The Journal of Mind and Behavior
Autumn 2008, Volume 29, Number 4
Pages 291–320
ISSN 0271–0137

The Two-Stage Model of Emotion and the Interpretive Structure of the Mind

Marc A. Cohen

Seattle University

Empirical evidence shows that non-conscious appraisal processes generate bodily responses to the environment. This finding is consistent with William James's account of emotion, and it suggests that a general theory of emotion should follow James: a general theory should begin with the observation that physiological and behavioral responses precede our emotional experience. But I advance three arguments (empirical and conceptual arguments) showing that James's further account of emotion as the *experience* of bodily responses is inadequate. I offer an alternative model, according to which responses (physical states) are perceived and *interpreted* by a separate cognitive process, one that assigns meaning to those responses. The non-conscious appraisal process and the interpretive process are distinct, hence a two-stage model of emotion. This model is related to Schachter and Singer's two-factor theory. Their often-discussed experiment showed that interpretation can play a role in producing emotions. But they do not show that interpretation is necessary for producing emotions in general, outside of the experimental conditions that generated unexplained arousal in subjects. My two-stage model supports this stronger claim by situating the interpretive process in a comprehensive model of emotion.

Keywords: emotion, William James, self-interpretation

In this paper I outline a theory of emotion, one that begins with William James's claim that physiological responses *precede* emotional experience. James relied on introspective evidence to argue for that claim, and recent research in experimental psychology supports his contention: for a set of emotions usually referred to as affect programs or basic emotions, physiological responses are generated by non-conscious, reflex-like appraisals in lower parts of the brain, prior to and without any emotional experience. There is good

The author thanks Gary Hatfield, Jeff Helmreich, the Editor of JMB, and an anonymous reviewer for valuable suggestions. Requests for reprints should be sent to Prof. Marc A. Cohen, Department of Management, Seattle University, 901 12th Avenue, Seattle, Washington 98122–1090. Email: cohenm@seattleu.edu

reason to generalize this account to include emotions that require cognitively richer appraisals. And as a result, the fundamental philosophical question about emotion concerns the relationship between these responses, emotions, and first-person experience.

I argue that emotions should not be identified with responses. But James's view, that emotions are the *experience* of responses, is also inadequate. It must be replaced by an account in which a second cognitive process — one that is distinct from the appraisal process — interprets responses and ascribes meaning to them. Emotions according to this two-stage model are physiological responses experienced as being about some situation in the environment. The two components, the response and the interpretation, are bound together in a single, conscious emotional state.

The argument proceeds in stages. In the first section, I summarize recent research on affect program emotions, which shows that non-conscious, reflex-like cognitive processes generate the (so-called) fear response. Paul Ekman's research on facial expressions offers support for extending this account from fear to other emotions. I then offer three arguments to support my claim that, in addition to the response-generating appraisal, an interpretive process is needed to produce an emotion: an empirical argument about emotional responses not being distinct; a conceptual argument about the source of intentionality that distinguishes emotions from other, purely physiological responses; and an argument that appeals to T.D. Wilson's (2002) more general account of the human mind, in particular, to Wilson's work showing that our first-person access to non-conscious cognitive processes is indirect — meaning interpretive and inferential.

In the second section I situate my model in relation to Stanley Schachter and Jerome E. Singer's (1962) well known paper, "Cognitive, Social and Physiological Determinants of Emotional State." Schachter and Singer injected subjects with adrenaline to cause a response; the subjects then "labeled" and "shaped" their responses and experienced emotions. Schachter and Singer take this to show that bodily states can be interpreted when subjects lack an appropriate explanation, but they offer no reason to think that a subject would be put in that same position — the position of having to interpret her responses — outside of the experimental setting. The two-stage model proposed here attempts to fill this gap by situating Schachter and Singer's point about the role played by interpretation in a broader, comprehensive model.

In the final section I show how my account can resolve a difficulty Michael Stocker raised about cognitive or judgment-oriented approaches to emotion, a criticism that is particularly pointed when applied to Schachter and Singer's account. And I outline the possible role played by consciousness in forming emotions.

The account of emotion presented in this paper, along with the broader theory of mind it implies, amount to a plausible reconstruction of Charles Taylor's constitutive thesis, his claim that our understanding of our emotions is constitutive of those emotions (see Taylor, 1985c, p. 101). This thesis underlies his conception of humans as self-interpreting animals, and it is an application of his more general claim, that:

> We are not simply moved by psychic forces comparable to such forces as gravity or electromagnetism, which we can see as given in a straightforward way, but rather by psychic "forces" which are articulated or interpreted in a certain way. (Taylor, 1985b, p. 36; see also Taylor, 1985c)

I do not have space for a systematic discussion of the relationship between Taylor's project and my two-stage model (for a more detailed discussion, see Cohen, 2002), but I want to note this relationship explicitly. My account differs from Taylor's in many respects, most notably in my appeal to empirical research. But the two-stage model of emotion preserves Taylor's insight, namely that our conscious life is made possible by meaning-infusing interpretations of our own actions and reactions. And my account amounts to an independent argument for that insight.

Enactivist accounts of consciousness distinguish their own approach from perceptual conceptions of consciousness. On the enactivist approach, consciousness organizes cognitive processes into patterns and manages goal-directed behavior (Ellis and Newton, 2002; Newton, 2000). On the perceptual conception, conscious states are reactive states of awareness that constitute a passive subject's experience. Wilson's work on the adaptive unconscious — discussed below — suggests that sophisticated, goal-directed behavior is possible without conscious processing, so the enactivist view seems to go too far: consciousness isn't necessary for sophisticated, goal-directed behavior. But my account of the binding that takes place in consciousness is closely related to the enactivist approach because I take emotions to be a particular kind of meaningful state created in consciousness.

The Two-Stage Model of Emotion

Affect Program Responses and Emotions

As noted, recent empirical work shows that responses precede emotional experience for affect program emotions.[1] Joseph LeDoux (1996) summarized one line of this research concerning fear: in rats conditioned to fear tones, the

[1]For a more extended discussion of the affect program approach and the problems surrounding it see Cohen (2005).

auditory thalamus directly triggers the action of the amygdala, which then triggers physiological and behavioral responses (like increased heart rate and freezing, respectively). Note here that LeDoux's conception of responses is broader than autonomic change; the terms "responses" and "physiological responses" are used in this broader sense throughout this paper (this broader sense is consistent with James's use; see footnote 9; and see the comment on behavior in the first part of the next section). LeDoux argues that the parallel organization of brain structure in humans — pre-conscious, reflex-like processes that detect, appraise and trigger responses to markers of danger — is an "evolutionary relic" (p. 163): the direct neural pathway in humans from the thalamus (which receives sensory input) to the amygdala (which triggers physiological changes) is a remnant of the way less-developed brains were organized. Although these neural structures lack the capacity to make fine distinctions among stimuli, this neural organization nevertheless serves (or served) a useful function in mammals and in humans: a system depending on cortical processes would slow an organism's ability to react. For this reason, because of this function, LeDoux describes this neural structure as a "quick and dirty system" (1996, p. 63; see also pp. 163–165).

LeDoux takes fear to be the prototype of the affect programs, and relying on the evolutionary argument outlined in the previous paragraph he generalizes the claim, suggesting that distinct modules involving the same kind of pre-conscious appraisal will generate other affect program emotions (see 1996, pp. 126–128). LeDoux sets aside the task of identifying the affect program emotions; lists vary but usually include happiness, sadness, fear, surprise, anger, and disgust.

Paul Griffiths (1997) follows LeDoux and offers a more precise characterization of the affect program modules, each as "a system akin to a reflex in its encapsulation and mandatory operation" (p. 94). These modules consist of: (i) a biased learning mechanism and/or a set of preprogrammed responses (like the disgust response, which seems to be present from birth); (ii) complex and invariant outputs that come as a unit — like the response to danger; and (iii) a direct, involuntary and non-conscious coordination between the appraisal and the production of the response (p. 94). Talk of appraisal in this context could be misleading: the appraisals taking place in these modules involve only the detection of salient features in the environment, a process best described as minimally-cognitive.[2] (I will broaden the account below, incorporating more cognitively rich forms of appraisal.)

This sort of psychological program could be distributed across a set of neural structures, and so on Griffiths' conception it is an open question whether

[2]This point is not intended as an argument *against* cognitivism with respect to emotions; cognitivism is ambiguous as to the requirement that appraisals be conscious. Zajonc's (1980, 1984) well-known experiments on the exposure effect support my characterization of this appraisal process as unconscious.

these programs will be easily identified with discrete neural structures. The relationship between these programs and the neural structures realizing them is not at issue here.

