What Emotions Really Are (In the Theory of Constructed Emotions)

Jeremy Pober*†

Recently, Lisa Feldman Barrett and colleagues have introduced the Theory of Constructed Emotions (TCE), in which emotions are constituted by a process of categorizing the self as being in an emotional state. The view, however, has several counterintuitive implications: for instance, a person can have multiple distinct emotions at once. Further, the TCE concludes that emotions are constitutively social phenomena. In this article, I explicate the TCE*, which, while substantially similar to the TCE, makes several distinct claims aimed at avoiding the counterintuitive implications plaguing the TCE. Further, because of the changes that comprise the TCE*, emotions are not constitutively social phenomena.

1. Introduction. Griffiths (1997) significantly advanced the philosophical study of emotions. Specifically, he argued that emotions had to be understood in terms of a causal theory of meaning (Putnam 1975; Kripke 1980) such that the term 'emotions' as well as subclasses of emotions such as 'anger' and 'fear' refer to whatever it is that causes so-called emotional behavior. Consequently, philosophical inquiry into the emotions is inextricably linked to neuroscientific inquiry into the brain basis of emotions.

In turning to neuroscientific work on emotions, Griffiths looked to a set of theories that can be grouped together as the affect program hypothesis (e.g., LeDoux 1996; Izard 1997; for philosophical analysis, see Prinz 2004;

Received February 2017; revised February 2018.

*To contact the author, please write to: Department of Philosophy, University of California, Riverside; e-mail: jeremy.pober@gmail.com.

†This article has benefited from the input of an unusually large number of people. I thank Lisa Feldman Barrett for piquing my interest in her work, as well as Matt Barker, Agnieszka Jaworska, Adam Kovach, Christopher Masciari, and Jim Sparrell for helpful discussions and comments. I especially thank Alex Madva, Carlos Montemayor, and Eric Schwitzgebel for comments on multiple earlier drafts of this article, as well as two anonymous reviewers for *Philosophy of Science*.

Philosophy of Science, 85 (October 2018) pp. 640–659. 0031-8248/2018/8504-0005$10.00

Kovach and DeLancey 2005; Scarantino 2012). According to this hypothesis, the so-called basic emotions (Ekman 1984) are realized in dedicated neurocircuits that motivate behavior without deliberate intervention. For example, the 'fear circuit' is centered around the amygdala and produces fight-or-flight responses in response to stimuli that are perceived as threatening (LeDoux 1996). Token activations of this circuit are the brain basis for token instances of fear.

The primary limiting factor to Griffiths's seminal work is that he turns to an outdated empirical theory of emotions to investigate their brain basis. Another theory of emotions, the Theory of Constructed Emotion (TCE; formerly the Conceptual Act Theory), has been put forth by Lisa Feldman Barrett and colleagues (Barrett 2006, 2009a, 2012, 2014, 2017; Wilson-Mendenhall et al. 2011; Lindquist et al. 2012; Barrett, Wilson-Mendenhall, and Barsalou 2015) and is gaining traction. While there are still defenders of the affect program hypothesis in both psychology (e.g., Izard 2007) and philosophy (e.g., Scarantino 2009, 2012), the TCE has been gaining in popularity to the point where one of the original major proponents of the affect program hypothesis has recanted (LeDoux 2012, 2015).

According to the TCE, emotions are what Searle (1995) calls social objects (Barrett 2012, 2014, 2017). Social objects are a class of objects that exist only because members of a community conceptualize them as such (Searle 1995). If Barrett is right, and emotions are social objects, then emotions are what I will call 'constitutively external' phenomena. In 'constitutive externalism', an entity's constitutive basis extends beyond its physical basis, in the way that what makes me an uncle extends beyond my body and to my brother and his son. As I will demonstrate, emotions' being constitutively external is an impediment to Griffiths's research program.

But before discussing how emotions' being social objects is problematic for the research program, there are more immediate issues. For the issues that motivate Barrett to identify emotions with social objects create conceptual problems that themselves reduce the TCE's power to explain emotional phenomena. In particular, the TCE entails that (1) a single organism can have multiple simultaneous distinct emotions for different perceivers, (2) we cannot be wrong about our own emotions, and (3) infants and animals cannot themselves have emotions but can have emotions created in them by a linguistically competent person.

In this article, I aim to provide a theory of emotions that agrees with the TCE on most of its major claims but addresses these concerns, which I will call the TCE*. I then contend that according to the TCE*, as opposed to the TCE, emotions are constitutively internal states, that is, states whose constitutive basis is entirely in the brain of the emoter. To be clear, though, my aim in this article is not to adjudicate the empirical case for the TCE or TCE* put forth by Barrett and others (e.g., Russell 2003); that task, while in-

teresting, has been done elsewhere (e.g., Barrett 2006; Scarantino 2009, 2012; LeDoux 2012; Lindquist et al. 2012). Rather, my aim is to clarify what emotions are if something largely like the TCE is correct.

The article proceeds as follows. In section 2, I discuss the TCE in detail. In section 3, I discuss each of the conceptual issues mentioned above and argue that each one depends on a specific motivation of Barrett's for identifying emotions with social objects; I then discuss what changes are needed to both assuage Barrett's motivations as well as solve the conceptual issues I introduce. In section 4, I turn to the question of whether emotions are constitutively external phenomena. I argue that Griffiths's project would be compromised if they are. However, I also demonstrate that, if the TCE* is correct, emotions are constitutively internal phenomena such that a token emotion is constituted by a token process of categorization wherein the emoter categorizes herself as experiencing an emotional state of a certain type.

2. The Theory of Constructed Emotion. In the affect program hypothesis, each (basic) emotion had a distinct neurocircuit dedicated to the realization of token emotional episodes of the appropriate kind. The TCE, however, starts from the assumption that there are no 'subpersonal' systems dedicated to any emotion category or even emotion itself: the subsystems which realize emotion also realize, for example, 'cold' cognitive processes. Rather, according to Barrett (2006, 2014) and Lindquist et al. (2012), emotions are categorizations of the self, based on multiple subpersonal systems such as perception, working memory, and affective state. At the heart of the TCE is the claim that a process of categorization, which renders an output with contents like "I am afraid," is constitutive of a token fear state.

