Emotion and Consciousness

S Sher and P Winkielman, University of California, San Diego, CA, USA

Glossary

Affect – The activation of valence-based representations reflecting the organism's assessment of 'goodness–badness' across multiple neural/psychological systems.

Affective priming – Experimental procedures in which subjects are exposed to valenced stimuli ('primes') and their effects on responses to simultaneously or subsequently presented other stimuli ('targets') are measured. For example, exposure to subliminal angry face primes may, on average, lower evaluations of the attractiveness of subsequently presented shapes or willingness to try a novel drink.

Attention – Process(es) resolving capacity limits in perceptual and cognitive processing, in which some stimuli or ideas are in effect 'selected' over others for preferential processing and/or awareness.

Attribution – An implicit or explicit determination of the source or cause of an observed event, that is, in the present context, a mood or emotional state.

Binding problem – When inputs are analyzed separately along multiple feature dimensions, the problem of knowing which features go with which is known as the binding problem. In visual perception, solving the binding problem allows one to perceive a yellow sun against a blue sky, rather than the other way around. In affective processing, the problem is to determine which affective reactions go with which eliciting stimuli.

Blindsight – An ability to guess, with above-chance accuracy, certain properties of visual inputs of which the subject is not consciously aware. Blindsight has principally been explored in patients with damage to primary visual cortex, though there have also been attempts to demonstrate blindsight in normal subjects. Similar phenomena (e.g., 'deaf hearing,' 'numb touch') in other sensory modalities have also been investigated.

Consciousness – Heated argument surrounds all attempts at definition; arguably indefinable. 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.

Emotion – Finer-tuned, qualitatively differentiated variants of affect (e.g., anger, disgust), motivating more specific reactions in more specific situations.

Implicit measures – Measured effects of external stimuli or internal states on physiological reactions or behaviors not involving explicit description of those stimuli or states. Implicit and explicit measures may partly or entirely tap into different pathways of information processing, and correspondingly may at times yield discrepant pictures of the subject's psychological state.

Metaconsciousness – Higher-order knowledge about one's own (past or present) conscious states. Metaconscious awareness comes in many shades and varieties; it may involve simple awareness that one is in the conscious state one is in, or more sophisticated beliefs about the relations between the conscious state and other states and stimuli.

Neglect – A neurological condition in which, owing to brain damage on one side of the brain, the patient shows a marked deficit in attention to and awareness of objects on the

opposite side of perceptual space. This spatial neglect may manifest in multiple reference frames. For example, the left half of visual space, or the left halves of objects located in both halves of visual space, may be neglected. Neglect often resolves over time into a phenomenon known as 'extinction,' in which objects on the neglected side are missed only when another object is simultaneously presented to the nonneglected side.

Subliminal – Refers to perceptual input, which, because of its presentation parameters, is not consciously accessible – that is, the subject would be unable to report explicitly about its properties if he or she tried. Subliminal stimuli can be delivered in different modalities (sight, hearing, touch, etc.), and subliminality is usually achieved by some combination of low stimulus intensity; very brief stimulus presentation; and forward or backward masking, in which another 'masking' stimulus, nearby in space and time, helps to obliterate awareness of a briefly exposed stimulus. Subliminality is generally verified in psychological experiments by subjective tests (e.g., "Did you see any face?") and/or objective tests in which subjects guess about some property of the stimulus (e.g., "A face was flashed – guess its gender").

Valence – A single, simple encompassing dimension of evaluation, ranging from negative ('bad') to positive ('good').

Introduction

"How do you feel right now?" is a question we often ask (sometimes meaningfully) and answer (sometimes honestly). Yet our answers to this particular question, unlike other questions, rarely elicit the rejoinder: "How do you know?" "Do you know?" is another question that might reasonably be asked, but usually is not. This article summarizes and comments on empirical research that attempts to seriously pose these unorthodox questions: How do you know how you feel right now. . .if you know?

Put in more sophisticated, if perhaps less revealing, terms, we describe research on the relationship between emotion (or affect) and consciousness. 'Affect' is usually taken to refer to the simultaneous activation of valence representations (i.e., representations of goodness or badness) across multiple neural/psychological systems. 'Emotion' refers to more fine-tuned variants of affect (anger, sadness, contempt, guilt), which motivate more specific reactions (screaming, weeping, turning up one's nose, apologizing) to more specific situations (injury, loss, perceived inferiority of another, transgression). We will sometimes use the terms 'affect' and 'emotion' interchangeably, but, as we note below, the consciousness requirements for specific forms of emotional processing may conceivably be more stringent than those for more general forms of affective response.