Paul Ekman's work on facial expression offers support for generalizing the affect program account of fear to a broader set of emotions. Ekman (1980; see also Ekman, Friesen, and Ellsworth, 1982) showed that there are pan-cultural expressions for a number of emotions, and he confirmed Darwin's suggestion that these expressions are shared with other animals.[3] Ekman explains these findings by arguing the following: facial expressions are components of emotional responses; the functions served by these responses proved to be adaptive for our evolutionary ancestors, and so the responses — including the facial expression component — were preserved over the course of the development of humans. I have challenged this account of facial expression elsewhere (Cohen, 2002, 2005), but the facts about facial expressions are consistent with the claim that there is a set of emotional responses each produced by a distinct affect program, that is, by reflex-like appraisals prior to and without the input of our conscious experience of emotion. Ekman identified pan-cultural expressions for five emotions, happiness, anger, sadness, fear, and disgust, suggesting that at least these five should be classified as affect programs. To be clear: Ekman's work predates LeDoux's, but logically, I begin with LeDoux's work because he identified the neural structures at work in generating the fear response; beginning there, Ekman's work on facial expression offers circumstantial (and evolutionary) reason to expect parallel structures for other affect program emotions.

Research on affect programs proceeds by identifying emotions with responses, and in the process this research sets aside conscious, first-person experience. The reasoning at work is this: the appraisal process is plausibly described as pre-conscious because of its location in the brain, and if we do not need to appeal to conscious feelings to understand how an emotion interacts with the world — to understand its function, causal history and causal efficacy — then conscious experience can be set aside as (potentially interesting but) merely epiphenomenal. LeDoux suggests this position (1996, p. 125). And Ekman makes this point explicitly; he argues, "The subjective experience of emotion, how each emotion feels, is for some at the center of what an emotion is . . . . [T]his is excluded because too little is known about how subjectivity maps on to other aspects of emotional experience" (1992b, p. 175). Moreover, the reliance on animals in empirical research reinforces this approach.[4]

---

[3]This body of work has been criticized for relying on English emotion terms and forced choices from among a limited set of terms used as translations. For a concise overview of this line of criticism see Barr–Zisowitz (2000). For more detail, see Russell (1994) and Haidt and Keltner (1999).

[4]This treatment of conscious experience as epiphenomenal is not peculiar to research on emotion; see Epstein and Hatfield (1994, esp. p. 170) on a lingering strand of behaviorist thinking in contemporary cognitive psychology and philosophy.

James's view fits neatly here (see James [1890/1981, chap. 25] and [1892/1984, chap. 24]). He challenged the (supposedly) common sense understanding of emotion, according to which the experience of fear, to use his most-cited example, generates a physiological response, like an increase in heart rate. Instead, James argued, the perception of some relevant stimuli directly produces a response, and the emotion is the subsequent conscious experience of the response.

Where James described physiological responses as following the direct perception of a stimulus, he is best understood as suggesting that there is some discrimination of information here, and so a perceptual process that can be described as cognitive in this limited sense. In the *Principles of Psychology*, he distinguishes between sensation and perception, in that the latter involves further cognitive processing: "The fuller of relations the object is . . . the more it is something classed, located, measured, compared, assigned to a function, etc., etc.; the more unreservedly do we call the state of mind a perception, and the relatively smaller is the part in it which sensation plays" (1890/1981, p. 651).[5]

Given this point about the role played by appraisals, for James emotional *experience* is best seen as the product of two cognitive processes, the first of which detects salient features in the environment and generates an appropriate physiological response, after which a second process perceives the effects of a response on one's body. To the degree that they are concerned with emotional experience, philosophers and psychologists working on affect program emotions could adopt this view, and extend their account to include emotional experience. My two-stage model is a further development of James's view.

Before proceeding, note that James's view is often dismissed in a routine way. Regardless of how the finer points are developed, James seems to get the intentional object of an emotion wrong: he claims that fear is the experience of the physiological changes that take place in my body when I see a bear in the woods. This seems implausible because on this account fear is about my body and not about the situation (the bear).[6]

---

[5]Ellsworth (1994) makes this point in order to correct oversimplified readings of James's position, in part relying on a James (1894/1994), a later and not-often cited paper. James himself intended that paper to clarify his view but without noticeable effect. This point may conflict with Hatfield (2007), who emphasizes the role of instinct in James's view: for James the perceptions involved in original emotions are instinctual *as opposed to cognitive*. On my approach, these instincts are nevertheless cognitive in the minimal sense described in the main text, because there is some discrimination of information. On this point I may already be departing from James.

[6]For example, in a recent review, McGinn (2003) dismisses Damasio (2003) as a restatement of James's position. McGinn's (supposed) refutation of all Jamesian views, Damasio's included, turns on the criticism outlined in the main text, that they assign the wrong intentional object to emotions, the body as opposed to some object or situation in the surrounding environment. I take James to assume that responses have the appropriate — outward directed — intentionality, and he could reply to McGinn by claiming that this intentionality is transferred in the experience of a response. Whether or not James would have endorsed this line of thought, his

footnote continued on next page ⟶

Whether or not this criticism cuts against James, against my view the criticism lacks traction: on the two-stage model, the process of interpretation assigns a meaning to a physiological response, a meaning that explains and makes sense of the response against the background of the situation taken to have caused it ("taken to have caused it" because mistakes are possible here). Because the emotion is constituted by the assignment of aboutness to a response, the emotion is not about the response. Instead, the emotion is the response understood in a certain way and the resulting subjective experience; this experience is emotional as opposed to being purely physical, like the experience of being cold. This amounts to a departure of James's view, according to which emotions just are the experience of responses, without this additional meaning component. Note that nothing about this model implies that an emotion cannot cause other mental states or play a broad causal role in our lives; in short, the two-stage model is not non-cognitive.

The Second Stage as Interpretive

My discussion of research on (so-called) affect programs was intended to show that some process in addition to the initial appraisal is needed to generate emotions. Three separate lines of argument support my characterization of the second stage as interpretive.

1. *Responses are not differentiated.* There is a prominent stream of research directed at identifying distinct responses across the set of affect program emotions. The most often-cited study in support of such differentiation was conducted by Paul Ekman and collaborators (Ekman, Levenson, and Friesen, 1983). But this study is at best suggestive because Ekman observed only partial differentiation across emotions. He used two techniques to elicit emotions: subjects were instructed to form facial expressions associated with specific emotions, and they were asked to think about an emotional experience of a certain sort. In the first, the facial expression task, Ekman was able to distinguish the responses associated with happiness, disgust, and surprise as a group from those associated with anger, sadness, and fear: the second group was characterized by larger increases in heart rate. And within that second group, the anger response could be distinguished from the responses associated with sadness and fear by changes in skin temperature. In order for the differentiation to

---

view does seem open to the problem discussed in the main text, that we cannot assume that intentionality is present in giving causal accounts of physiological responses. McGinn offers another line of criticism against both James and Damasio, asking why, if emotions are just the experience of bodily states, isn't the awareness of one's body position and temperature an emotion? This criticism cuts against Damasio for two reasons, because he assimilates emotions to homeostatic processes more generally, and because Damasio is not clear about the relationship between responses and emotions. Note that this second criticism does not threaten my account: we do not experience all bodily states as emotions because we do not ascribe meaning to our awareness of them all.

be complete, however, further discriminations are still required; at this point Ekman appealed to the re-lived emotion task, in which he was able to distinguish between the responses associated with sadness and fear using skin resistance. Ekman therefore claimed to have provided a complete differentiation of anger, sadness, and fear, though one that still leaves happiness, disgust, and surprise, undifferentiated.

Even though Ekman and others (including Griffiths, 1997) take the study to suggest that responses are differentiated, the study actually shows that discrete emotions can occur in the absence of autonomic differentiation. This finding is well supported in the literature on differentiated responses. Cacioppo, Berntson, Larson, Poehlmann, and Ito's (2000) meta-analysis of 22 papers reached this conclusion. And Ekman's own study makes an especially powerful case for this point: Ekman asked his subjects to rate the intensity of their emotional experiences, so he could limit his investigation to the cases in which subjects actually experienced emotions. But he found that his subjects had strong emotional experiences without distinguishable autonomic responses, which suggests that a differentiated autonomic response is not intrinsic to or a necessary part of an emotion.

Moreover, using the re-lived emotion task, Ekman failed to find any consistent differentiation between happiness, disgust, and surprise as a group and anger, sadness, and fear as another. So, again, his own experiment supports the claim that differentiation across responses is inconsistent.

