

City University of New York (CUNY)

CUNY Academic Works

Dissertations, Theses, and Capstone Projects

CUNY Graduate Center

10-2014

Constitutively Embodied Emotions

Daniel Shargel

Graduate Center, City University of New York

[How does access to this work benefit you? Let us know!](#)

More information about this work at: https://academicworks.cuny.edu/gc_etds/472

Discover additional works at: <https://academicworks.cuny.edu>

This work is made publicly available by the City University of New York (CUNY).

Contact: AcademicWorks@cuny.edu

CONSTITUTIVELY EMBODIED EMOTIONS

by

DANIEL SHARGEL

A dissertation submitted to the Graduate Faculty in Philosophy
in partial fulfillment of the requirements for the degree of Doctor of Philosophy,
The City University of New York

2014

This manuscript has been read and accepted for the
Graduate Faculty in Philosophy in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

Barbara Montero

Date

[required signature]
Chair of Examining Committee

John Greenwood

Date

[required signature]
Executive Officer

Jesse Prinz

David Rosenthal

Noel Carroll
Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

CONSTITUTIVELY EMBODIED EMOTIONS

by

Daniel Shargel

Adviser: Professor Jesse Prinz

My primary thesis is that emotions are partially constituted by bodily states. My view derives from the James-Lange tradition, but contrary to Neo-Jamesian theories, I claim that emotions are partially constituted by integrated peripheral bodily states and brain states, rather than bodily perceptions. This view may seem vulnerable to two obvious critiques: emotions, unlike bodily states, have intentional objects, and neuroscientists have already identified the neural basis of emotions, so there is no reason to look for constituents outside of the brain. I argue on the basis of social psychology research that emotions are not intentional states, since they do not influence thought and behavior in a manner that is specific to any purported objects or contents. In my discussion of intentionality I point to a number of factors that collectively give the impression that emotions have objects. Regarding the second critique, I argue that current research in affective neuroscience tells us much less than it seems to about the constituents of emotions. A great deal of confusion in that literature results from failure to carefully distinguish emotions from desires. I also illustrate different ways that positions on the embodiment of emotions influence interpretations of the empirical data.

Acknowledgements

First of all, I would like to thank my advisor Jesse Prinz. I initially developed my interest in researching emotions after reading his *Gut Reactions*, and he has always encouraged me not to waver from my accidental radicalism. I'm grateful to my committee members Noel Carroll, Eric Mandelbaum, Barbara Montero, and David Rosenthal for their support, and for making my defense a career highlight. I also thank David Rosenthal for his guidance throughout my graduate studies, and for the countless subtle ways he has influenced my approach to philosophy.

I received valuable feedback from emotion researchers too numerous to name, often at conferences where I delivered talks based on these chapters. Thanks for giving me a much needed sanity check, particularly on my interpretation of empirical work.

My friends in the program have given me all sorts of invaluable help, from ruthless feedback during talks to the Cognitive Science Symposium, to detailed notes on drafts of my papers and chapters, to long sessions in front of laptops at coffee shops and bars. I could always rely on their insight, generosity and good company.

I can't thank my brother and parents enough for their confidence in me, and their support throughout this lengthy process. My interest in philosophy also traces back to countless conversations over dinner together.

Finally, my wife Adina was there with me on the front lines when things got rough, and nobody else had more to do with my actually finishing this dissertation. To say that I am extremely grateful to her would be accurate.

Table of Contents

Abstract	iv
Acknowledgments	v
1 The Emotional Effects of Bodily Perceptions	1
1.1 Introduction	1
1.2 Lange's Premise	3
1.3 The James-Lange Theory	11
1.4 The Inter-systemic View	23
1.5 Conclusion	30
2 Inter-systemic States	31
2.1 Introduction	31
2.2 The Constitution of Emotions	32
2.3 Marks of the Mental	39
2.4 Conclusion	45
3 Emotions without Objects	47
3.1 Introduction	47
3.2 The Empirical Case	48
3.3 Philosophical Responses	64
3.4 Conclusion	71
4 Looking for Emotions in the Brain	73
4.1 Introduction	73
4.2 From Folk Psychology to Affective Neuroscience	75
4.3 The Ventromedial Prefrontal Cortex	86
4.4 The Amygdala	95
4.5 Conclusion	100
Conclusion	101
Bibliography	103