2.1. Categorization in the TCE. According to Barrett, one of the main functions of the brain is to categorize its 'internal' and 'external' milieu (its body and environment). She defines categorization as "comprising two processes: (1) accessing and activating a relevant category representation and binding it to a perceived instance, and (2) drawing inferences from knowledge associated with the category and applying them to the instance" (Barrett et al. 2015, 89). The brain performs these categorizations based on input from various domains, including perception and memory. And the brain can use multiple domains to categorize a specific bit of input: for instance, as I will explain shortly, to categorize perceptual input into objects of perception, the brain uses information from memory as well as perception itself. Barrett (2009a) calls the perceptual input the 'focus' of the categorization; it also renders a categorization about this input, that is, objects in the perceptual field. The focus is something like the type of input that triggered a token act of categorization.

Barrett illustrates the categorization process she considers the basis of emotions with a description of the phenomenon philosophers term *seeing as*. When seeing, for instance, a bee, for the visual system to go from the array of photons on the retina, or even the basic shape-like representations it creates (Marr 1982), to seeing something *as a bee*, which requires "categoriz[ing] the sensory input using conceptual knowledge from past experiences" (Barrett et al. 2015, 88). Specifically, it requires an organism to have information available about the concept, BEE, and apply it automatically and effortlessly to its perceptual representation.

According to Barrett, the brain categorizes for an evolutionary reason: it does so with an eye to promoting behaviors that help an organism achieve *allostasis*, which Barrett defines as the process of "efficiently ensur[ing] resources for physiological systems within an animal's body . . . so that an animal can grow, survive, and reproduce" (2017, 3). Specifically, categorizing an object in one's perceptual field promotes what Barrett (2012) calls *situated* action, that is, behavior tailored to promote allostasis in the current environment of the organism. She uses her example of a bee to explain how categorizations support allostasis by suggesting behaviors: "For . . . people who have been stung . . . seeing a bee might mean freezing. . . . Or they might wave their arms and run away" (Barrett et al. 2015, 89). The reason seeing a bee prepares us for action is because the information stored in an organism's memory about bees is not just perceptual, it also includes information about stings, which are bad for allostatic success.

Barrett does not commit herself to a specific computational account of how the categorization process works, but she at several times (e.g., Wilson-Mendenhall et al. 2011; Barrett et al. 2015) notes the compatibility of her account with the theory of situated conceptualization promoted by Barsalou (1999, 2009). According to Barsalou's view, the way the brain gets from an act of categorization to a behavior is by searching memory for behaviors that helped achieve the goals of the organism during past experiences when that category was active (e.g., Barsalou 2009). The visual representation of a bicycle might promote riding behaviors for those who enjoy riding bikes because a token activation of the concept BICYCLE will activate a search among token memories wherein bicycles were present to find instances of riding behaviors.[1]

Crucially, it is the same process of categorization responsible for creating emotions that is responsible for perceptual inferences as just described: "In

1. Barsalou, as best I can tell, does not—as Barrett does—phrase the aim of situated action in terms of allostasis but rather in terms of goal achievement (Barsalou 2009, 1283). For those uncomfortable with the concept of allostasis, Barsalou's idea of goal achievement might prove an adequate replacement concept. I discuss this possibility more in my concluding remarks.

the same way that your brain used prior experience to [categorize] . . . visual sensations . . . it uses such knowledge to [categorize] bodily sensations" (Barrett et al. 2015, 89). While the brain makes categorizations based on all sorts of inputs (e.g., perceptual inputs, memory-based inputs), when it is focused on *affective* input—usually, but not necessarily, a large change in affective valence—the resulting categorizations are emotions (Barrett 2006, 2009a, 2017; Barrett et al. 2015). However, while affective states are the focus of the categorization process, they are not all that is needed for it: at times, brain states involving the same affective state can be correctly categorized as states of either fear or anger (Lindquist and Barrett 2008).[2] Consequently, more than affective input is needed: information from other sources (such as memory and perception) is required as well.

Affect—our conscious, ever-present sense of feeling—is an internal sensation that, according to Barrett, makes us aware of information about our bodily state pertinent to allostasis. Affect is the way in which we are aware of "interoceptive sensations," which are themselves "the representation and utilization of . . . the relevant statistical regularities . . . of the internal milieu" (Barrett 2017, 6). For example, a negative affective feeling of which we can become aware might plausibly indicate an increase in heart rate and blood pressure of which we are not aware directly and that, from the perspective of allostasis, would be better at lower levels. Categorizing this as a 'fear' state prepares us for action to alleviate the fear in order so that we, as organisms, achieve allostasis at that time. In this way, the psychological process I have described is, in Griffiths's terminology, the causal basis of so-called emotional behavior.

There are clearly some differences between the categorization involved in seeing as and the categorization constitutive of an emotion (other than the very fact that one is constitutive of something and the other is not). First, the categorization constitutive of an emotion does not categorize its focus, our affective state, as being an emotional state; rather, it categorizes the whole organism as being in that state. And this sort of object of categorization is distinct from perceptual categorizations, which 'focus' on perception and categorize the objects of perception.

2. The cited experiment involves priming people into an affective state by being asked to think of a past situation in which they were in that affective state. It is worth noting that the subjects could have recalled an anger or fear state as a result of this priming without actively categorizing their current selves as being in that state. I am thankful to an anonymous reviewer for pointing out this alternate interpretation of the data. That said, both interpretations of the experiment support the conclusion that emotion categories are not reducible to specific types of affective state, as in the alternate interpretation, the same affective state led to recalling states of either anger or fear across individuals.

Second, when I categorize my perceptual field as containing a bee, there is, if things go right, an object, the bee, which my perception is about, and this object existed before, and independent of, my perceiving it. Even if I am wrong about there being a bee, my perceptual representation is making a claim about what is out there independent of me: it is claiming there is a bee, even though there is not. Yet in the case of emotions, the input to the categorization process does not include anything, according to the TCE, that counts as an emotion before the act of categorization (Barrett 2006, 2012, 2014, 2017).