Definitions of 'consciousness' vary widely in their pretensions and limitations, ranging from the austere but simplistic (e.g., 'what the subject reports') to the evocative but 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 interrelated properties: (1) widespread and coordinated availability of information to a broad range of response systems, including those involved in verbal reports, button-presses, etc.; (2) flexibility in dealing with and learning from novel situations; and (3) the reported feeling that the current experience is in some sense 'real.' 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'). To explore these phenomena systematically, our treatment is structured into three sections, which address three different levels of consciousness which may or may not be associated with an emotion.

Emotional states need not be triggered by simple stimuli in simple ways. But when they are, we can pose three questions about the relationship between emotion and consciousness: (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, is he or she 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, using methods drawn from cognitive psychology, neuropsychology, and neuroimaging. In a typical experiment using a paradigm called 'subliminal affective priming,' developed by Robert Zajonc and his colleagues, the subject looks at a screen on which an unfamiliar and affectively ambiguous target shape – for example, an ideograph from an unknown language – is presented. The subject's task is to form an overt evaluation of the target – for example, to rate the attractiveness of the ideograph on a numerical scale. Unbeknownst to the subject, however, a photograph of a face is very briefly flashed on the screen just before the target shape appears at the same location, replacing it. Under the right timing parameters, the target shape has the effect of 'masking' the briefly flashed face – that is, preventing it from entering consciousness. Nonetheless, numerous studies have found that subjects evaluate the target shape more favorably, on average, when the subliminal face that preceded it had a happy rather than an angry or fearful expression. Despite the fact that, as indexed by subjective reports or objective forced-choice tests, the subject appears to be unaware of the faces, the facial expressions are found to bias evaluations of subsequent target shapes. As discussed in the section 'Conscious and unconscious emotion,' subliminal facial expressions have also been found to have effects on behaviors other than overt evaluation, including risk-taking in gambling situations and consumption behavior.

Evidence from affective blindsight further corroborates the notion that affective properties of a visual stimulus can be partly computed in the brain even when the subject is not conscious of the stimulus. 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 – for example, "Is the line segment vertical or horizontal?" – such patients may make inspired guesses, a phenomenon known as 'blindsight' – for example, correctly guessing the orientation of the unseen line segment far over 50% of the time. Recently, some cases of 'affective' blindsight have been reported. When images of emotionally expressive faces are presented to these patients' blind fields, they are sometimes nonetheless able to guess, with high accuracy, the affective valence of the faces. 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 normal subjects, with higher amygdala activation upon exposure to emotionally charged – and especially to fearful – faces, even when these faces are not consciously perceived.

Thus at least some affect-related processing takes place for stimuli that, because of brain damage or brief stimulus exposures, fail to reach awareness. An important related question concerns the role of attention in affective evaluation. It is widely accepted that human perception is subject to steep capacity limits, such that enhancing processing in one part of the perceptual field – by allocating attention to one object or spatial region – impairs, in certain respects, the processing of other parts of the field.[1] But there has long been controversy about the relationship between attention and awareness – and in particular about the fate of nonattended perceptual input. Is attention to a stimulus necessary for conscious awareness of that stimulus? Further, can attention influence processing of an unconscious stimulus? Cognitive experiments on change blindness – in which subjects are oblivious even to gross changes between

[1]It is useful to distinguish between attention to objects or locations, on the one hand, and attention to different properties of a fixed object at a fixed location, on the other. The studies summarized here involve the allocation of attention to different objects or locations, though in some cases these shifts of attention are confounded with shifts of attention between different kinds of properties.

one visual display and another – and inattentional blindness – in which subjects, attending to one part of a display, seem to have no awareness of a stimulus that pops up in another part of the display – have been advanced as evidence that we have no consciousness of what we do not attend to. However, the interpretation of these results, as well as the larger question of the relationship between attention and awareness, remains hotly disputed. Furthermore, whether or not awareness itself requires attention, recent behavioral and neuropsychological studies in cognitive psychology suggest that attending to a location influences the processing of stimuli at that location, even when these stimuli remain unconscious.