Chapter 1: The Emotional Effects of Bodily Perceptions

1.1 Introduction

Going back to William James (1922), the primary philosophical question about emotional phenomenology is whether it is exhausted by bodily perceptions. James (1992) took it as a datum, apparent through introspection, that bodily perceptions do exhaust emotional phenomenology. I will refer to this for convenience's sake as *James' Premise*. If you accept James' Premise, and you also accept that emotions just are what we feel during emotional episodes, then it follows that emotions are bodily perceptions. I will refer to this as the *perceptual view*. However, if you deny one or both of those premises, then various other theoretical positions become available.

The prominence of James' Premise has obscured a related consideration raised by his contemporary Carl Lange (1922).¹ Lange (p. 66) says that,

As a matter of fact, it is not difficult to prove now, and by means of the most ordinary and well known experiences, that emotions may be induced by a variety of causes which are utterly independent of the disturbances of the mind, and that, on the other hand, they may be suppressed by pure physical means.

Lange (pp. 66-8) goes on to discuss some of the ways that bodily manipulation influences emotions, such as when we drink alcohol, take hallucinogenic drugs or have a cold shower. As he says, it is easy to think of more examples: e.g. the calming effect of taking deep breaths, or going on a morning run to reduce frustration throughout the day. I will refer to this claim, that one can

¹ In fact, James and Lange each make both of these points, though James emphasizes the former, and Lange the latter.

produce or suppress emotions by manipulating the bodily periphery², as *Lange's Premise*.

According to Lange, the best explanation for these effects is that emotions are identical to peripheral states. I will refer to this as the *peripheral view*. The peripheral effects that Lange discusses are mediated by bodily perceptions, so Lange's Premise can be construed as a claim about the causal effects of emotional phenomenology. In this chapter I will show that, construed in this manner, Lange's Premise disrupts a wide range of theories, his own included.

Lange's Premise is most conspicuously threatening to theories which identify emotions with judgments or judgment-like propositional attitudes, which include cognitivist theories in philosophy and some appraisal theories in psychology.³ It is difficult for these theories to explain why certain varieties of peripheral manipulations would produce emotional changes. James and his contemporary followers can provide a much more natural explanation. James and Lange agree that emotional phenomenology is exhausted by bodily perception, and bodily perception produces the psychological features that are distinctive of emotional episodes. They differ over whether bodily perceptions are the effects of emotions or are identical to them. It seems, therefore, that Lange's Premise and James' Premise together serve as evidence in favor of one of their views being correct, and other considerations altogether would be needed to choose between them.

² Since the brain is part of the body, I will use the terms 'bodily periphery', 'peripheral', and 'peripheral state' in place of the more familiar 'body', 'bodily', and 'bodily state' when discussing the body beyond the brain. I will use 'bodily perception' to refer to perceptions of one's own peripheral states.

³ Solomon (1988) and Nussbaum (2001) defend cognitivist theories, and Ellsworth and Scherer (2003) survey appraisal theories.

In this chapter I will not discuss cognitivism or appraisal theory further, since chapter 3 will focus on undermining those theories. Instead, I will argue here that Lange's Premise actually undermines the leading theories in the James-Lange tradition. I will propose a new theory of emotions which is inspired by Lange's and informed by the problems that I identify.

In part 2 I will review some of the empirical evidence in support of Lange's Premise. In part 3 I will develop critiques of the most prominent theories in the James-Lange tradition. I will begin by discussing the Neo-Jamesian theory developed by Jesse Prinz (2004), before turning to Lange's own theory. Finally, in part 4 I will present a variant of Lange's theory that I call the inter-systemic view, and argue that it is best suited to account for Lange's Premise.

1.2 Lange's Premise

When Lange claimed that it is possible to elicit or alter emotions by manipulating the body, his evidence consisted largely of personal observations and popular platitudes. However, today there is a diverse body of empirical research that backs up his point. In this section I will survey this evidence, before considering the implications for conventional theories of emotions.