Barrett is certainly aware of these differences: because of them, she claims (2012, 2014, 2017; Barrett et al. 2015) that emotions are objects *created by* the process of categorization. That is, when I categorize myself as afraid, I am creating an object, fear. By claiming we create emotions by a process of categorization, what Barrett means is that we add novel epistemic properties—such as those that suggest action—to a brain state that did not have them before the categorization. A state of affect does not itself suggest any behavior; once it 'becomes' an emotion, then it does. A token emotion "creates meaning about the . . . value of . . . physical sensations, over and above [the sensations'] immediate sensorial valence and arousal . . . when physical sensations . . . are conceptualized as emotions" (Barrett et al. 2015, 103).

2.2. Emotions in the TCE. According to Barrett, emotions are the same kind of entity as what Searle (1995) calls social objects. For Searle, there is a class of objects that exist in the physical world but cannot be reduced to their physical basis, such as flowers and weeds being irreducible to plants.[3] As Barrett rightly notes: "When a plant serves as a flower or a weed, this creates meaning about the value of the plant: referring to a plant as a flower communicates that it is to be admired and cherished, while experiencing it as a weed brands it as something to be discarded. Flowers and weeds prescribe actions that mere plants cannot: flowers are to be cultivated and weeds are to be pulled from the ground" (2012, 417).

The way Searle understands these 'social objects' is that they exist when a community of people categorizes them as such: what makes a plant count as a flower or weed is not biological but a decision made by a community or society. Formally, a social object Y exists when some X, a physical object, counts (is categorized) as Y in context C (Searle 1995, 43), where a community is constitutive of the context. More generally, by 'social' I mean

3. There are also social objects, such as governments, that have no obvious physical basis in the way that flowers do; emotions are therefore identifiable with the latter kind of social object, as the physical basis is the emoter's brain state. See Thomasson (2003) for discussion of this distinction.

something predicated over multiple people or a community (Bechtel 2009). Searle's social objects count as 'social' in this sense insofar as they require a community or group as context; I will elaborate on their social nature in section 4.2.

The relevant point for the moment is that Barrett reaches the conclusion that emotions are social objects because she (rightly) believes that social objects are the sort of thing that is compatible with three major claims Barrett makes about emotions: (1) we can 'create' emotions not only in ourselves but in others; (2) emotions are, in a sense I will explicate, two-place predicates; and (3) emotion concepts are culturally relative.

First, as I noted previously, emotions are objects that we 'create'. Moreover, emotions are created not just in the self but in others when, for example, I perceive you to be afraid, you are afraid for me (Barrett 2009a, 2012; Barrett et al. 2015). That we can create emotions in others allows a (sort of) explanation of emotions in animals. For animals lack the conceptual repertoire to categorize themselves as afraid, but animals can count as afraid when we conceptualize them as such (Barrett 2012, 2014). Likewise, at least some social objects, such as flowers, are things we can create outside ourselves.

Second, emotions are *perceiver-dependent* phenomena (Barrett 2009a, 2012, 2017; Barrett et al. 2015), that is, objects that are 'created' by a perceiver (or, given the differences between perception and emotion creation, a quasi-perceiver). Barrett therefore claims (2012, 2017) that existential statements about emotions are only truth evaluable when expressed as two-place predicates, that is, with a predicate place for an emoter as well as a predicate place for a perceiver/creator.[4] That the emoter and perceiver-creator need to be expressed with distinct predicates reflects not only their perceiver dependence but that the perceiver and creator need not be the same organism. Again, social objects are likewise perceiver dependent, as perceivers are constitutive of a social context, C, in Searle's formulation of social objects. For Searle, if there is no context of categorizers to, say, decide there is a cocktail party, then there cannot be a cocktail party.

Barrett sees a third similarity between social objects and emotions: that the categories of both are socially constructed. Barrett notes that "emotion categories . . . vary as a function of learning, and in particular, how emotion words shape concept learning" (Barrett et al. 2015, 96). Barrett documents cultural variation in emotions as follows:

4. The number of predicates in question is the number of organisms necessary for the mental state to exist, not the number of predicates in the content of that state. In the sense of predication relevant to this article, mental states as generally discussed in philosophical literature are truth evaluable as one-place predicates. X's belief that A Φ's B if C, for instance, is a certain belief predicated over X.

> Cultural variation in emotion categories takes various forms. Some emotion categories exist only in specific cultures. For example, "ligit" is the experience of intense, euphoric aggression that occurs during head hunting in the Ilongot tribe from the Philippines (Rosaldo 1980). Some emotion categories appear to be universal, but their content and relational themes vary. . . . The experience of "sadness" is more akin to physical agony in Russian but the experience of loss in the USA (Wierzbicka 2009). . . . The same mental content can exist across cultures but be differentially configured as emotion categories. . . . In the USA, sadness and anger are experienced as separate and distinct emotions . . . whereas in Turkey sadness and anger are properties of a single emotion category called "kizginlik" (Mesquita 1993). (Barrett 2009b, 1284–85)

That emotional concepts vary across cultures in these three ways can be easily explained by the TCE if we posit that emotion categories such as anger, sadness, and so on, are culturally learned, or *socially constructed*, which Barrett does (Barrett 2009b; Barrett et al. 2015). If the concepts are culturally learned, then it is no mystery that cultures will either (*a*) develop different concepts to describe their affective states or (*b*) develop different logical relations among the same concepts.

3. Conceptual Issues in the TCE. There are three conceptual issues afflicting the TCE, which I will discuss in the following order: (1) a single person can have in theory infinite distinct emotional states, (2) we cannot be wrong about our own emotional state, and (3) infants and animals cannot have emotions without a linguistically capable adult to 'create' their emotions for them. In this section, I show that each of these three issues is an unavoidable implication of one of the three major claims discussed in section 2.2. I then discuss in each case what needs to be changed to make the TCE tenable; these changes comprise the beginning of my version of the TCE, the TCE*.

3.1. Having Multiple Emotions and Creating Emotions in Others. Picture the following scenario. A patient, Portnoy, comes into his therapist's office to talk about the latest maladaptive episode of fear he has experienced. His therapist, however, claims that Portnoy is not afraid but is in fact angry. Clearly, one of the two of them is wrong, and any good theory of emotion needs to account for this fact. Yet according to the TCE, neither is wrong: because we can create emotions in others, Portnoy is afraid for himself yet angry for his therapist.