To be sure, responses include changes other than autonomic ones, and a more complex study focusing on other measures *could* identify distinct responses. But at this point there is no empirical basis for thinking responses are differentiated. And, further, Richard J. Davidson's (1993) work explains why we should not expect to find differentiated responses: on the basis of empirical work he argues that there are only two basic responses across the affect program emotions, approach and withdrawal, each located in different parts of the brain.[7]

If this is the case, if autonomic responses are not differentiated, then James's account of emotional experience is inadequate. On his view, as described above, the experience of physiological and behavioral responses is the emotion, and so his account seems to depend on there being perceptibly distinct responses for different emotions — emotions feel different from one another because they are the perception of different (and distinct) patterns of arousal.[8]

[7]On the point that approach/withdrawal tendencies are the fundamental adaptive advance, Davidson cites Toobey and Cosmides (1990), who emphasize the role of emotions in sorting stimuli into categories.

[8]Ellsworth (1994) criticizes Jamesian views for reifying emotions, for treating emotions and emotion-components as things as opposed to processes. She seems, on my reading, to confuse two points: we could emphasize the fluid nature of our emotions and emotional experience, and the interpretations that go along with them, but nevertheless take those to be constitutive of emotions and see the underlying cognitive processes as just that, underlying processes.

Without distinct responses, James, and with him advocates of basic emotions, cannot explain why emotions feel different.[9] If, for example, the physiological responses accompanying happiness and sadness are the same, we cannot explain why those emotions feel different in terms of the perception of their responses. And the fact that different instances of fear can occur with different physiological responses leads to the same conclusion. Both examples suggest that a cognitive process other than the perception of physiological changes must be at work.[10]

In short: Ekman's evidence about responses not being distinct across emotions presses us to abandon James's account of the second stage, his claim that emotions are the experience of these responses. At a minimum, this line of argument suggests that another cognitive process will be involved, one that can bridge the gap between indistinct arousal and particular kinds of emotional experience.

2. *Responses lack intentionality.* A more conceptual argument supports the characterization of the second cognitive process as interpretive. As noted, work in philosophy and psychology on affect programs identifies emotions with responses: emotions just are responses, with our conscious experience set aside as a separate phenomenon or problem, one that may or may not be interesting depending on one's perspective.

But this identification is not defensible. Responses lack the intentionality that distinguishes emotions from other bodily states, where I am using intentionality in the strong sense of *experienced aboutness*. To see this, notice that there is no basis for distinguishing between the response to danger and the physiological response to extreme cold: both serve a clear function; both involve autonomic and physiological changes (patterns of blood flow in the body, changes in heart rate); and both trigger action-readiness and changes in muscle tone/activity. And both sets of physiological changes are equally meaningless.

---

[9]James is not at pains to make these distinctions precise. He resists associating an emotion with a "sacramentally or eternally fixed" response, and he makes disparaging remarks about the project of categorizing emotions as of secondary importance to a general theoretical account (1890/1981, p. 1069; see also 1892/1984, p. 331). That said, James does not take emotion to be an undifferentiated class of experience, nor does he limit his attention to visceral responses (on this last point, again see Ellsworth, 1994). And I do not think James would object to my line of argument about differentiated responses. In both the *Principles* and *Briefer Course* he notes, "The various permutations of which these organic changes [the ones associated with emotional responses] are susceptible make it abstractly possible that no shade of emotion should be without a bodily reverberation as unique, when taken in its totality, as is the mental mood itself" (1890/1981, p. 1066; 1892/1984, p. 328). Deigh's (1994) discussion on this point is helpful.

[10]Note that the lack of differentiation across responses is problematic for another reason, separate from the point about different kinds of emotional experience: each emotional response serves (or served) a different function, and so the responses should differ accordingly. See Cohen (2005).

To be sure, instances of both responses will be intentional in the weaker sense of involving mental representations, but this sense is inadequate when applied to our *experience* of emotion, as about situations. There is also a more general implication here: we cannot assume that intentionality, in the stronger sense of experienced aboutness, is present in a mental state just because we can give a causal history of that mental state. This line of thought could be controversial; it could be taken to suggest that perceptual states are not intentional in the stronger sense in virtue of their causal history. Whether or not this is the case, and whether or not my view has that implication, the kinds of physiological responses at issue in this paper are not representational states and so my point should apply in the context of emotions without controversy.

Therefore, because responses lack the necessary intentionality, (i) the response to danger is not an emotion or even emotional — unless we are willing to classify the experience of being cold as an emotion, which would amount to giving up the category, and would amount to giving up the characterization of emotion as intentional. And (ii) the experience of fear, to take one case, requires some psychological process in addition to the one that generates the response to danger.

The change in terminology is important here: identifying the physiological response as a response to danger, not as fear, is intended to prevent the philosophical move that takes place without argument or explanation (or even notice), namely the treatment of the response as itself already emotional — in this case, as already about or part of fear — just because danger is present. But responses should be thought of as pre-emotional: only when meaning is ascribed to them are responses experienced as about something, and only then can they become emotions.

This conceptual argument suggests that the second cognitive process is best characterized as interpretive and meaning-ascribing. In short, a two-stage model of emotion can account for the physiological responses and also for the ascription of meaning to those felt responses — meaning that a two-stage model can account for our experience of emotions. Emotions, on this model, arise from interaction between the sensation of bodily states and cognitive processes that are, at least in part, about those sensations. The production of an emotion could therefore be described as having four stages: an initial appraisal, the (usually unconscious) sensation of a physiological response and its effects on one's body, a second cognitive process that interprets and ascribes meaning to physiological responses, and the resulting emotion (the experience). But only two of these stages are of interest as cognitive processes — the initial appraisal and the subsequent interpretation of it. For this reason I refer to the account offered here as a two-stage model.

Two points of clarification about the process described here. First, the interpretations and beliefs ascribed to responses could be present before those

responses. Moreover, the person's desires and preparation for action could constitute a background of meaning against which responses and environmental factors are interpreted (these could also constitute a background against which initial appraisals are made). Allowing preparation for action to play this sort of role does not suggest that the action itself is sufficient for an emotion (a point I return to in a moment), nor does it suggest that overt action is necessary — imagined actions could play the same role. And more broadly, I do not mean to suggest that the responses have a simplistic stimulus–response structure. The response could be caused or partially caused by factors internal to the subject.[11]

James is typically read as a non-cognitivist with respect to emotions, but Prinz (2003, 2004) appeals to contemporary work in informational semantics, arguing that content can be ascribed to bodily reactions in virtue of the relationship between those responses and their environmental triggers. The responses can therefore be seen as indicators that those triggers are present. So, responses can be described as cognitive at least in that sense, and as a result the perception of responses can interact with other cognitive states — transforming James's account into a bodily *and* a cognitive account of emotion. But the argument just presented applies again: content could be ascribed to the response to cold in exactly the same manner, rendering that state cognitive in the same way. We cannot conclude that bodily states to which content has been ascribed are emotions. Something else is missing.[12]

This point about Prinz clarifies one aspect of the two-stage model: the model, like Prinz's proposal, seeks a characterization of emotion that is both cognitive and appropriately physical/somatic/experiential. But the argument presented shows that the cognitive content must be explicit rather than implicit or ascribed. This is a necessary condition, and the two-stage model accounts for this component of emotion.

The line of argument just presented focuses on autonomic responses, but the same conclusion follows when responses are understood in a broader sense, as including an action-readiness component. Preparation for action is not an

[11]I thank an anonymous reviewer from JMB for helping me to clarify these two points.

[12]See Prinz (2004, p. 244): he defines emotions as having two components, embodied appraisals and valence markers, and my point is that the response to cold *has both*. Prinz might respond by appealing to core relational themes, suggesting that the response to danger "detects without describing" (p. 243) "immediate, concrete and overwhelming physical danger" (p. 16, taken from Lazarus), while the cold response detects no such theme (see Prinz, chapter one for references to Lazarus's work). But the response to cold does detect a relational theme, though one that does not appear on Lazarus's list, a core relational theme we could maybe articulate in terms of physical threats to homeostasis. The question is then, why do some core relational themes constitute emotions and others do not? Prinz might respond that the point is empirical, starting with emotional states we can make a list, and the core relational theme detected by the cold response is not on it. But this leaves the difference unexplained, and suggests that more needs to be said.

emotion or emotional. There is no basis for distinguishing preparation for jumping into a pool (taking a deep breath, tensing muscles against the expectation of cold, and so on) from the preparation for action in response to danger; there is no basis on which the second can be distinguished as emotional.