1.2.1 Empirical Support

I will start with studies showing that peripheral manipulation produces emotional effects. Facial feedback effects are the most-researched method of peripheral manipulation. Strack et al. (1988) manipulated participants facial expressions by having them hold pencils in their mouths in two different ways - one that produced a smile, and another that produced a frown. The

participants, unaware that they were making these expressions, judged stories more amusing when smiling than when frowning. Soussignan (2002), attempting to replicate Strack et al.'s results, found that only Duchenne smiles (which involve orbicularis oculi eye muscles) produce emotional effects.

There is also significant evidence for other varieties of peripheral feedback. For example, Stepper and Strack (1993) demonstrated that posture influences expressions of pride, and frowning of brow muscles influences subjective effort. Philippot et al. (2002) identified patterns of breathing that are distinctive of joy, anger, fear and sadness, and demonstrated that producing those breathing patterns is sufficient for inducing corresponding emotional responses. This is perhaps the most intuitive of all of these results, since folk psychology recommends using breathing patterns to cope with unwelcome emotional arousal.

Other evidence on peripheral effects is drawn from neuroscience. The insula, a brain region which receives input from the body, is typically active during emotional episodes, and is deeply interconnected with a set of brain areas that are involved in motivation (Craig 2009). A recent study by Berntson et al. (2011) found that patients with insula lesions reported less arousal in response to both positive and negative stimuli than normal participants, and reported weaker positive and negative reactions as well. Damage to the somatosensory cortex, which is involved in bodily perception, impairs visual recognition of different types of facial expressions (Adolphs et al. 2000).

This evidence, taken as a whole, supports two conclusions. First, the evidence from peripheral feedback studies, combined with the evidence about the role of the insula, suggest that bodily perception produces characteristically emotional motivational effects. Second, the

research on the roles of the insula and somatosensory cortex suggests that bodily perception plays a critical role in making us aware of our emotions. The former confirms Lange's Premise, and the latter will also play a significant role in the theory evaluation to follow.

1.2.2 Peripheral Specificity

There are two main empirical challenges that both the perceptual and peripheral theories face. The first could be called peripheral specificity. If emotions are peripheral states, then each different type of emotion must correlate with a different type of peripheral state. The perceptual view also requires each type of emotion to correlate with a different type of peripheral state, since otherwise perceptions of those peripheral states will not themselves be distinct. On both views, there must be enough variation among the relevant peripheral states to explain the variation among emotions.

I will discuss the relevant research shortly, but first it is worth clarifying how emotions are individuated. The peripheral view states that every type of emotion is a type of peripheral state, but is it also committed to the view that every type of peripheral state is a type of emotion? If not, how do we distinguish emotions from other sorts of peripheral states? Similarly, does the perceptual view consider all bodily perceptions emotional states? If not, how do we distinguish emotions from other bodily perceptions?

I think the most charitable way of answering these questions is to focus on the functions of emotions. Theories in the James-Lange tradition deny that emotions are cognitive states which represent relations to one's environment such as danger, loss, and contamination — relationships that Lazarus (1988) calls 'core relational themes'. However, embodied theories can still assert

that the function of emotions is to help us cope effectively with themes, and that emotions are normally elicited by systems which have the function of detecting themes. In order to distinguish emotions from other similar states, we just need to determine whether a given type of peripheral state is normally elicited in response to some theme. If so, and only if so, those peripheral states, or the perceptions of those states, count as emotions. The peripheral states and bodily perceptions related to drives such as hunger and thirst would not qualify, since they have the function of maintaining some interior balance (the correct supply of energy or hydration) rather than helping us relate to some external condition.

Returning to the question of specificity, embodied theorists need to show that there is a one-to-one relationship between emotion types and peripheral state types. It would be a problem if there were two peripheral states for one emotion, or one peripheral state for two emotions. However, if there seems to be a mismatch, it is difficult to tell whether we should reject the theory, or if we should alter our method for individuating emotion-types. Imagine that fear correlated with two quite different peripheral state types, or that jealousy and anger had the same peripheral correlates. It is open to the theorist to propose that fear comes in two distinct varieties, or that jealousy and anger are not intrinsically different. In order to settle this kind of issue we need to have an independent grip on how to individuate emotions. This problem is not entirely intractable, but approaches to individuation in the literature are currently so diverse and contradictory that no solution is immediately at hand.