And there seems to be no theoretical limit to how many emotional states poor Portnoy can have at the same time: if there are enough perceivers, he can have exactly as many emotions as we have emotion concepts. And this

conclusion seems extremely counterintuitive to say the least. Moreover, on some philosophical views of the emotions, the claim is not just counterintuitive, it is a conceptual impossibility. Helm (2007) argues that emotions are subject to 'tonal rationality', such that positive and negative emotions contradict each other in the same way beliefs in P and not-P would. On this view the idea that I am afraid for me and happy for you is not only incredibly implausible, it is downright incoherent. The issue, whether one of counterintuitiveness or actual incoherence, is directly tied to the idea that we can create emotions in others, and the only solution I can see is jettisoning the claim itself.

And this issue, of whether we can create emotions in others, is itself tied to the issue of whether emotions ought to be understood as two-place predicates. To understand why they ought not to be, consider this. There are two distinct senses of 'creation' relevant to this discussion, one that we might call metaphysically interesting and one that we might call metaphysically boring. Even if Searle's analysis of social objects is ultimately incorrect, it is, I think, quite clear that he identified a metaphysically interesting sort of phenomenon: a class of objects whose constitutive basis is outside of a person or people but that is created by people through acts of categorization. But when the creation in question is simply a single organism creating novel mental states in its own head, the creation is metaphysically boring. For I 'create' new mental states in my mind all the time. Just now, I created the belief that I had typed a new sentence ending in "all the time." What it means to 'create' these states is simply to perform a brain process constitutive of (the existence of) these states.

The difference between the two senses of creation is that, in the first instance, the creator can create an object distinct from herself, or create a new property of an object distinct from herself, whereas in the latter case, the perceiver or creator can only create something within herself. In the first case, genuine two-place predication is required: since the creator and the created are not analytically the same thing, it is informative to specify both. In the latter case, however, the creator and (possessor of) the created are analytically the same entity, so it is uninformative to specify both.

Whether there is any reason to consider emotions to be two-place predicates turns on whether we can create emotions in others. And I have argued that we cannot. I can now articulate the first two changes to the TCE that comprise TCE*: (1) we cannot create emotions in others, and, consequently, (2) emotions are best understood as one-place rather than two-place predicates.

It is true that, when perceiving other organisms in our environment, we impute to them emotion terms and, moreover, do so using the same general categorization process—and the same concept—by which we create emotions in ourselves. And we do indeed 'create' a mental state when we cat-

egorize another as afraid. However, the relevant mental state exists (and motivates behavior) in ourselves, not the person we are perceiving, and is better called 'imputing an emotion'. Further, genuine 'fear' and 'imputing fear' generate different sets of behaviors in the perceiver/creator, one properly characterized as 'fear behaviors' and the other as 'behaviors appropriate to responding to fear in another organism'.

3.2. Being Wrong, Believing, and Seeming. In the previous section, I noted that it had to be the case that either Portnoy or his therapist was wrong about his emotional state. What I did not emphasize is that it also must be the case that *either* one (or both) of them can be in error. For Portnoy might have a problem in admitting he gets angry: we can suppose that he is the sort of person who regularly proclaims, red faced and aggressively, "I am not angry!" when it is plain to all others that he is. A theory of emotions must therefore be able to explain Portnoy's disposition.[5]

If emotions are perceiver dependent in the way described by Barrett, it seems that we cannot be wrong about our own emotions. For if Portnoy categorizes himself as angry, then he is angry. And, as Barrett repeatedly points out, the categorization process is the same as the one underlying perceptual inference, so he is clearly conscious of its outputs if not its inner workings.

There is a natural way to cash out the intuition that a person can necessarily be wrong about her emotions even if categorizing oneself as being afraid is constitutive of being afraid. What it is to be wrong about one's own emotional state, according to this line of thinking, is to have a belief about one's emotional state. There is a theoretical construct in philosophy of perception called 'seemings' (Cullison 2010; Pace 2017); roughly, seemings are the truth-evaluable contents of visual experience that persist alongside, but can contradict, beliefs about the distal world based on experience. For instance, the Muller-Lyer illusion involves a seeming of two lines of unequal length that can persist alongside a belief that the two lines in front of you are equal. And the idea of seemings has been extended beyond perception to other mental phenomena; for instance, Oddie (2005) suggests desires are seemings of goodness, and more recently Carruthers (2017) suggests the same about affective valence.

If this line of thinking is correct, then the seeming would then necessarily be true: for each token instance of anger, Portnoy would have a seeming of being angry and would in fact be angry, insofar as the emotion itself and the seeming are generated by the same mechanism.[6] It would nonetheless still

5. I am thankful to an anonymous reviewer for helping me frame the issue in this way.

6. There is also a sense in which the seeming is necessarily false: the existence of a seeming state plausibly implies that there seems to be some object that exists independent of the seeming, and as I have noted, there is not in the case of emotions. And it may

be open to an agent to form a belief that she is not angry on other grounds. Portnoy could have a seeming that he is angry and refuse to believe it. Such a belief would be necessarily irrational, but that fact fits the phenomenon it is explaining: Portnoy's refusal to believe that he is angry is irrational.

The TCE* should then add a third claim, that the categorization mechanism constitutive of emotions delivers in seemings of being in an emotional state, not beliefs that one is in that state. After all, the mechanism is the same one that goes on in perceptual inferences, which (given the specific sort of perceptual inference Barrett refers to) delivers seemings as well: when I see something as a bee, it seems to me that there is a bee, even if I know that I am really seeing a wasp. Barrett herself argues that the mechanism is also responsible for generating occurrent beliefs (when the categorization focuses on the content of our thoughts; Barrett 2009a). But it seems to me it would not be a substantive change for the TCE to claim that the categorization mechanism issues in seemings, and something more is needed to get to belief.