Affective processing of perceptual input has often been conceived as 'automatic' – that is, 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).[2] 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, evidence that affective processing can be modulated by attention has been accumulating.

The relationship between emotion and attention has recently come under intensive focus in neuroimaging experiments using functional magnetic resonance imaging (fMRI). As noted earlier, 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 recent 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 attention is directed away from the faces. (Measures are generally taken to ensure that shifts of attention are not accompanied by movements of the eyes.) The picture emerging from these studies is still a somewhat complex work in progress, with conflicting results between different studies using different attention tasks. But some general conclusions can be drawn. While several studies failed to find any attentional modulation of affect-related amygdala activation, numerous other studies have observed such modulations. In one influential study by Luiz Pessoa and colleagues, none of the affect-related activations found when attention was focused on happy or angry faces survived when attention was allocated to other items in the same displays. The results of this and a number of other neuroimaging studies indicate that attention can modulate amygdala responses to emotional faces under some conditions, and sometimes dramatically – but it is not entirely clear how general these effects of attention are.

This developing literature is actively seeking to identify factors that will explain and thus reconcile the different results found in different experiments. These 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 (e.g., people with

[2] In general, it is important to note that the relationships between affect and attention, and between affect and awareness, are two-way streets: Just as attention and 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.

anxiety disorders). It is important to remember that the levels of amygdala activation revealed in an fMRI study supply only a crude measure of what the amygdala may or may not be computing. Other studies have used event-related potentials, recorded at the scalp, to try to separate out the effects of attention on different components of the brain's response to emotional stimuli. 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; attentional constraints on affective processing in the two cases may (or may not) be correspondingly different. In any event, while the roles of the various factors described above remain to be fully sorted out, it has by now repeatedly been shown that, under suitable conditions, the allocation of attention can substantially modulate the amygdala's response to emotional faces.

Another relatively underexplored problem in the relationship between attention and emotion concerns affective binding. According to Anne Treisman's feature integration theory, one essential function of spatial attention is to solve the visual system's 'binding problem.' In distributing your attention across a visual display, you may perceive that redness, greenness, a circle, and a square are all present. 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 – may require 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.

Finally, a fascinating study by John Marshall and Peter Halligan illustrates the possibilities for some forms of affective processing when attention and awareness are jointly impaired. These researchers studied a 'neglect' patient who, as a consequence of right-hemisphere brain damage, persistently failed to attend to left visual space and to report awareness of objects appearing there. This patient also had a partial left visual field defect, but the experiment was conducted under free viewing conditions in which the patient, by moving her head, could put 'left space' (e.g., the left half of an object) into her right visual field. The patient was repeatedly shown two house drawings, vertically arrayed. In one of the drawings, flames emerged from a window on the left side of the house, but the two drawings were otherwise identical. While the patient, ignoring the left-side flames, judged the two houses to be exactly the same, she almost always decided (when forced to make what struck her as an entirely arbitrary choice) that she would prefer to live in the house without the flames. A burning house is not an appealing place to live, apparently even when the flames are neither attended not fully perceived.

Taken together, the research summarized above indicates that affective processing of a stimulus can take place in the absence of awareness of that stimulus. Nonetheless, while it is unlikely that focused attention is necessary for all forms of affective processing, such processing can be,

under conditions which have yet to be fully mapped out, strongly affected by, and perhaps dependent on, attention.

Conscious and Unconscious Emotion

In the previous section we considered whether the external triggers of an emotion need to be consciously represented. But what about the emotional response 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 valenced event across several components – perceptual, cognitive, motivational, expressive, bodily, and experiential. Therefore, the key empirical question in this area is which components of a large-scale affective reaction can be activated, but remain unconscious.