Action-readiness is not yet purposeful behavior; a third cognitive process — one distinct from the initial appraisal and from the interpretation — will generate behavior. This claim about there being a third cognitive process is an empirical hypothesis, but it is not an essential part of the two-stage model. Even if there is a behavioral component of responses, this component does not offer the basis for a criticism of the preceding line of argument because intentionally directed behavior is neither necessary nor sufficient for an emotion. It is not necessary because one can feel afraid without acting in any way. It is not sufficient because intentional behavior, even if facilitated or prepared for by a response, is itself not emotional. To see this consider an emotional zombie, one who acts in the same way as everyone else but without any of the underlying emotions. Such a person could be fully conscious and could generate the full range of complex, intentional behavior, all without experiencing emotions. Such a person might have the appropriate physiological (and behavioral) response to seeing a bear in the woods and flee, without feeling afraid. The plausibility of this case shows that intentional behavior is not sufficient for emotion. In short, responses serve as action potential or action-readiness, but no behavior need occur for there to be an emotion, and no resulting behavior is sufficient for an emotion.

Note that an appeal to zombies could seem to beg the question. But I do not mean to appeal to a hypothetical creature with exactly the same physiology and no emotion (so this is not a traditional zombie argument). Instead, we can easily imagine a person who does not experience fear in, say, a wartime situation, and we could describe this person as an emotional zombie because the perception of danger and subsequent physiological reactions occur without the emotion of fear. This zombie and another person in the same situation who does experience fear do not have to have the exact same physiology for my argument to work; indeed, interpretation and the conscious binding described below are both psychological processes, so their presence will cause physiological differences between the two persons.

This line of thought about intentionality relies in a fundamental way on first-person evidence, and on our (or at least my) conceptual commitments regarding emotion, and so for this reason it is difficult to defend through further argument; the line of thought does, however, explain the mistake in taking responses to be emotions.[13] Note that in labeling this second stage *inter-*

[13]The claim that emotions are the product of an interpretive process suggests that non-human animals will not experience emotion, because non-human animals are generally thought to not possess the required capacity for higher-order thought, and so will not be able to have the

footnote continued on next page ⟶

pretive I do not mean to be relying on or referring to a particular theory of interpretation. The process at work involves the assignment of meaning to a response or state of arousal in the context of a situation taken to have caused it, and this process is reasonably characterized as a form of interpretation.[14] Below I will further refine this characterization, and explain why this interpretive process is not conscious, though conscious reflection on it could affect the resulting emotion.

We interpret two different physiological responses to danger as fear — correctly — on the basis of an interpretation, on the inferred relationship between a set of physiological changes and events in the surrounding environment. But note also that there is space here for *mis*-interpretation: emotions arise from the interpretation or misinterpretation of a response in light of an actual or perceived cause. One could be mistaken about the connections between arousal and the environment, and in the binding make mistakes, mis-attributing our arousal and experiencing the "wrong" emotion. One could interpret a stomach ache and feel despair. That is, wrongness would lie in a misunderstanding of the relationship between responses and the surrounding environment, in a conflict between the content of the initial appraisal and the interpretation about the salience of events in the environment — though there is no internal standard of correctness here because the content of the initial appraisal is unavailable (on this point see below).

Despite the potential for misinterpretation, the results of the interpretive process are not arbitrary, so this potential for misinterpretation is not a reason to dismiss my account of emotion. Expecting otherwise, that is, demanding

necessary meaning-ascribing interpretations. For a number of emotions this should not be surprising: dogs, for example, cannot feel shame because they cannot understand the underlying social norms against which shame is possible. But this should not be taken to suggest that animals cannot have subjective experiences, form social bonds, and experience a kind of loss when those bonds are broken. I agree, then, with the main lines of Carruthers (2004), namely that animals can suffer because of the role of somatosensory feelings in their lives, even if those feelings are not conscious. But, as the foregoing should have made clear, I resist the association of those somatosensory states with emotion. That said, Carruthers' account of the experience of the purely negative (or presumably positive) aspect of bodily changes could be a significant component of human experience, and so it could help explain the too-quick slide from that experience to talk of emotion proper. To clarify: I take emotion to be responses understood and experienced in a certain way. Animals have some of the components of emotion, and so will have some analog of emotion but not emotions per se. Note also that some states we call emotions in humans might also fail to count as emotions on my view; for example, a hostile response might not properly be counted as anger.

[14]Shweder (1994) offers a related account. He argues that an emotion is a story about a physiological response, and the construction of this story is the "emotionalization" of the response. This view is, in my opinion, too purely cognitive; it fails to account for the psychological arousal or agitation that distinguishes emotion from other narrative cognitions. Note that when Shweder talks of emotionalizing feelings, he refers to the emotionalization of the feeling of a response — the feeling of being aroused in a certain way, and not the feeling of, say, being angry, which is produced through emotionalization of the response/arousal.

that there be no room for *mis*interpretation, amounts to expecting a system that perfectly reproduces the narrative structure of the world without personal perspectives, biases, and confusions. Expecting that is unreasonable.[15]

Mistakes of this sort provide an alternative explanation for cases in which some want to appeal to unconscious emotions. For example, an office worker could experience agitation because of conflict with his boss, agitation that would normally be experienced as anger at that boss, but because of a mis-attribution the underlying agitation is displaced; the agitation is expressed as anger at his spouse. This sort of case seems best described as one of mis-directed agitation rather than a case in which anger at the boss is unconscious. The two-stage model makes it possible to explain both stages of the process, the generation of the agitation and its subsequent mis-direction.

Leaving the matter at this point may seem unsatisfying, but that is only because of the limits of conceptual analysis: our concept of emotion suggests that emotions must be experienced, but nothing prevents emotions from being re-conceptualized and identified with non-conscious responses (except for the damage to common sense). The appeal to first-person experience in the distinction between the response to cold and fear is meant to shake one's confidence in such a re-conceptualization (if one is even tempted in that direction), but it can do no more. The same line of thought holds for the case of emotional zombies, mentioned above. And, the possibility of an empirical account of consciousness should help eliminate the impulse to set aside conscious experience in order to give an adequately scientific account.

3. *Response-generating processes are opaque.* According to the affect program literature, the cognitive processes responsible for generating responses (the initial appraisals) are informationally encapsulated modules that were selected for their adaptive value over the course of evolution. In addition to being encapsulated, these modules are inaccessible to other parts of the mind, meaning that the content of these appraisals will not be accessible to other modules, to an executive control module, to global cognition, or to phenomenal consciousness.[16] We can, then, perceive physiological responses in our bodies,

[15]Richard Moran (2001) criticizes Charles Taylor's constitutive thesis, mentioned in the introduction, along these lines. Misunderstandings create very complicated cases. On my view someone who bangs his fist on a table and shouts "I'm not angry" is correct. Anger is different from frustration, agitation, and hostility in that it requires an identification of a wrong; a hostile action in response to a wrong not yet articulated is not yet anger. To be sure, there is something counterintuitive here; we want to say that the person in this example is angry. But saying so is (on my view) a demand for reinterpretation, a demand for the subject to better understand her own reactions.

[16]See also Griffiths (1997) and Carruthers (2005). Carruthers uses the term *inaccessible* where Griffiths uses *opaque*, and he, Carruthers, argues that modules must be both encapsulated and inaccessible in order to keep mental processing tractable. In a more recent paper, however, Carruthers (2006) changes his mind about encapsulation, allowing that heuristics could keep non-encapsulated mental processes tractable.

but we lack direct access to the content of the appraisal that triggered the response.

As a result, interpretive modules or an executive control module must infer the content of these appraisals in order to construct a coherent narrative of events and judgments — a narrative that makes sense of a response. In situations we classify as emotionally-charged, responses trigger this interpretive process to account for and explain the (conscious or unconscious) experience of physiological changes. This explanation ascribes meaning to the response, and the experience of the response understood in a particular way constitutes the emotion. In addition to constituting emotions, this interpretive process makes the (inferred) content of the appraisal and the broader narrative surrounding it available for global cognition. And so, in addition to making it possible to experience a physiological response as being *about* some cause in the surrounding environment — meaning, in addition to making possible the conscious experience of an emotion — the interpretive process makes it possible to modulate our responses.