3.3. Innate Emotion Concepts and Emotions in Infants and Animals. Humans are not the only kind of organism capable of having emotions. When anyone in earshot sets off fireworks, my dog looks for a space with a low ceiling, such as the space beneath my desk, hides in it, and begins to tremble. To interpret her behavior as issuing from anything other than an instance of fear seems counterintuitive, since our concept of fear explains perfectly what she does. In short, animals have emotions, and an account of emotions needs to explain how that is the case. And, without making further claims, the TCE* would seem to fail in this respect. For although Barrett's insistence that emotions are two-place predicates got her into trouble, it also provided her a way to account for emotions in animals by saying we (language-using organisms) create emotions in them. Whether or not her solution is adequate (and I believe it is not, as it would be unclear how animal research involving, e.g., fear homologues is to be understood if animals could not have emotions in their own right),[7] by removing the two-place

well be true that, if the categorization process constitutive of emotion also delivers seemings, then the seemings it delivers are necessarily false in this sense. But I am interested in explaining Portnoy's issue and therefore the sense in which his beliefs about his emotions are wrong, and that sense is necessarily one in which he can be either right or wrong. Because this is the sense in which I am interested in whether beliefs can be right or wrong, it is therefore also the sense in which I am interested in whether seemings can be right or wrong.

7. Further, if we can 'create' emotions in organisms that do not have emotion concepts, it is unclear why we should then be restricted to only 'creating' emotions in living organisms and not, say, our dead ancestors. I am thankful to an anonymous reviewer for making this point.

predication claim to fix the first issue, I have rendered that avenue unavailable.

I therefore want to suggest that some basic emotion concepts are innate and possessed by mammals. If, for instance, the concept 'fear' is innate and shared widely among mammals, then it is no mystery how dogs can be afraid: they have the concept for fear. Of course, not all concepts are innate, but claiming that the concepts for the basic emotions such as anger and fear are innate nicely supplements the TCE* by providing it with a plausible account of the difference between basic and complex emotions.

As mentioned above, however, emotion concepts exhibit significant cross-cultural variation. Yet I think that distinguishing between the idea (plausibly associated with the affect program hypothesis) that emotions themselves are innate in a way that they come with a prepackaged associated affective state, concepts, and patterns of behavior and the idea I propose, that emotion concepts are innate, can account for the cross-cultural variation.

By an innate concept, I mean something like the concepts of 'object' or 'agent' that we are born with according to the Core Knowledge theory of development (Carey and Spelke 1996; Spelke 2003). Roughly, concepts that are innate in this sense have a 'preprogrammed' intension but a variable extension and variable logical associations with other concepts. Given this conception of innate, we would expect for an innate concept that cultures have the same basic concepts but with different logical connections. For instance, the concept SELF is plausibly part of Core Knowledge, insofar as Core Knowledge includes agential concepts that require a concept SELF such as RECIPROCITY (Spelke and Kinzler 2007). Yet individualist (largely Western) societies and collectivist (largely Eastern) societies have different understandings of SELF; the latter, but not the former, define SELF in part in terms of relations to others (Markus and Kitayama 1991).

And, at least according to the findings Barrett uses that I summarized above, the same-concept-different-associations idea seems to explain cultural variation in emotion concepts. It would be completely commonplace for cultures to have different concepts associated with sadness, as, for example, Americans and Russians do (loss and agony, respectively). All that is required is that both cultures have the concept sadness. And it is plausible, if not entirely expected, that some cultures would create similar enough associations to the concepts 'anger' and 'fear' that they would decide that the two belong under the same umbrella concept, as, according to Barrett, the Turkish culture does. We can therefore add a fourth claim to the TCE*: that the concepts for basic emotions are innate and possessed by at least mammals.

4. The Social Dependence of Emotions. In light of the differences between TCE and TCE*, it is worth examining whether it is appropriate to identify emotions with social objects in the TCE*. In order to do so, I artic-

ulate a way of judging whether a phenomenon counts as a social object and argue that emotions in the TCE* do not. That they do not is fortunate, for, I argue, if emotions were social objects, then one of the major advantages of Griffiths's method—that it allows us to explain emotional phenomena mechanistically—would be called into doubt.

4.1. Constitutive Externalism. I claim that a phenomenon counts as a social object if it is *constitutively external* and, more specifically, constitutively social. Constitutive externalism is distinct from the more familiar content (Putnam 1975) and vehicle externalism (i.e., extended mind; Hurley 1998). In constitutive externalism, the properties that make something the kind of thing that it is—what I am calling the constitutive basis of that thing—go beyond its physical basis of that thing. An example will help illustrate this point. I am an uncle, which means that when my nephew says "uncle," he is referring to me. The physical basis of his uncle is my body: its boundary is my skin. But what makes me an uncle—what is constitutive of my uncle-hood—is to be found outside my body.

Constitutively internal phenomena are those wherein the properties that make them the sort of thing that they are can be found entirely within the physical basis. For instance, what makes gold count as gold is that its atoms have 79 protons in their nuclei; gold is constitutively internal.

The constitution of some phenomenon, the issue of what makes something the sort of thing that it is, can be contrasted with the causal-etiological history of that phenomenon (Salmon 1984; Ylikoski 2013). What constitutes, for example, a glass's being fragile is a question of its current structure, and the causal history of the glass only comes into play indirectly. Of course, it was its casual-etiological history—likely a manufacturing processes—that made the glass have the microstructural configuration that makes it fragile, but that is not part of the constitutive basis. The glass's fragility is constituted by properties internal to the glass, its microstructure, even if the history of those properties is external to the glass. Constitutive internalism is compatible with what we might call causal-etiological externalism.

4.2. Social Externalism. We might consider 'social externalism' a special class of constitutive externalism wherein the phenomenon's constitutive basis is social, that is, predicated over multiple people at the same time (Bechtel 2009). A phenomenon would be constitutively internal and causal-etiologically social if the constitutive basis is predicated over its physical basis, yet what made the constitutive/physical basis the way it is was influenced by social phenomena predicated over multiple people.

Social objects on Searle's account and emotions in the TCE (and TCE*) are both constituted by a process of categorization as such. Yet these pro-

cesses of categorization are distinct. Crucially, the categorization process is itself performed by a whole community in the case of social objects and by an individual in the case of emotions.[8]

I will first discuss social objects. For Searle, categorizations constitutive of social objects are committed by a group process of categorization. Crucially, this process of categorization is very different from the one described in section 2. It requires very few of the members to actually categorize via a psychological process; rather, it requires all of them to play certain roles. We do not all consciously (or unconsciously) decide that our government exists; rather, we act like it exists by respecting its regulations, paying taxes, and so on, and therefore it does. Even in examples in which it seems like only one person is doing the categorization, and doing it as a psychological process, the rest of the community is implicated not by categorizing but by playing the appropriate role.