For many writers, the most critical element of emotion is the intrinsically conscious, subjective, experiential component. In other words – the essence of feelings is 'the feeling.' Indeed, it is initially difficult even to think about anxiety without conjuring the phenomenal experience of apprehension, worry, loss of control, or impending doom, along with the subjective sense of trembling and sweating. Similarly, it appears that the very essence of love is the subjective feeling of care, attachment, and warmth toward another. Conscious feelings are also powerful motivators of behavior. Few clinical patients complain of unconscious anxiety or depression (though their partners might). Similarly, few people would drink or take recreational drugs if they made them only unconsciously happy or relaxed. Othello's poignantly conscious jealousy poisons his mind and eventually drives his destruction, whereas Romeo is motivated by conscious love and compassion. Unsurprisingly, understanding conscious feelings plays a central role in both research and clinical practice. Thus, emotion researchers have spent many years delineating the various meaning dimensions of conscious feelings (e.g., appraisal and attribution theories), whereas philosophers have explored their phenomenological structure (e.g., Husserl, Brentano, Sartre, Solomon). On the practical end, psychiatrists have focused on pharmacological interventions into affective neurochemistry that alter conscious experience (e.g., benzodiazepines like Valium and SSRIs like Prozac).

However, the idea that emotion can also be unconscious has a long history in emotion research. It goes back at least to Darwin, who described many 'instinctive' (fast, rigid, involuntary) emotional behaviors and speculated about their origins in our remote evolutionary ancestry. Early prototypes of complex emotion presumably evolved to spur appropriate reactions to positive or negative events. Accordingly, many basic behavioral reactions associated with human emotion are widely shared by animals, including reptiles and fish. The evolutionarily old neurocircuitry and neurochemistry underlying basic emotional reactions (fear, liking) is wired into subcortical brain structures, such as the amygdala, nucleus accumbens, hypothalamus, and even brainstem parabrachial nucleus and pons. Indeed, the most direct and effective neural manipulations of basic emotional reactions involve electrical or chemical intervention into subcortical structures. For example, Kent Berridge and colleagues showed that brain microinjections of drugs that activate opioid receptors in subcortical nucleus accumbens elicit increased 'liking' responses for sweetness. Importantly, these effects do not depend on 'higher-order' neural machinery as they occur even in decorticated animals. Similarly, 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. In short, affective neuroscience highlights the role of subcortical structures in basic emotional reactions. This raises the possibility that some causes of human emotion, and perhaps even some emotional reactions themselves, might not be accessible to full-blown conscious awareness.

Data from psychological studies with normal subjects support this possibility. As mentioned in the section 'Affective processing of unconscious and unattended stimuli,' there is now extensive evidence that affect, and perhaps even emotion-like states, can be triggered by stimuli of which

the subject is unaware. But can the emotional reaction itself be unconscious? Experimental evidence suggests that, at least under some circumstances, people are unable to report any shift in conscious emotion even as a consequential behavior appears to reveal the presence of a covert affective reaction. For example, in one series of studies by Winkielman and colleagues, participants were unobtrusively exposed to several subliminal happy or angry facial expressions. Immediately after the subliminal elicitation of affect, participants reported their conscious feelings (mood and arousal) and also consumed and rated a novel beverage. The ratings of conscious feelings were unaffected by subliminal faces. However, participants consumed more beverage after happy rather than after angry faces, and rated the beverage more favorably. Not only was their overt behavior indicative of affective change, but follow-up studies using psychophysiological measures, such as affective startle and facial EMG, revealed that responses of low-level approach/avoidance systems were influenced in an affect-congruent way by the subliminal faces. 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, there remain many open questions about the conceptualization and mechanisms of unconscious emotion. In addition to probing for neural substrates of conscious and unconscious emotional reactions, ongoing research in several laboratories is examining whether the critical property of unconscious motivating states is simply positive–negative valence (unconscious affect), or whether there are unconscious states that drive behavior in differentiated fashion associated with specific emotions (fear, anger, disgust, sadness, etc). It is also worth exploring the tricky possibility that what sometimes presents as 'unconscious emotion,' as suggested by the failure of emotion self-reports, might sometime represent a failure of constructing or updating an appropriate higher-order self-description of an emotional state (i.e., a problem with meta-awareness, as we discuss shortly). It will take further research with clever designs to address these possibilities. But for now, it seems likely that at least in some conditions not only the processing of emotion triggers, but also to some extent emotional responding, may unfold without reaching full awareness.