The fact that the initial appraisal process is opaque supports the argument above against Prinz's view, or it supports an argument for reading Prinz's view as a variant of mine. On my view we must ascribe content to responses (in the form of an interpretive explanation) because the content of the initial appraisal is opaque; this enables us to experience a response as meaningful. Even if it is not fully articulated, the interpretation creates an explicit content, explicit in the sense that it is available to the subject. Prinz, in contrast, takes initial appraisals and the physiological responses to lack cognitive content in the usual sense of cognitive, but ascribing content (by appeal to informational semantics) enables him to transform James's body-oriented account of emotion into one that accommodates aspects of cognitive views: responses can therefore be said to have content in virtue of causal connections with the surrounding environment. But this leaves open the question of whether the content is available to the subject: Does the subject experience the response as meaningful? The argument in this section of the main text concerning the cold response suggests that responses must be experienced as meaningful in order to be emotions (or emotional). But then Prinz's appeal to informational semantics is irrelevant: with Prinz we do not need to argue that the response has content intrinsically, or prior to an interpretation; even if it did, that content would not be available.

Timothy D. Wilson (2002) offers a general account of mind in these terms, isolating the inferential nature of our access to the set of non-conscious processes that make possible much of our perception, thought and action — "pervasive, adaptive, sophisticated mental processes that occur largely out of view" (p. 6). These processes include the perception of the world in three dimensions, and also the analysis of information about the social world, the

formation of judgments about others' personalities and intentions, and the explanation of our own behavior and that of others in terms of causes and reasons. To be sure, we can have conscious access to the outputs of these processes, but not to the processes themselves.

Wilson's more general account of the role played by inferential processes supports the contention of the two-stage model, namely that interpretations of non-conscious appraisal processes play a central role in the formation of emotion. Moreover, below I will extend the two-stage model to include appraisals not characterized as minimally-cognitive. If response-generating appraisals can incorporate conceptual knowledge and background knowledge, then responses may not be generated by a set of discrete and opaque modules. But Wilson's work shows that even if this is the case, even if responses are generated by fully cognitive processes not localized in discrete modules, there is good reason to expect that more complex response-generating appraisals will remain inaccessible to global cognition and interpretation will still be necessary.

Wilson's work therefore offers an argument for the two-stage model of emotion that does not depend on the line of argument above about the distinctness of responses or on the point about the kind of appraisal at work (cognitive, minimally-cognitive, or non-cognitive): given the conceptual point, that emotions are responses experienced as contentful, Wilson's account of mind shows that an interpretive process will be necessary to supply the content.[17]

Further Evidence for the Role of Physiological Responses

Two further lines of evidence support the account offered here of the relationship between physiological responses and emotions. First, Paul Ekman (1992a) observed that forming facial expressions produces both the physiological changes associated with emotions as well as emotional experience (see also the more detailed discussion of this finding in Levenson, Ekman, and Friesen, 1990). This phenomenon is exploited in experiments on emotional responses, one of which was discussed above; in that study (as in others) subjects were told to contract certain muscles — that is, they were not told to make the facial

[17]Separate from the more general characterization of self-directed higher-order thought processes as inferential, Wilson identifies affective reactions with emotions (2002, p. 112) — which I think is an error. Following James, Wilson argues that our emotions can generate adaptive responses prior to conscious experience, meaning that non-conscious emotions generate adaptive responses. But the discussion in the main text shows that Wilson's point should be put in terms of responses and not emotions. The intuition underlying Wilson's claim is preserved with this refinement: the adaptive unconscious produces responses that enable us to deal with problems posed by the environment, all without input from conscious processes. But on this refinement, the example of emotion no longer supports his claim that we can have unconscious feelings. See pp. 124–125.

expression associated with some emotion — and the resulting physiological responses were studied. Ekman offers a series of very general explanations for why forming facial expressions would trigger the other autonomic changes associated with an emotion. His own view is that there is a "hard-wired connection between the motor cortex and other areas of the brain involved in directing the physiological changes which occur during emotion" (1992, p. 35). This is all he says on the matter, and he concedes that there is no empirical evidence for this hypothesis or the alternatives he mentions. The relevant point here is this: in one study, 78% of the subjects reported feeling an emotion. This finding suggests that feedback from either facial expressions or facial expressions combined with certain patterns of physiological arousal can provide the basis for, and so be part of the sufficient condition for, emotional experience.

Second, patients with spinal cord injuries have some degree of impairment in their emotional experience, and the higher the injury on the spine the more that experience is impaired. In a recent study, Montoya and Schandry (1994) measured spinal patients' ability to perceive their own cardiac activity (by having them count their own heartbeats and comparing the counts to the measured heart rate).[18] Montoya and Schandry found that the ability to accurately perceive heart rate varied with the impairment of emotional experience, and they concluded from this finding that the experience of emotional feelings depends on feedback about changes in bodily states — which supports the contention here that physiological responses are a necessary component of emotional experience.[19]

---

[18]This evidence, along with the first work on the subject by Hohmann (1966), has been challenged; for an overview of the debate see Cohen (2002).

[19]For a detailed discussion of this line of research, see Damasio (1999), which describes an extreme version of this kind of impairment, one called locked-in syndrome (now subject of a recent film, Julian Schnabel's *The Diving Bell and the Butterfly*). As noted above, Damasio offers an explicitly Jamesian theory of emotion. Though I do not have space for a detailed overview of Damasio's work and the points of similarity and difference, two points are important to mention. First, the relationship between my account and his is partially obscured by terminological differences: Damasio uses the term "emotion" for what I call a pre-emotional response, and he uses "feeling" where I use emotion. Second, Damasio assimilates responses to the broader set of homeostatic processes developed over the course of evolution. But given the line of argument in the main text showing that responses are not emotional, I take it to be important to distinguish emotions from other experiences of homeostatic change that will not be intentional. Moreover, Damasio assumes that bodily responses differ across emotions, and so he adopts James's view of emotion as the experience of those responses. Despite my claim that ANS responses are not differentiated, Damasio's point may be defensible because his conception of responses is especially broad, including, for example, the low rates of image production and hyper-attentiveness to images in the case of sadness, as opposed to the rapid image change and short attention span that goes with happiness. If what he calls "bodily ways" include factors such as these, then responses could turn out to be differentiated, and the experience of these bodily states could produce distinct emotional experiences. But even if this is the case, the experience of such bodily ways will still be unable to produce the intentionality we associate with emotion, and for this reason an interpretation-oriented account of the second stage is required.

Extension of the Two-Stage Model to the (So-Called) Higher Cognitive or Non-Basic Emotions

Up to this point I have relied on the empirical evidence on affect programs to suggest a role for an interpretive process in generating emotions. The central experiments were conducted on animals, primarily rats, for which reflex-like and minimally-cognitive appraisals are at work. The generalization of the account to humans depends on parallels in brain structure across humans and rats and also on plausible assertions about evolutionary development preserving pre-existing structures.

But for humans there is good reason to think that the non-conscious appraisal processes could also involve both conceptual knowledge and background knowledge of facts and social norms. For example, a response to danger could occur when I encounter a gun; this response requires the classification of the object as a gun and background knowledge about the dangers surrounding guns — so this response requires conceptual knowledge at least in that minimal sense. Conditioning alone cannot account for the response, because it is possible even for those of us who have never had an actual bad experience with a gun.

Whether and to what degree responses in non-human animals can also be generated by appraisals involving conceptual distinctions is an empirical question. There are suggestive cases, and it seems plausible to expect conceptual distinctions to be at work in the appraisals of some primates. Further, we should expect that structures would gradually evolve that can trigger physiological responses on the basis of cognitively richer appraisals. And these structures would evolve on top of the capacity for minimally-cognitive appraisals described by LeDoux and Griffiths (and others). My appeal to the account of appraisals in the affect program literature — an account of appraisals as reflex-like — was not meant to suggest that such appraisals, which generate the bodily responses and their subsequent sensation, cannot involve conceptual content and background knowledge, at least in humans.

A parallel case: driving a car requires a number of reflex-like reactions. These are not learned though conditioning (fortunately), and they can involve conceptual knowledge at least in the form of the categorization of objects — e.g., I won't swerve out of my lane to avoid a plastic bag in the road but I will for an animal. But these reactions are nonetheless plausibly described as both non-conscious and reflex-like. My suggestion is that pre-emotional responses can be generated by similar appraisals that are also non-conscious and reflex-like but nevertheless involve conceptual distinctions and background knowledge.

Allowing that response-generating appraisals can have a richer cognitive component in humans marks a break from the literature on affect programs, and

it amounts to an empirical hypothesis, one I take to be plausible for the reasons just described. But the central elements of the account offered thus far — the distinction between the response and the subsequent emotion, and the claim that responses themselves are not intentional — are preserved while admitting a role for conceptual knowledge and background knowledge in appraisals.

The literature on affect programs usually makes a distinction between affect programs and what are called "higher cognitive" or "non-basic" emotions, which seem to lack stereotypical responses and seem to require relatively complex cognitive appraisals — appraisals that would not be possible if neural programs of the sort described in the affect program literature were at work. These emotions include shame, jealousy, and pride, as well as some instances of anger and fear (those instances of fear not easily detected using perceptual clues).