Consider an agent at the Department of Motor Vehicles (DMV) who validates licenses by stamping them. Each time she stamps a piece of plastic with a picture and the right information, it becomes a license. In a sense, licenses are licenses because she categorizes them as such. First, it is unclear whether she actually has to be performing the psychological process outlined in section 2: she could be creating licenses by mindlessly stamping pieces of plastic while on the phone. But even if she does need to perform some psychological process that counts as categorization, it is not the case that she is the only person required for her to be able to create licenses. For some governmental agency had to put her in the position, and could take her out of it, and (almost) all of us must agree to participate in a civil society for a government to exist. And, people must care whether someone has a valid license for licenses to count as anything other than pieces of plastic—if, for instance, all the police stopped checking licenses on traffic stops, it is unclear whether the stamped pieces of plastic would still count as licenses. Social objects are, therefore, constitutively external: their constitutive basis is a categorization process predicated over multiple people. Further, it is not even clear whether the DMV operative actually needs to perform what I have been calling the psychological process of categorization; she could be stamping licenses mindlessly. The process of categorization that is predicated over multiple people is therefore necessarily distinct from the psychological process implicated in emotions.

8. In sec. 2, I claimed that social objects count as social insofar as they include a social community as context. This answer was incomplete, as it was left unclear in what sense the context of the social object counts as a part of it. I can now complete the answer: the community counts as part of the constitutive basis of the social object insofar as it is constituted by a process of categorization and that process is performed by (and therefore predicated over) the whole community.

4.3. Emotions Are Constitutively Internal. Emotions, unlike social objects, are created by an individual. In articulating her account of Conceptual Acts, Barrett writes: "Plants . . . become flowers or weeds . . . *in a human mind that exists in consensus with other human minds*. . . . Other minds might be absent at the proximal moment of perception [of an object to be categorized], but even then, the realness of flowers and weeds nonetheless depends on those minds in some distal way, because *other minds were necessary to transmit the categories in the first place*" (Barrett 2012, 417–18, italics added). Note that Barrett writes of the conceptualization happening "in a human mind" and "at [a] proximal moment of perception" (where perception is something an individual agent does) and that the role of the social group is essential only insofar as "other minds were necessary to transmit the categor[y] in the first place."

The psychological process that is itself constitutive of an emotion is being performed by (and therefore is predicated over) a single person. That person is using information she (at least may have) learned socially. Emotions are therefore an instance of a constitutively internal phenomenon that is etiologically social: the constitutive basis is a psychological process within a single person's brain, but what made the brain able to perform that process with that concept is the information the person learned from a community during her development.

Crucially, we can only conclude that emotions are constitutively internal from the fact that the categorization is performed by a single person *if emotions cannot be created in others*. I have been, for ease of explication, claiming that the categorization process is the entire constitutive basis of an emotion. And it is for the TCE* but not for the TCE, as the constitutive basis in the latter account includes both a perceiver and an emoter even though the categorization process is performed in the perceiver's mind. In virtue of being predicated over both organisms, the constitutive basis of emotions would be an external phenomenon: in the case of creating emotions in another person, it would count, under my definition, as social.

4.4. Why Constitutive Internalism Matters: The Search for Natural Kinds. According to Griffiths's method, to understand what emotions really are, one must not merely know where in the world the thing that instantiates the schema of emotions is—in the brain—but how to describe those brain processes in a way that supports generalizations. In his words, "emotions are the referents of the [natural] *kind terms* of theories that deal with emotional phenomena" (Griffiths 1997, 171, italics added).

Natural kinds are, roughly, categories in the sciences whose members share a cluster of interesting properties in virtue of a common underlying causal mechanism (Griffiths 1997; see also Boyd 1991; Wilson, Barker, and Brigandt 2007; Pober 2013); natural kind terms are the concepts used to rep-

resent those categories. This view of natural kinds is called the Homeostatic Property Cluster (HPC) view; for each natural kind there exists a cluster of properties, wherein each member has some number of those properties but not all members must have the same ones. What makes the properties count as a cluster, rather than a random grouping, is that they are all generated by a common underlying causal mechanism; the mechanism keeps the properties in 'homeostasis'.

If a class of objects has many stable properties in common, and has them in common because of a social practice, then these objects can be a 'natural' kind, and the practices can count as their underlying 'mechanisms' (Boyd 1991; Mallon 2003). Social objects such as flowers might plausibly count as 'natural' kinds in this sense: they have properties in common, such as their desirability, in virtue of our social practices of how we treat flowers. While it might be more appropriate to call such a category a 'social kind' rather than a natural kind, since the mechanism is itself something predicated over multiple people or a community, the distinction is merely semantic. For members of these 'social kinds' have a cluster of common properties in homeostasis just like members of bona fide natural kinds whose basis is, for example, the brain.

Now consider Griffiths's (1997) criticism of another view of emotions. According to the 'ecological' view, each emotion category evolved as an adaptive response to a certain sort of pressure; for instance, an ecological theorist might say fear states are those brain states that motivate behavior and evolved in response to pressure from predators. Emotion categories such as fear may well be natural kinds if the ecological theory were correct wherein "the causal homeostatic mechanism of each ecological category is a particular set of adaptive forces" (234). As Griffiths rightly notes, on the one hand, "specifying an adaptive problem tells us very little about the detail of the mechanism that solves the problem" (219), but on the other hand, "one of the main objects of psychology is to get behind the behavior of humans and other organisms and discover . . . the underlying mechanisms" (234–35). Likewise, if the mechanisms through which each instance of 'fear' or 'anger' (or each instance of an emotion generally) shared what properties they do were really social practices, then we would know nothing about the psychological mechanism underlying emotions.[9]

9. Mallon (2003) suggests that social practices themselves are reducible to beliefs about those social practices held among members of the community, but even if this is true, we would still lack a mechanistic story of emotion generation, since the relation between beliefs and emotions would still be mediated by the practices themselves. For instance, suppose that 'anger' instances are unified by social practices of sanctioning others for norm violations, which are in turn constituted by beliefs about what counts as a violation of social norms. My instances of anger are not constituted by my beliefs about norm vi-

Knowing the psychological mechanisms underlying emotions seems crucial not only if we value keeping the aim of psychology but also if we are to answer philosophical questions posed by emotions. For instance, to what extent we are morally responsible for either having or acting on our emotions plausibly depends at least in part on how much (self-)control we have over whether we experience or act on emotions. Yet how much control we have over either experiencing or acting on emotions is going to depend at least in part on the connections between the brain systems that underlie our capacity for self-control and those that underlie our emotions.