I will now turn to the empirical evidence. The peripheral states that embodied theories address are governed by the peripheral nervous system (PNS). The PNS is divided into the autonomic nervous system (ANS) and the somatic nervous system (SNS). The SNS governs the

types of muscles that we can voluntarily control. This system is responsible for our facial expressions, vocalizations, level of muscle tension, and coordinated bodily movements such as walking, all of which change in characteristic ways during emotional episodes.

The ANS governs the level and manner of bodily arousal. Sympathetic changes increase arousal and parasympathetic changes decrease it, though arousal is not an entirely unified phenomenon (Folkow 2000). A state of arousal will increase some combination of heart rate, perspiration, blood flow to various parts of the body, and visceral activity, and each of these changes occur during certain emotional episodes (Kreibig 2010). As I mentioned earlier, there are two branches of the PNS: the ANS and the SNS. James-Lange theorists do not need to demonstrate autonomic specificity (a one-to-one relationship between emotion types and autonomic state types) or somatic specificity (the same, but with somatic state types). It is fine in principle for two emotions to correlate with the same autonomic state as long as they correlate with different somatic states, and vice versa. However, research on peripheral specificity has tended to focus on one system or the other, so I will discuss them in turn.

The most prominent proponent of somatic specificity is Paul Ekman, who argues that there are universal facial expressions that correlate with basic emotions. Ekman (Ekman and Friesen 1971) found that all participants, including those who had not been exposed to western civilization, associated the same facial expressions with the same emotions. Specifically, this holds for the six emotions - fear, anger, sadness, joy, disgust and surprise - that he argued are basic.

The standard critique of Ekman is that he used a forced choice paradigm (Russell 1994). He got his results by giving participants a limited set of emotion-types to choose from, and

results are much more varied when participants are free to use whatever words they like to describe facial expressions. However, as Prinz (2004) argues, it is suggestive that even a forced choice paradigm was able to find any culturally invariant correlations between facial expressions and emotions. The perceptual theorist is not committed to somatic state types, by themselves, perfectly correlating with emotion types. In addition, embodied theories are not committed to emotions being entirely innate. Facial expressions might combine with other somatic and autonomic changes to collectively differentiate between types of emotions, and as Prinz emphasizes, the nature of emotions might be partially a function of cultural influence.

In addition to studies of facial expressions, there has been some research on other somatic correlates such as whole-body movement and vocal inflection.⁴ However, there is a great deal of research on autonomic changes, such as heart rate, blood pressure and skin conductance. A number of studies, including Ekman et al. (1983), attempted to show that different autonomic changes correlate with different emotions. A prominent review of such research by Cacioppo et al. 2000 reported mixed results, with some emotions easier to distinguish than others. For example, there is greater heart rate acceleration in anger, fear and sadness than in disgust, greater skin conductance (the result of sweating) in disgust than happiness, greater vascular activity in anger than fear, and greater cardiac activity in fear than anger. Some recent reviews, such as Sequeira et al. 2009 and Kreibig 2010 are confident that emotions have distinct autonomic signatures, though many emotion researchers remain skeptical about these results.

Skeptics of the embodied views can point to the vast number of studies and inconclusive results. Embodied theorists, on the other hand, can emphasize the methodological difficulties in

⁴ See e.g. Sauter et al. 2012 and Dael et al. 2011.

doing this research, and the variety of correlations that have been demonstrated. Perhaps we will soon develop better methods of individuating and reliably eliciting emotions, but at this point it is not possible to draw any strong conclusions about peripheral specificity. Until then, embodied views will have to rise or fall based on other issues.

1.2.3 Eliminating Peripheral Inputs

Another ground for concern is rooted in an old debate about patients with severed spinal cords. Walter Cannon (1927) famously argued that these patients had normal emotional experiences. Since they lacked bodily feedback, emotions must not be identical to bodily perceptions. This critique has just as much force against peripheral theories, since bodily perception would on that view mediate the effects of emotions on motivation. More recently, Hohmann (1966) showed that spinal cord injuries partially but not entirely diminish emotional responses, a result that was disputed by Chwalisz et al (1988).