Thinking about Feelings and Their Causes

When an emotion and the stimulus that elicits it are both consciously accessible, it is an open question whether the subject will appreciate the causal connection that links them. The broken air conditioner in the museum may suffice to explain the art critic's vague discomfort, but what will prevent the critic from attributing this unease to the paintings on display? 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? (The affective influence of unconscious stimuli, summarized above, would suggest that this could not universally be the case.) Or may affective 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 problem of affective attribution is not solved flawlessly. Researchers commonly employ 'mood manipulations' – uncomfortable temperatures, cramped postures, sad music, etc. – to modulate subjects' overall mood states in more or less subtle ways. Just as the scowls of unseen faces depress attractiveness ratings for subsequent novel shapes (see section 'Affective processing of unconscious and unattended stimuli'), so broad shifts in mood, caused by consciously accessible stimuli and conditions, tend to bleed into subjects' affective evaluations of different objects that subsequently come within the focus of attention. The prior fear-related induction of arousal may make a potential romantic partner seem more enticing, and on relatively gloomy days respondents tell survey researchers that their overall life satisfaction is relatively low.

To be sure, such effects need not automatically be interpreted in terms of defective attribution. A grey day may render more salient the greyer aspects of one's existence. However, there is

abundant experimental evidence that such mood manipulation effects can derive from, or at least be strongly modulated by, active processes of attribution. For example, in a classic study by Norbert Schwarz and Gerald Clore, a telephone interview about life satisfaction was conducted. As noted above, survey respondents give higher life satisfaction ratings when they are contacted on sunny than on gloomy days. However, when the weather was mentioned explicitly by the interviewer just before the life satisfaction question was asked, life satisfaction ratings were no longer affected by ambient gloom. Schwarz and Clore interpret this finding in terms of their 'feelings-as-information' model, according to which affect experienced while attending to an affectively ambiguous stimulus is, by default, assumed to convey information about the stimulus – unless an alternative, and more plausible, attribution for the affect is brought to the subject's attention. Life satisfaction is a particularly nebulous concept, which takes on different forms when viewed from different perspectives, and hence can accommodate radically discrepant interpretations under different conditions. According to the model, when the lousy weather outside is explicitly highlighted before respondents are asked about life satisfaction, they attribute the lousy feelings they are currently experiencing to the weather and hence not to the disappointments that taint their life satisfaction. Note that, in this model, the subjective feeling of affect is itself assumed to be information-bearing: it is not taken to be a mere inert correlate of some other underlying computation. But the information the feeling carries is not obvious, and it is subject to systematic misconstrual under suitable experimental conditions.

However, such misconstrual should be viewed as the exception rather than the rule. According to the feelings-as-information model, after all, feelings are informative. The use of one's present affective state for default attribution is a 'heuristic' (similar to others studied in research on judgment and decision making), a rough but useful rule-of-thumb that generates reasonable judgments under typical conditions. For instance, if you feel vaguely uneasy when you enter a new apartment, you might be better off, on average, renting a different one. Because conscious deliberation is slow, effortful, and often relatively insensitive to fine-grained nuance, deferring by default to less informationally transparent but faster and more sensitive affective reactions may often, many researchers believe, be a wise general policy. Indeed, some evidence from experimental studies of neurological patients suggests that impairments in affective processing may lead to systematically poorer choices in 'rational' choice domains like decision making under risk.

It is important to note that a given situation potentially raises multiple problems of affective attribution at multiple levels of processing, and the extent to which their solutions overlap is a nontrivial question. The attribution 'implicit' in a classical conditioning experiment – where learning systems may (as John Garcia demonstrated) 'blame' recently ingested food, rather than a red flashing light, for presently experienced stomach discomfort – need not dovetail, in process or output, with the attribution a subject makes 'explicit' in filling out a questionnaire. To more fully capture the complexity of typical real-world attribution problems, imagine an experiment (1) involving a subtle unpleasant mood manipulation, (2) in which multiple affectively ambiguous items are simultaneously exhibited to the subject, and (3) in which multiple probes – some involving explicit measures like ratings of attractiveness, others involving implicit measures like skin conductance response (SCR) or heart rate – are employed to assess affective reactions to each of the stimuli presented. Affective reactions and attributions may or may not run in parallel across the different levels of processing revealed by different kinds of probe. Even if the subject consciously decides upon a single primary source of current mood – this painting, perhaps, and not that sculpture, is the main culprit – we cannot assume that implicit measures of affect like SCR will draw the same distinctions – with, for example, a raised SCR for the painting, subsequently viewed, but not for the sculpture. The ways in which processes of considered affective attribution may interact, or fail to interact, with less consciously accessible and controlled implicit processing of affective stimuli is an important question in research on attribution. At present, researchers are actively contesting the role of conscious awareness in affective conditioning – in which an initially neutral

'conditioned stimulus' inherits the affective properties of an attractive or aversive 'unconditioned' stimulus with which it is paired. While there is substantial diversity of opinion, some researchers believe that effective conditioning of affect requires consciousness of the conditioned–unconditioned pairing, which would imply that 'higher-level' representations can feed down to influence seemingly uncontrolled responses.