The issue of constitutive internalism versus externalism is directly relevant here, for there is a conceptual connection between a phenomenon's constitutive basis, on the one hand, and the mechanism underlying the unity among members of a natural kind on the other. The constitutive basis of some phenomenon can be understood as what makes the phenomenon the sort of thing that it is. If the phenomenon in question is a natural kind on the HPC view of natural kinds, then the mechanism underlying the cluster of properties its members have in common is what makes it what it is. Thus, if the constitutive basis of some 'natural' kind is a set of social practices, then the mechanisms sustaining the unity of its members are going to be exactly those social practices.

I do not want to commit myself in this discussion to comprehensive claims about what natural kind terms the TCE* (or TCE itself) warrants using. I do suggest, however, that emotion is a natural kind. Its members share a great many properties, including the crucial property of motivating emotional behavior to achieve or maintain allostasis, and they do so in virtue of a common causal mechanism: a process of categorization, focused on affect, that outputs an emotion category. That the categorization is focused on affect differentiates emotions from other acts of categorization leading to situated action.

According to the TCE, then, the constitutive basis of emotions, the perceiver/categorizer and the emoter, is social, so if 'emotion' is any sort of kind, then it is a social kind. But according to the TCE*, the process of categorization is a normal brain process. Emotions therefore have precisely the sort of mechanistic basis they were lacking in the TCE, and emotion is a natural kind with a psychological, rather than social, underlying causal mechanism.

5. Concluding Remarks. The TCE* started with four specific departures from the TCE: (1) We can only create emotions in ourselves. (2) Emotions are one-place predicates. (3) The categorization process constitutive of

olation; rather, they are constituted by something that is itself constituted by our beliefs about norm violation.

emotions delivers seemings of being in an emotional state. (4) The concepts for basic emotions are innate and possessed by at least mammals. These claims solved three major conceptual issues. The first two claims made it so that we can only have one emotion at a time. The third claim made it so that we can be wrong about our own emotions. And the fourth claim made it so that animals have at least basic emotions. From these claims—specifically the first two—I added a fifth claim: (5) Emotions are constitutively internal phenomena, not social objects. And from it, a sixth: (6) Emotion is a natural (and not social) kind whose underlying causal mechanism is the process of categorization.

Despite these differences, the TCE* retains much of the TCE. It retains virtually everything discussed in section 2.1: a process of categorization, focused on affect, which results in categorizing the self with an emotion concept, is constitutive of token emotional states. This process of categorization is quite like the one implicated in the phenomenon of *seeing as*: indeed, insofar as they both deliver in seemings, the process on the TCE* might be more like the perceptual categorization process than the one posited by the TCE itself. More generally, the *brain basis* of an emotion—what is happening in an organism's brain when it is in an emotional state—is the same in the TCE and TCE*.

I am tempted to add a seventh claim to the TCE*. Specifically, I must admit concern about the claim that all situated action, including emotional action, is aimed at achieving or maintaining allostasis. I am more sympathetic to Barsalou's (2009) claim that situated action is aimed at goal achievement more generally and am myself inclined to think that situated action is aimed at something like maximizing received reward. Yet while determining what emotional action, or situated action more generally, is aimed at is a worthwhile project, all that is necessary for the TCE* to be tenable, and more specifically for claim 6 to be true, is that motivating actions aimed at something similar must be in the property cluster of the natural kind 'emotion'.

And there is still more to be discussed to fully answer the question of what emotions are. In addition to specifying what natural kind terms the TCE* warrants other than 'emotion', there is still the issue of figuring out exactly how emotions qua mental states relate to brain states. For if emotions are not social objects, what are they? I have argued that a token emotional state is constituted by a token act of categorization. Yet how to understand that constitution relation would seem to depend on one's prior metaphysical commitments: for reductionists, a token emotion is reducible to a token brain state involving an act of categorization, for role functionalists, token emotions would supervene on just those brain states, and so on. Nonetheless, I believe progress has been made. Whatever one's preferred mental/physical relation, I have articulated which physical states are the appropriate relata. And I have done so in a way that both renders the TCE*

free of conceptual issues plaguing the TCE and allows for the search for natural kinds in the brain to proceed.