But if conceptual distinctions and background knowledge can play a role in non-conscious, reflex-like appraisals, then the distinction between affect program and higher cognitive emotions may disappear. If, for example, I react to a drop in stock market indices with an immediate physiological response, then the affect program structures are presumably at work, relying on background knowledge of the effects of changes in those indices.

There is another way in which the class of affect program emotions could be expanded to include emotions usually classified as higher or cognitive. To begin with an example: if humans recognize and respond to instances of social conflict in an immediate way with physiological arousal, we could then interpret that response and experience a variety of more particular emotions, like shame, or guilt, or some other emotion for which the appraisal seems more complex. It is possible that there is a limited set of basic appraisals — maybe as few as three, danger/surprise, social conflict, and positive social bonding — and the responses produced by each would then be interpreted to produce the complete range of human emotions, including the social emotions, anger, shame, guilt, envy, jealousy, as well as some instances of happiness and sadness.

If this proposal is correct for some set of emotions, and this, too, is an empirical question, then the difference between basic and non-basic emotions again disappears: all emotions, or at least the affect program ones and some of the so-called higher ones, are generated by the interpretation of reflex-like responses. The qualification, that at least some of the non-basic emotions are generated in this way, is meant to acknowledge the possibility of there being another class of emotion with no physiological component. If there are such emotions, they would stand outside of the debate about how emotions are related to autonomic or physiological responses. But such emotions would not present the basis for an argument against my proposal.

At root this second suggestion is a point of speculative ethology, namely that some small set of appraisals/responses can explain the capacities mammals

have to respond to both threatening situations and social relationships appropriately. This proposal accommodates the empirical research, according to which emotional responses are not distinct, and it allows room for an evolutionary account. There could be a set of discrete structures for recognizing the two kinds of situations, or there could be a more general response system that produces physiological arousal in stressful or conflictual situations. But, again, the neural realization of such a system is not an issue here.

This line of thought and the preceding point, about the place for conceptual knowledge and background knowledge in response-generating appraisals, are not mutually exclusive. My goal in outlining them is only to show how an account beginning with reflex-like responses could generalize to be an account of emotion more broadly.

Schachter and Singer's Hypotheses and Results

As noted, Schachter and Singer (1962) suggested that emotions are a "function" of both physiological arousal and cognitive factors — these are the components of their "two-factor theory." More specifically, they argued (i) that unexplained physiological arousal generates an "evaluative need," and (ii) in response to that need, the arousal is "labeled" and "shaped by" an interpretation in the particular context in which it occurred. This labeling or shaping of an aroused state results in an emotion.

Schachter and Singer's experiment proceeded as follows: subjects were recruited to participate in a study of the effects of vitamins on perception. They were injected with either adrenaline or a placebo (the supposed vitamins), and were told to wait twenty minutes for the vitamins to take effect. The subjects were placed in a room with a researcher posing as another subject, and during this period Schachter and Singer tried to manipulate their emotional states in two ways. In one set of cases subjects were given a survey containing upsetting, personal questions to answer (for example, "With how many men (other than your father) has your mother had extramarital affairs?"); the confederate reacted angrily, encouraging the subject to do so as well. In the other set of cases the confederate engaged in increasingly silly behavior (making paper airplanes, playing with a hoola-hoop) in order to bring about a mood of euphoria. Some of the subjects were told (correctly) that the injection could produce an increased heart rate and shaking hands (the informed group), others that the injection could produce side effects like itching feet (the misinformed group), and others were not told anything (the uninformed group).

In the experimental setting for euphoria, subjects who were not informed or who were misinformed about the effects of the injection showed and reported stronger emotional reactions than the subjects who were aware of the connec-

tion between their arousal and the injection, and the same held for observations of the subjects in the anger setting.[20]

The fact that the same physiological arousal was labeled differently in different contexts shows that subjects labeled their arousal on the basis of an interpretation of the context, which varied with the two experimental settings. In Schachter and Singer's words, "emotional states are a function of the interaction of such cognitive factors [analyses of the situation] with a state of physiological arousal" (1962, p. 381).

The subjects who understood that their physiological arousal was produced by the injection did not report experiencing emotions, and this part of the result supports Schachter and Singer's second proposition, that the interpretation or labeling is triggered by an evaluative need: these subjects were informed about the effects of the injection; they therefore understood the cause of their feeling and so did not interpret/label their arousal; and as a result they did not experience an emotion. This suggests that the labeling is productive, that the physiological arousal is converted into an emotion by the interpretation/labeling. If this is the case, then it seems clear that different labelings can produce different emotions.

In part, Schachter and Singer's proposal serves as a direct test of the second stage of my two-stage model: by isolating and exposing the role played by interpretation in an experimental setting, Schachter and Singer offer some empirical support for thinking of emotion as dependent on and partly constituted by the meaning ascribed to a response in an interpretive process. As noted above, the interpretive process at work here is one that could be best thought of as explanatory, as directed at explaining (and therefore understanding) how the response is connected with and caused by certain aspects of our surroundings.[21]

But Schachter and Singer only conclude that subjects will interpret and label their responses when they are un-explained. They suggest that interpretations may play a role more generally, but they offer no account of how or why subjects could experience unexplained responses outside of the experimental setting.

---

[20]Self-report data for the anger group was not useful because it was discovered that the subjects, who were college students, volunteered for the experiment in order to receive extra points on their final exams, and, as a result, they were apparently unwilling to endanger those points by reacting angrily or voicing their anger to the experimenter. Nevertheless, as noted, the observations of the subject's behavior confirmed the result for the anger group.

[21]Schachter and Singer describe this process of labeling as cognitive, and they describe the resulting cognitions as explanatory (1962, p. 381). In a later discussion of this experiment Schachter (1966) offers more explicit support for describing the interpretive process in terms of explanation: he writes, "it is suggested that one labels, interprets, and identifies this stirred-up state in terms of the characteristics of the precipitating situation and one's apperceptive mass" (p. 50), and he describes his experiment as "manipulating an appropriate *explanation* [for the subject's arousal]" (p. 54). For a more general overview of the debate on how to interpret Schachter and Singer's use of the term "label," see Reisenzein (1983).

Schachter and Singer do take arousal to precede emotion. But because they offer no account of the appraisal process and its role in producing physiological responses, they cannot explain how or why a subject would *normally* be confronted by physiological arousal standing in need of explanation, and so they cannot explain why an interpretive process would be necessary in our everyday experience of emotion, outside of the laboratory setting. This limitation is implicit in their talk of arousal, as disconnected from external events, as opposed to responses to events. (I use the two terms interchangeably.) And this limitation is a product of Schachter and Singer's idiosyncratic version of cognitivism with respect to emotions: the cognitive factor they isolated in their experiment is the labeling/interpretation, *not* the appraisal responsible for the arousal in the first place. This makes their work quite different from cognitivism in the philosophical context, which is focused on the environment-directed appraisal process that generates the response.

In short, Schachter and Singer show how interpretation can play a role in producing emotions, and even that interpretations are necessary in the case of unexplained arousal. But they do not show how or whether interpretation should be necessary to produce emotions in general. By situating the interpretive process in a general theory, the two-stage model shows why interpretation is a necessary component of emotion in general. Moreover, their account leaves open the question of how an interpretation — a cognitive process about a state of arousal — can generate an emotion. I turn to this question and the broader issue of consciousness at the end of this paper.

Philosophical Criticisms of Schachter and Singer

Schachter and Singer presented their account as an alternative to the prevailing non-cognitive views of emotion, and their particular account of the role of cognition in generating emotion produced a great deal of controversy. In an essay published twenty years after Schachter and Singer's original study, Rainer Reisenzein (1983) reviewed more than 100 articles about it, many of which try to replicate Schachter and Singer's findings using different experimental frameworks.

For example, in one study by White, Fishbien, and Rutstein (1981), male subjects ran in place for two minutes (the high arousal group) or 15 seconds (the low arousal group), and were then shown a videotape of a woman being interviewed. Half of the subject viewed a tape in which the woman was attractive and enthusiastic, the others a tape featuring an unattractive and dull woman.[22] The subjects rated the woman in the video on nine traits, four of

[22]Zillman (1978) describes a series of similar experiments on the transference of arousal to situations involving aggression.

which were used to measure romantic attractiveness: how physically attractive the woman was, how sexy she was, how much the subjects would like to date her, and how much they would like to kiss her. High-arousal subjects rated the attractive woman as being more attractive than the low-arousal ones. White and his collaborators take this to show that the "misattribution of arousal facilitates romantic attraction" (1981, p. 56) — that is, they take this result to show that high arousal will produce "passionate love" if the arousal is interpreted as caused by the qualities of the attractive other. This account of passionate love is, of course, the application of Schachter and Singer's model to that emotion.