A related limitation of the preceding discussion is that it artificially separates the process of causal attribution from the causal process for which the attribution is being made. That is, we have presupposed the existence of a separate and prior causal relation between external stimulus and internal affect; holding this relation fixed, we then ask whether the subject's judgment of casual attribution corresponds with this actual relation. In this way, the subject's causal judgments about his or her affective states are taken to be just as isolated from those affective states as they would be if the subject were instead making causal judgments about the origins of another person's affective states. This idealized division and comparison of the causal judgment and the causal process judged is often useful. However, when the affective state which the subject contemplates is ongoing, it is a misleading simplification. Because our various mental states intimately and dynamically impinge on and interact with each other, beliefs about our own states can readily become self-fulfilling or self-defeating. Explicit affective attributions fall within this class of problematic beliefs. An initial affective attribution may proceed to make itself true or make itself false: now that I have attributed my present gloominess (set initially into motion by the overcast sky overhead) to this odd black sweater, the sweater itself may begin to make me feel gloomy. This may now feed back into and reinforce or otherwise modify the process of attribution, rendering it partly accurate. Furthermore, in addition to the outcome of an explicit attribution process, the mere process of stepping back to explicitly think about one's affective states may have significant effects on those states.

It is important not to overstate this important point. There are, in particular, striking and well-documented parallels between the ways people respond to cues about their own affective states and the ways people respond to cues about the affective states of others. The self-fulfilling, -defeating, and otherwise-altering effects of explicit affective judgment should not be viewed as a chaos that completely engulfs and disfigures the normal process of attributive inference. Nonetheless, the process and outcome of affective judgment can have significant effects on the affective states being judged.

Jonathan Schooler has argued for a tripartite distinction between the unconscious, the conscious, and the 'metaconscious.' This distinction derives from the claim that one can have a subjective experience without knowing that one is having it. Awareness that one is in an experiential state is then said to render that state (transiently) 'metaconscious.' On this view, the transition from consciously being in state X to being in a state where one also metaconsciously knows that one is in state X requires additional computational steps that are only intermittently executed and are potentially fallible. It is important to appreciate that 'metaconsciousness' may come in many varieties and can vary in scope and intensity – it is not an all-or-nothing proposition. There is a distinction, while experiencing a red image, between (1) having an inarticulate awareness of one's experience; (2) privately articulating to oneself (in English) "Here, I see red"; and (3) explicitly entertaining a sophisticated causal account of the experience (e.g., 'this present experience is the product of light of a certain wavelength reflecting off a surface with certain properties, impinging my retinas, and triggering a specific cascade of axon potentials in my brain, associated with this experience'). While certain problems of interpretation may arise at the far extreme – where dim metaconsciousness needs to be differentiated from none – the concept is useful over a broad range, where metaconscious contents (thoughts referring to current or past conscious states) can patently vary in articulateness, sophistication, and veridicality.

As noted earlier, explicit metaconscious attributions of affective content ("My present annoyance is at the smug expression on your face") can feed back and modify both the affective state and the process of affective attribution that explains it. In addition, Schooler and his colleagues have reported some evidence suggesting that merely engaging in

metaconscious processing can alter the affective state itself. In one study, their subjects listened to Stravinsky's *The Rite of Spring* – a striking and jarring piece of music, which the researchers assumed would be affectively ambiguous. 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 (just as reflecting on elusive qualities of a fine wine may disrupt our perception and memory of those very qualities). While further investigation is needed in this area, this and other evidence – for example, a number of studies showing that mood inductions have altered effects in the presence of a mirror, which presumably encourages the subject to more closely monitor his or her own overt reactions – indicates that the intensity and direction of affective metaconsciousness can systematically alter conscious affective states.