Heims et al. (2004) focused on a different population of patients: those with autonomic failure, rather than spinal cord injuries. They found that these patients performed normally on the Iowa Gambling Task, which is considered by some to test normal emotional responses, but worse at attributing emotions to characters in shorts stories. I will argue in chapter 4 that deficient desires, not emotions, cause deficits on the Iowa Gambling Task, so we should not expect an emotion deficit to necessarily influence performance on that task. On the other hand, evidence that autonomic failure makes it more difficult to attribute emotions does support perceptual or peripheral theories. As the authors suggest, “The mild underperformance of patients on this task suggests that autonomic bodily responses may play a role in predicting, perhaps through

empathetic emulation, the subjective emotional feeling states of others” (p. 1984). These patients had only autonomic deficits, with somatic responses intact, which may explain why the effects were not particularly strong. It is hard to explain, on traditional emotional theories, why autonomic failure should interfere with effective empathetic emulation.

However, it is still necessary for perceptual or peripheral theories to explain why emotional phenomena are not entirely eliminated by spinal cord injuries. Prinz (2004) and Damasio (1999) provide a variety of replies. First of all, these patients do have some peripheral input from their faces, or other peripheral input that remains after the injury, so embodied theories would predict that they would have emotions. Their emotions would just implicate a narrower range of peripheral states. Second, it is also possible that these patients, when they have emotional experiences, are actually hallucinating peripheral states. On the perceptual view, this would mean that they were in real emotional states which consist of hallucinatory perceptions. On the peripheral view, the patients would not be in real emotional states. By hallucinating the relevant peripheral states, they would be hallucinating that they are in emotional states.

Again, the research on the effects of impaired peripheral inputs are suggestive but not quite conclusive. Preventing peripheral states from influencing the brain does, as embodied theories predict, seem to interfere with normal emotional functioning. However, the research does not make it clear whether all emotional effects are dependent on these inputs, as perceptual and peripheral theories would predict. Given the strength of the arguments for embodied theories, and the inconclusiveness of the evidence levied against them, their position seems strong. Pending further evidence or better counter-arguments, the question is not whether, but in what manner emotions are influenced by peripheral inputs.

1.3 The James-Lange Theory

Prominent theories in the James-Lange tradition seem well-suited to explain Lange's Premise, since they predict that peripheral manipulation would result in altered emotional states. However, I will argue that these theories cannot explain how and why these changes occur. I will take Jesse Prinz as a representative of James' branch, before turning to Carl Lange as a representative of his own view for lack of a recent proponent.

1.3.1 Embodied Appraisals

In *Gut Reactions* (2004), Prinz combines a perceptual theory of emotions with a teleosemantic theory of reference. Consider an individual, Ben, who is afraid of spiders. Ben sees a spider, or concludes for some reason or other that a spider is present. As a result, he will have a spider-perception, or an occurrent belief with the content 'there is a spider here'. Either of these states would be in what Prinz calls a calibration file. Calibration files are sets of mental states that cause the same type of peripheral change. Each of these types of peripheral change is correlated with a different emotion. So if Ben has a spider-perception or the thought 'there is a spider here', the presence of that state in the relevant calibration file will cause him to enter into the peripheral state that is correlated with fear.

That state is not, according to Prinz, fear itself. Ben's fear is instead his subsequent perception of that peripheral state. This identification is, on the face of it, counter-intuitive. Ben is not afraid of his accelerated heart-beat or any of the other physical changes he is undergoing.

He's afraid, we would ordinarily say, of the spider. If that is so, how can fear be a perception of a bodily state? Prinz replies to this objection by appealing to his theory of representation.

He defends a version of the causal theory of representational content. A state carries information about whatever it reliably predicts, and it represents whatever it has the function of reliably predicting (p. 53). Prinz distinguishes between the real and nominal contents of a representation (pp. 67-9). Mental states represent objects by carrying information about their features. When those are merely features of that object, given the natural kind that it falls under, then those features are the nominal content of the representation. The real contents are the essential features of the natural kind under which that object falls. When the state carries information about essential features of the object it represents, then those features are themselves the real content of the representation.

