Boomsma, D. I. (2014). Evolutionary mechanisms for loneliness. *Cognition and Emotion, 28*(1), 3–21. https://doi.org/10.1080/02699931.2013.837379
5. Clore, G.L. & Huntsinger, J.R. (2007). How emotions inform judgment and regulate thought. *Trends in Cognitive Sciences, 11*, 393-399.
6. Cohen, M. A., Cavanagh, P., Chun, M. M., & Nakayama, K. (2012). The attentional requirements of consciousness. *Trends in Cognitive Sciences, 16*(8), 411-417.
7. Givon, E., Udelsman-Danieli, G., Almagor, O., Fekete, T., Shriki, O., & Meiran, N. (2022). Can feelings “feel” wrong? Similarities between counter-normative emotion reports and perceptual errors. *Psychological Science, 33*(6), 948–956. https://doi.org/10.1177/09567976211063915
8. Haun, A. M., & Tononi, G. (2025). The unfathomable richness of seeing. *Trends in Cognitive Sciences, 29*, 892-902. doi: 10.1016/j.tics.2025.06.001

9. Kihlstrom, J. F., Mulvaney, S., Tobias, B. A., & Tobis, I. P. (2000). The emotional unconscious. In E. Eich (Ed.), *Cognition & Emotion*. New York: Oxford University Press.

10. Newen, A., & Montemayor, C. (2023). The alarm theory of consciousness: A two-level theory of phenomenal consciousness. *Journal of Consciousness Studies, 30*(3), 84–105. https://doi.org/10.53765/20512201.30.3.084

11. Ohman, A., Flykt, A., & Lundqvist, D. (2000). Unconscious emotion: Evolutionary perspectives, psychophysiological data and neuropsychological mechanisms. In R.D. Lane, L. Nadel & G. Ahern (Eds.), *Cognitive neuroscience of emotion* (pp. 296-327). New York: Oxford University Press.

12. Pessoa, L. (2022). The entangled brain: How perception, cognition, and emotion are woven together. The MIT Press.

13. Plum, F., Schiff, N., Ribary, U., & Llinás, R. (1998). Coordinated expression in chronically unconscious persons. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 353*(1377), 1929-1933.

14. Schooler, J. W., Mrazek, M. D., Baird, B., & Winkielman, P. (2015). Minding the mind: The value of distinguishing among unconscious, conscious, and metaconscious processes. In M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (Eds.), *APA handbook of personality and social psychology*, Vol. 1. Attitudes and social cognition (pp. 179-202). Washington, DC: American Psychological Association.

15. Smith, R., & Lane, R. D. (2016). Unconscious emotion: A cognitive neuroscientific perspective. *Neuroscience and Biobehavioral Reviews, 69*, 216–238. https://doi.org/10.1016/j.neubiorev.2016.08.013

16. Yoshimi, J., & Ford, J. (2025). The symphony of consciousness and the minimal experience of self. *Philosophy and the Mind Sciences*,6. https://doi.org/10.33735/phimisci.2025.11830 .

17. Zajonc, R. B. (2000). Feeling and thinking: Closing the debate over the independence of affect. In J. P. Forgas (Ed.), *Feeling and thinking: The role of affect in social cognition* (pp. 31-58.). New York: Cambridge University Press.