Summary

This article has described multiple dissociations between consciousness and emotion. Stimuli of which we are unaware can elicit affective states of which we are aware. In some cases, the elicited affective states, while systematically modifying our physiology and behavior, may fail to reach full awareness. Furthermore, even when both the eliciting stimulus and the emotion are conscious, we may be unaware of the relationship between them. Yet, while affect can in these ways be multiply dissociated from awareness, affective reaction and conscious awareness do not occupy distinct and hermetically sealed compartments, our conscious selves looking on ineffectually as mere observers of our approaches and avoidances, only able to guess what and why. Rather, our deliberate allocation of attention can strongly modulate the affective effects of valenced stimuli, and the process and products of metaconscious reflection about affective experience can partly reshape that experience.

Or, returning to our original question: How do you know how you feel right now. . .and do you know? The research described here shows that these questions are not just perversely skeptical. You may not always know how you feel or what causes you to feel that way. And your knowledge of how you feel derives at least in part from general-purpose patterns of reasonable but fallible inference from imperfectly informative cues. Nonetheless, what you think you know with regard to how you feel, about what, and why, can exert a profound influence on how you feel, about what, and why.

Suggested Readings

Barrett LF, Niedenthal P, and Winkielman P (eds.) (2005) *Emotion and Consciousness.* New York: Guilford.

Berridge KC (2003) Pleasures of the brain. *Brain and Cognition* 52: 106–128.

Clore GL and Huntsinger JR (2007) How emotions inform judgment and regulate thought. *Trends in Cognitive Sciences* 11: 393–399.

Kihlstrom JF, Mulvaney S, Tobias BA, and Tobis IP (2000) The emotional unconscious. In: Eich E (ed.) *Cognition & Emotion.* New York: Oxford University Press.

Marshall JC and Halligan PW (1988) Blindsight and insight in visuo-spatial neglect. *Nature* 336: 766–767.

Murphy ST and Zajonc RB (1993) Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology* 64: 723–739.

Ohman A, Flykt A, and Lundqvist D (2000) Unconscious emotion: Evolutionary perspectives, psychophysiological data and neuropsychological mechanisms. In: Lane RD, Nadel L, and Ahern G (eds.) *Cognitive Neuroscience of Emotion*, pp. 296–327. New York: Oxford University Press.

Pegna AJ, Khateb A, Lazeyras F, and Seghier ML (2005) Discriminating emotional faces without primary visual cortices involves the right amygdala. *Nature Neuroscience* 8: 24–25.

Pessoa L (2005) To what extent are emotional visual stimuli processed without attention and awareness? *Current Opinion in Neurobiology* 15: 188–196.

Schooler JW (2002) Re-representing consciousness: Dissociations between experience and meta-consciousness. *Trends in Cognitive Sciences* 6: 339–344.

Schooler JW, Ariely D, and Loewenstein G (2003) The pursuit and assessment of happiness can be self-defeating. In: Brocas I and Carrillo JD (eds.) *The Psychology of Economic Decisions,* vol. 1, pp. 41–70. Oxford: Oxford University Press.

Vuilleumier P (2002) Facial expression and selective attention. *Current Opinion in Psychiatry* 15: 291–300.

Winkielman P and Berridge KC (2004) Unconscious emotion. *Current Directions in Psychological Science* 13: 120–123.

Winkielman P, Berridge KC, and Wilbarger JL (2005) Unconscious affective reactions to masked happy versus angry faces influence consumption behavior and judgments of value. *Personality and Social Psychology Bulletin* 1: 121–135.

Zajonc RB (2000) Feeling and thinking: Closing the debate over the independence of affect. In: Forgas JP (ed.) *Feeling and Thinking: The Role of Affect in Social Cognition*, pp. 31–58. New York: Cambridge University Press.

Biographical Sketch

Shlomi Sher is a postdoctoral scholar in psychology at the University of California, San Diego. He received his BA in mathematics from Harvard University and his PhD in psychology from Princeton University. He is interested in conceptual and experimental problems in the study of consciousness, perceptual attention, rationality, and moral judgment.

Piotr Winkielman is a professor of psychology at the University of California, San Diego. After undergraduate study at the University of Warsaw and the University of Bielefeld, he received his PhD in social psychology from the University of Michigan and postdoctoral training in social neuroscience at the Ohio State University. His current research explores the relation between emotion, cognition, body, and consciousness using psychological and psychophysiological approaches. His research has been supported by the National Science Foundation and National Alliance for Autism Research and he has served as an associate editor of *Psychological Review* and *Emotion*.