Despite results of this sort, there are a number of criticisms directed at Schachter and Singer and a great deal of controversy surrounding their work. In the remainder of this section I respond to one of their more philosophically oriented critics, whose work forces a clarification with respect to the interpretive process. But separate from the detailed points on both sides of the debate, I take Schachter and Singer's work to be suggestive; and for this reason I situated my view with respect to theirs. In terms of defending the two-stage model, however, I take the arguments presented above to be complete; they do not depend on an appeal to Schachter and Singer's work.

Paul Griffiths argues that, "with hindsight there is a single, devastating objection against [Schachter and Singer's conclusions]" (1997, p. 82). Appealing to the experimental literature on confabulation, he argues,

> One would expect Schachter and Singer's subjects to confabulate in order to explain the abnormal arousal caused by adrenaline injections. The results obtained do not discriminate between this null hypothesis and the hypothesis that subjects were observing the normal arousal associated with the emotions they reported. (1997, p. 83)

Griffith's null hypothesis is that the subjects confabulated explanations for their arousal, meaning that they didn't really experience an emotion, and they only *described* their arousal and their behavior in emotional terms because they were uninformed about the effects of the supposed-vitamin injection. This suggestion fails, however, because it cannot account for the differences Schachter and Singer observed in the *behavior* of the informed and uninformed subjects. This difference suggests that an emotion was present in the uninformed subjects — it is, for example, because of this emotion that the euphoria subjects behaved as they did.

That said, Griffiths is right in a sense, because confabulation is at work. But Schachter and Singer are trying to show that in the realm of emotion, confabulation is productive: interpretations are not *just* an inert explanation or a description of arousal in emotional terms, but instead result in the production of an emotion. Contrary to Griffiths' suggestion, then, the null hypothesis ruled out by Schachter and Singer's experiment is that the context and inter-

pretations of it play no role in the labeling and experiencing of an emotion, and — as noted above — the lack of emotion reported by the informed group compared to the uninformed groups seems to rule this hypothesis out.

The other explanation Griffiths offers for Schachter and Singer's result seems to be this: the subjects in the anger cases just experienced anger, and the ones in the euphoria cases experienced euphoria — so there is no need to talk of labeling the arousal produced by the adrenaline. This explanation is inadequate, however, because it does not explain the different levels of emotion experienced by the informed and uninformed subjects. That is, claiming that the euphoria subjects just experienced euphoria does not explain why the informed and uninformed subjects experienced and demonstrated different levels of emotion.[23]

A more sophisticated criticism along these lines would be the following: the experimental situations were only weak elicitors of emotions, so the subjects who received adrenaline experienced emotions only because the adrenaline had heightened their sensitivity to the normally weak emotion experienced in each situation. In other words, the anger subjects got angry only because they were already aroused and the euphoria subjects experienced euphoria only because they were already aroused. This would explain why subjects who received injections experienced emotions while those who received placebos did not (or, more precisely, why they experienced stronger emotions than the placebo group), but this alternative explanation still cannot explain why there were differences across the informed and uninformed groups.

Emotion and Consciousness

According to the two-stage model an emotion is a response understood and experienced in a certain way, a response to which meaning has been ascribed. On this account an appraisal triggers a physiological response because of some stimulus in the environment. The (conscious or unconscious) sensation of the arousal then triggers the interpretive process to connect or relate the response to some environmental cause, directing the response at some object and informing it with meaning.

The argument for the two-stage model is part conceptual and part empirical. The claim that emotions are responses experienced as meaningful is a

[23]In a later essay, Griffiths (1998) repeats this criticism but changes the second sentence to read: "The results obtained do not discriminate between this hypothesis and the hypothesis that the experiment simulated normal emotions" (p. 199). I do not know what to make of his use of the word "simulated" because it suggests that the subjects did not actually experience the emotion, contrary to the suggestion in his book, as quoted in the main text. Griffiths might mean that the experiment simulated conditions in which anger and euphoria are experienced and in virtue of that simulation subjects experienced real emotions. Or the text should have read stimulated, not simulated.

conceptual one (supported by the argument about the cold response above). The role played by meaning here shows that a number of competing accounts are incomplete. And the further claim, that this meaning is supplied by an interpretive process that is distinct from the initial appraisal, is supported by two lines of empirical research: work on affect program emotions and also Wilson's more general account of the mind. Wilson's account in particular offers support for the somewhat counter-intuitive claim that emotion-generating interpretations must infer the content of our initial response-generating appraisals.

But another point remains to be clarified: What is the difference between an interpretation that produces an emotion and one that merely accompanies and explains an instance of arousal? This is an application of a problem raised by Stocker (1987), who notices the gap between a judgment and an emotion that has the same appraisal-content, e.g., the gap between the judgment "that patch of ice is dangerous" and being afraid of falling on the ice. Judgment-oriented theories of emotion appeal to desires to account for the gap here: an emotion is an evaluation tied to a situation about which we have some strong desire. So, we experience fear given the judgment (or evaluation or appraisal) that the ice is dangerous combined with the desire to stay safe.

But Stocker cites other examples where the appeal to desires will not account for the difference between judgments and emotions: he can alternate between fear and relative comfort while flying, all without changing his evaluation of the danger of being in an airplane or his desire to survive the flight. The conceptual analysis fails to provide an explanation of the difference here, and it leaves Stocker only able to name the problem, characterizing emotions as "emotionally held thoughts."

The two-stage model of emotion can account for the gap Stocker identifies by explaining what it means to hold a thought emotionally. The evaluation of some situation will produce a judgment about it. In the context of a physiological response generated by a prior appraisal, this judgment will amount to an explanation of that response, and so a subject will experience an emotion. Outside that context, without the prior response/arousal, a judgment will be non-emotional. Put in Wilson's terms, where the adaptive unconscious has appraised the ice as dangerous and triggered a bodily response, the subsequent judgment that the ice is dangerous will, in Stocker's terminology, be emotionally held, and so the subject will experience an emotion.[24]

[24]Ellsworth (1994) makes this point in almost identical terms in her discussion of James, cited above: "the sense of bodily changes provides the emotionality to what would otherwise be a neutral perception or interpretation of the situation" (p. 223). But because she does not distinguish between the two cognitive processes identified in my two-stage model, she is unable to address the fundamental question identified here of how responses and emotions are related.

This line of thought could be open to a further criticism, an extension of Stocker's original point: How are we to understand the difference between the simultaneous presence of a judgment and arousal on one hand and an emotional experience of the judgment on the other? This extension applies directly to Schachter and Singer's account and to the two-stage model: What is the difference between an interpretation or explanation of arousal and an emotion?

Although I do not have space to develop the point systematically, the response lies in the binding of the two components. And, on my view, this binding is experienced in consciousness. According to Carruthers' (2000a, 2000b) account, the capacity for higher-order thought offers part of a basis for phenomenal consciousness. Given first-order perceptual content (i.e., red), representations of that content are generated by the higher-order thought systems (i.e., seems red to me). Carruthers then appeals to work suggesting that the representational content of a state depends at least in part on the nature of the system that interprets the state.[25] For example, data about patterns of light do not contain representational content for a being without the neural structures capable of interpreting and using it. Applied to the case of consciousness, the human ability to have first-order perceptual representations at the same time as — *bound together with* — the higher-order ones produces the subjectivity we identify with phenomenal consciousness.

The argument is one of empirical possibility: a system with the capacity to consume representations in the right way — meaning, with the capacity to bind and experience multiple representations together — *could* explain phenomenal consciousness, and so it is possible to give a naturalistic explanation. In short, the capacity to read bound representations in that way is the capacity for conscious experience.

A similar capacity for binding the affect associated with a response and a judgment about the environment could explain how an emotion differs from the co-presence of arousal/affect and a judgment. Arousal must be bound to — experienced together with, as connected to — a judgment about some feature in the environment, and without this binding the judgment is a cold, non-emotional thought. Or, put the other way, an interpretation or judgment about the environment and its relationship to arousal must be bound to the arousal that has already occurred, enabling us to experience the arousal as having the meaning of the interpretation. Otherwise, the arousal is just that, a non-emotional physiological response. If this account is correct, and ultimately this is an empirical question, then emotions are compound states composed of two elements, and compound states of this sort will be experienced differently from the two components merely coinciding.

---

[25]On this point Carruthers cites Millikan (1984) and Botteril and Carruthers (1999, chap. 7).