REFERENCES

Barrett, Lisa Feldman. 2006. "Solving the Emotion Paradox: Categorization and the Experience of Emotion." *Personality and Social Psychology Review* 10:20–46.
———. 2009a. "The Future of Psychology: Connecting Mind to Brain." *Perspectives in Psychological Science* 4:326–39.
———. 2009b. "Variety Is the Spice of Life: A Psychological Construction Approach to Understanding Variability in Emotion." *Cognition and Emotion* 23:1284–1306.
———. 2012. "Emotions Are Real." *Emotion* 12:413–29.
———. 2014. "The Conceptual Act Theory: A Precis." *Emotion Review* 6:292–97.
———. 2017. "The Theory of Constructed Emotion: An Active Inference Account of Interoception and Categorization." *Social Cognitive and Affective Neuroscience* 12:1–23.
Barrett, Lisa Feldman, Christine D. Wilson-Mendenhall, and Lawrence W. Barsalou. 2015. "The Conceptual Act Theory: A Road Map." In *The Handbook of Psychological Construction*, ed. Lisa Feldman Barrett and James A. Russell, 83–110. New York: Guilford.
Barsalou, Lawrence W. 1999. "Perceptual Symbol Systems." *Behavioral and Brain Sciences* 22: 577–660.
———. 2009. "Simulation, Situated Conceptualization, and Prediction." *Philosophical Transactions of the Royal Society of London* B 364:1281–89.
Bechtel, William. 2009. "Explanation: Mechanism, Modularity, and Situated Cognition." In *Cambridge Handbook of Situated Cognition*, ed. Philip Robbins and Murat Aydede, 155–70. Cambridge: Cambridge University Press.
Boyd, Richard C. 1991. "Realism, Anti-foundationalism, and the Enthusiasm for Natural Kinds." *Philosophical Studies* 61:127–48.
Carey, Susan, and Elizabeth Spelke. 1996. "Science and Core Knowledge." *Philosophy of Science* 63:515–33.
Carruthers, Peter. 2017. "Valence and Value." *Philosophy and Phenomenological Research*. doi:10.1111/phpr.12395.
Cullison, Andrew. 2010. "What Are Seemings?" *Ratio* 23:260–74.
Ekman, Peter. 1984. "Expression and the Nature of Emotion." In *Approaches to Emotion*, ed. Klaus Scherer and Peter Ekman, 319–43. New York: Erlbaum.
Griffiths, Paul. 1997. *What Emotions Really Are: The Problem of Psychological Categories*. Chicago: University of Chicago Press.
Helm, Bennett W. 2007. *Emotional Reason: Deliberation, Motivation, and the Nature of Value*. New York: Cambridge University Press.
Hurley, Susan. 1998. "Vehicles, Contents, Conceptual Structure, and Externalism." *Analysis* 58:1–6.
Izard, Carroll E. 1997. "Emotions and Facial Expressions: A Perspective from Differential Emotions Theory." In *The Psychology of Facial Expression*, ed. James A. Russell, Fernández Dols, and José Miguel, 57–77. New York: Cambridge University Press.
———. 2007. "Basic Emotions, Natural Kinds, Emotion Schemas, and a New Paradigm." *Perspectives on Psychological Science* 2:260–80.
Kovach, Adam, and Craig DeLancey. 2005. "On Emotions and the Explanation of Behavior." *Nous* 39:106–22.
Kripke, Saul. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.
LeDoux, Joseph. 1996. *The Emotional Brain*. New York: Simon & Schuster.
———. 2012. "Rethinking the Emotional Brain." *Neuron* 73:653–76.
———. 2015. "Afterword: Emotional Construction in the Brain." In *The Psychological Construction of Emotion*, ed. Lisa Feldman Barrett and James A. Russell, 459–64. New York: Guilford.
Lindquist, Kristen A., and Lisa Feldman Barrett. 2008. "Constructing Emotion: The Experience of Fear as a Conceptual Act." *Psychological Science* 19:898–903.

Lindquist, Kristen A., Tor D. Wager, Hedy Kober, Eliza Bliss-Moreau, and Lisa Feldman Barrett. 2012. "The Brain Basis of Emotion: A Meta-analytic Review." *Behavioral and Brain Sciences* 35:121–43.
Mallon, Ron. 2003. "Social Construction, Social Roles, and Stability." In *Socializing Metaphysics*, ed. Frederick Schmitt, 327–53. Lanham, MD: Rowan & Littlefield.
Markus, Hazel Rose, and Shinobu Kitayama. 1991. "Culture and the Self: Implications for Cognition, Emotion, and Motivation." *Psychological Review* 98:224–53.
Marr, David. 1982. *Vision*. New York: Freeman.
Mesquita, Batja. 1993. "Cultural Variations in Emotions: A Comparative Study of Dutch, Surinamese, and Turkish People in the Netherlands." PhD diss., University of Amsterdam.
Oddie, Graham. 2005. *Value, Reality and Desire*. New York: Oxford University Press.
Pace, Michael. 2017. "Experiences, Seemings, and Perceptual Justification." *Australasian Journal of Philosophy* 95:226–41.
Pober, Jeremy M. 2013. "Addiction Is Not a Natural Kind." *Frontiers in Psychiatry* 4:123.
Prinz, Jesse J. 2004. *Gut Reactions*. New York: Oxford University Press.
Putnam, Hilary. 1975. "The Meaning of 'Meaning.'" *Minnesota Studies in the Philosophy of Science* 7:215–71.
Rosaldo, Michelle. 1980. *Knowledge and Passion: Ilongot Notions of Self and Social Life*. Cambridge: Cambridge University Press.
Russell, James A. 2003. "Core Affect and the Psychological Construction of Emotion." *Psychological Review* 110:145–72.
Salmon, Wesley C. 1984. "Scientific Explanation: Three Basic Conceptions." In *PSA 1984: Proceedings of the 1984 Biennial Meeting of the Philosophy of Science Association*, vol. 2, 293–305. East Lansing, MI: Philosophy of Science Association.
Scarantino, Andrea. 2009. "Core Affect and Natural Affective Kinds." *Philosophy of Science* 76: 940–57.
———. 2012. "How to Define Emotions Scientifically." *Emotion Review* 4:358–68.
Searle, John R. 1995. *The Construction of Social Reality*. New York: Free Press.
Spelke, Elizabeth S. 2003. "What Makes Us Smart: Core Knowledge and Natural Language." In *Language in Mind: Advances in the Study of Language in Thought*, ed. D. Gentner and S. Goldin-Meadow, 277–311. Cambridge, MA: MIT Press.
Spelke, Elizabeth S., and Katharine D. Kinzler. 2007. "Core Knowledge." *Developmental Science* 10:89–96.
Thomasson, Amie. 2003. "Realism and Human Kinds." *Philosophical and Phenomenological Research* 67:580–609.
Wierzbicka, Anna. 2009. "Language and Metalanguage: Key Issues in Emotion Research." *Emotion Review* 1:3–14.
Wilson, Robert A., Matthew J. Barker, and Igno Brigandt. 2007. "When Traditional Essentialism Fails: Biological Natural Kinds." *Philosophical Topics* 35:189–215.
Wilson-Mendenhall, Christine D., Lisa Feldman Barrett, W. Kyle Simmons, and Lawrence W. Barsalou. 2011. "Grounding Emotion in Situated Conceptualization." *Neuropsychologia* 49: 1105–27.
Ylikoski, Petri. 2013. "Causal and Constitutive Explanation Compared." *Erkenntnis* 78:277–97.