Above I advanced the conceptual point that emotions must be consciously experienced. Though much remains to be said, the analogy between the two binding processes described here could be taken to suggest that the binding of the two components of emotion is experienced in or as consciousness.

References

Barr–Zisowitz, C. (2000). 'Sadness' — Is there such a thing? In M. Lewis and J.M. Haviland–Jones (Eds.), *The handbook of emotions* (pp. 607–622). New York: Guilford.

Botteril, G., and Carruthers, P. (1999). *The philosophy of psychology*. Cambridge: Cambridge University Press.

Cacioppo, J.T., Berntson, G.G., Larson, J.T., Poehlmann, K.M., and Ito, T.M. (2000). The psychophysiology of emotion. In M. Lewis and J.M. Haviland–Jones (Eds.), *The handbook of emotions* (pp. 173–191). New York: Guilford.

Carruthers, P. (2000a). The evolution of consciousness. In P. Carruthers and A. Chamberlain (Eds.), *Evolution and the human mind* (pp. 254–275). Cambridge: Cambridge University Press.

Carruthers, P. (2000b). *Phenomenal consciousness: A naturalistic theory*. Cambridge: Cambridge University Press.

Carruthers, P. (2004). Suffering without subjectivity. *Philosophical Studies, 120*, 99–125.

Carruthers, P. (2005). The case for massively modular models of mind. In R. Stainton (Ed.), *Contemporary debates in cognitive science*. Oxford: Blackwell. Retrieved May 8, 2005, from http://www.philosophy.umd.edu/Faculty/pcarruthers/Articles-a.htm

Carruthers, P. (2006). Simple heuristics meet massive modularity. In P. Carruthers, S. Laurence, and S. Stich (Eds.), *The innate mind: Culture and cognition*. Oxford: Oxford University Press. Retrieved May 8, 2005, from http://www.philosophy.umd.edu/Faculty/pcarruthers/Articles-a.htm

Cohen, M.A. (2002). Self-interpretation and emotion (Doctoral dissertation, University of Pennsylvania). *Dissertation Abstracts International, 63*, 1858.

Cohen, M.A. (2005). Against basic emotions. *The Journal of Mind and Behavior, 26*, 229–254.

Damasio, A. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. New York: Harcourt.

Damasio, A. (2003). *Looking for Spinoza: Joy, sorrow, and the feeling brain*. New York: Harcourt.

Davidson, R.J. (1993). Parsing affective space: Perspectives from neuropsychology and psychophysiology. *Neuropsychology, 7*, 464–475.

Deigh, J. (1994). Cognitivism in the theory of emotion. *Ethics, 104*, 824–854.

Ekman, P. (1980). Biological and cultural contributions to body and facial movement in the expression of emotion. In A.O. Rorty (Ed.), *Explaining emotions* (pp. 73–102). Berkeley: University of California Press.

Ekman, P. (1992a). Facial expression of emotion: New findings, new questions. *Psychological Science, 3*, 34–38.

Ekman, P. (1992b). An argument for basic emotions. *Cognition and Emotion, 6*, 169–200.

Ekman, P., Friesen, W.V., and Ellsworth, P. (1982). What are the similarities and differences in facial behavior across cultures? In P. Ekman (Ed.), *Emotion in the human face* (pp. 128–144). New York: Pergamon Press.

Ekman, P., Levenson, E.R., and Friesen, W.V. (1983). Autonomic nervous system activity distinguishes between emotions. *Science, 221*, 1208–1210.

Ellis, R., and Newton, N. (2002). The unity of consciousness: An enactivist approach. *The Journal of Mind and Behavior, 26*, 255–280.

Ellsworth, P.C. (1994). William James and emotion: Is a century of fame worth a century of misunderstanding? *Psychological Review, 101*, 222–229.

Epstein, W., and Hatfield, G. (1994). Gestalt psychology and the philosophy of mind. *Philosophical Psychology, 7*, 163–181.

Griffiths, P. (1997). *What emotions really are: The problem of psychological categories*. Chicago: University of Chicago Press.

Griffiths, P. (1998). Emotions. In W. Bechtel, D.A. Balota, and G. Graham (Eds.), *Companion to cognitive science* (pp. 197–203). Oxford: Blackwell.
Haidt, J., and Keltner, D. (1999). Culture and facial expression: Open-ended methods find more expressions and a gradient of response. *Cognition and Emotion, 13*, 225–266.
Hatfield, G. (2007). Did Descartes have a Jamesian theory of emotions? *Philosophical Psychology, 20*, 413–440.
Hohmann, G.W. (1966). Some effects of spinal cord lesions on experienced emotional feelings. *Psychophysiology, 3*, 143–156.
James, W. (1981). *Principles of psychology*. Cambridge, Massachusetts: Harvard University Press. (Originally published 1890)
James, W. (1984). *Psychology: Briefer course*. Cambridge, Massachusetts: Harvard University Press. (Originally published 1892)
James, W. (1994). The physical basis of emotion. *Psychological Review, 101*, 205–210. (Originally published in *Psychological Review, 1*, 516–529, in 1894)
LeDoux, J. (1996). *The emotional brain: The mysterious underpinnings of emotional life*. New York: Simon and Schuster.
Levenson, R. W., Ekman, P., and Friesen, W.V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology, 27*, 363–384.
McGinn, C. (2003, February 23). Fear factor. *New York Times Book Review*, p. 11.
Millikan, R. (1984). *Language, thought and other biological categories*. Cambridge, Massachusetts: MIT Press.
Montoya, P., and Schandry, R. (1994). Emotional experience and heartbeat perception in patients with spinal cord injury and control subjects. *Journal of Psychophysiology, 8*, 289–296.
Moran, R. (2001). *Authority and estrangement: An essay on self-knowledge*. Princeton: Princeton University Press.
Newton, N. (2000). Conscious emotion in a dynamic system: How I can know how I feel. In R. Ellis and N. Newton (Eds.), *The caldron of consciousness: Motivation, affect, and self-organization* (pp. 91–108). Amsterdam: John Benjamins.
Prinz, J. (2003). Emotion, psychosemantics, and embodied appraisals. In A. Hatzimoysis (Ed.), *Philosophy and the emotions* (pp. 69–86). Cambridge: Cambridge University Press.
Prinz, J. (2004). *Gut reactions: A perceptual theory of emotion*. Oxford: Oxford University Press.
Reisenzein, R. (1983). The Schachter theory of emotion: Two decades later. *Psychological Bulletin, 94*, 239–264.
Russell, J.A. (1994). Is there universal recognition of emotion from facial expression: A review of the cross-cultural studies. *Psychological Bulletin, 115*, 102–141.
Schachter, S. (1966). The interaction of cognitive and physiological determinants of emotional state. In L. Berkowitz, (Ed.), *Advances in Experimental Social Psychology, 1*, 49–80.
Schachter, S., and Singer, J.E. (1962). Cognitive, social and physiological determinants of emotional state. *Psychological Review, 69*, 379–399.
Shweder, R. (1994). "You're not sick, you're just in love": Emotion as an interpretive system. In P. Ekman and R.J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 32–44). Oxford: Oxford University Press.
Stocker, M. (1987). Emotional thoughts. *American Philosophical Quarterly, 24*, 59–69.
Taylor, C. (1985a). Self-interpreting animals. In C. Taylor, *Human agency and language: Philosophical papers volume one* (pp. 45–76). Cambridge: Cambridge University Press.
Taylor, C. (1985b). What is human agency? In C. Taylor, *Human agency and language: Philosophical papers volume one* (pp. 15–44). Cambridge: Cambridge University Press.
Taylor, C. (1985c). The concept of a person. In C. Taylor, *Human agency and language: Philosophical papers volume one* (pp. 97–114). Cambridge: Cambridge University Press.
Toobey, J., and Cosmides, L. (1990). The past explains the present: Emotional adaptation and the structure of ancestral environments. *Ethology and Sociobiology, 11*, 375–424.
White, G., Fishbien, S., and Rutstein, J. (1981). Passionate love and the misattribution of arousal. *Journal of Personality and Social Psychology, 41*, 56–62.
Wilson, T.D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious*. Cambridge, Massachusetts: Harvard University Press.

Zajonc, R. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35*, 151–175.

Zajonc, R. (1984). On the primacy of affect. *American Psychologist, 39*, 117–123.

Zillman, D. (1978). Attribution and misattribution of excitatory reactions. In J.H. Harvey, W. Ickes, and R.F. Kidd, (Eds.), *New directions in attribution research*, volume 2 (pp. 335–368). Mahwah, New Jersey: Lawrence Erlbaum.