Ben's fear, according to Prinz, is his perception of being in a certain peripheral state. The nominal content of that representation is something to the effect of 'my bodily periphery is altered in such and such a way'. However, the real content is something else entirely. It may be paraphrased as 'I am in danger'. This is an example of what, following Lazarus (1991), he calls core-relational themes (which I will call 'themes', for short), by which he means "a relation that pertains to well-being" (Prinz 2004, 15). The real contents of emotions, according to Prinz, are themes. Emotions represent themes by way of the peripheral states that carry information about them. The peripheral states, in turn, carry information about themes because any mental state that carries information about a given theme will be in the calibration-file that causes the corresponding type of peripheral state to occur.

In short, emotions are bodily perceptions, which through registering those peripheral features represent a theme. Hence his description of emotions as embodied appraisals, or more colloquially, gut reactions. Themes are of vital importance, so it is in our interest to keep appraised of them. That is the job of the emotions.

1.3.2 Bodily Perceptions

My first critique of Prinz's view is that he mischaracterizes the role of bodily perceptions. Consider someone who has just been fired, and whose awareness of that fact causes a tightness in his chest, and a general feeling of weakness in his limbs. He doesn't need to learn that he has suffered a setback which may affect his well-being. By the time he perceives the peripheral change he is already responding to that setback. Instead, what the perception tells him is that he is in an intense state of fear or anxiety. If he wants to avoid manifesting his emotion through his appearance and actions, certain precautions will have to be made. This suggests that contrary to Prinz, bodily perceptions do not have the function of telling us about themes, they tell us about our emotions.

There are cases in which bodily perceptions seem to tell us about themes. It is familiar from popular culture that feeling the hair stick up on the back of your neck means you are in danger, and feeling a warm glow while on a date means that you have found "the one". Bracketing the reliability of these judgements, these are actually cases where we infer that we are in a given theme on the basis of a perceived emotion. Piloerection is a peripheral correlate of fear, and flushing is a peripheral correlate of sexual arousal. The systems which normally elicit fear and sexual arousal presumably have the function of identifying the presence of danger and a

suitable mate, respectively. If you trust that those systems are reliable you can infer that you are in the relevant themes on the basis of the emotions that the systems produce, but that is an inference you need to make. The bodily perception itself just tells us that we are in a given type of emotional state.

Prinz could reply that I have not addressed his teleosemantic theory. If piloerection carries information about danger, then perception of that piloerection will carry that information as well. However, it is not enough for a state to carry information. Information, after all, is cheap. On Prinz's view, a representational state needs to have the function of carrying that information. We do not have direct access to the function of a state, since on Prinz's (pp. 54-5) view this is dependent in part on evolutionary history. However, we would have no basis for thinking that bodily perceptions have the function of representing themes unless this is reflected in their functional role. Specifically, they should play this role under normal conditions - the sort of conditions that explain the fact that we have inherited the faculties that allow us to enter such states from our distant ancestors (Millikan 1993). However, the actual role that bodily perceptions normally play, which is as far as we can tell the role that similar states played under the relevant historical conditions, is to inform us about our current emotional states. On a teleosemantic view, this implies that they have the function of representing emotions, not themes.

This raises an interesting question: why would bodily perceptions represent emotions? There are two main alternatives that we should consider: first, the peripheral states may partially or entirely constitute the emotions, so by perceiving them we are literally perceiving our emotions; and second, peripheral states may be distinct from emotions, but reliably caused by them. This second possibility is consistent with a conventional cognitivist or appraisal theory,

since on these views it is a contingent feature of emotions that they frequently produce peripheral states.

1.3.3 Valence

My second critique of Prinz's perceptual view concerns his account of valence. On his view, emotions are not just perceptions of themes, they are valenced perceptions. The valence basically tells us "more of this," or "less of this" (p. 174). Fear typically has a negative valence, so it not only represents that we are in danger, it motivates us to escape from that danger.

He does not frame valence in terms of desire, but he does characterize it in terms of imperative content, and desires are the type of mental state which corresponds to imperative speech acts. Positive valence could be explained by the presence of a positive desire, and negative valence by an aversion or negative desire. Positive emotions would be complexes of two states: a perception of a theme and a desire that this theme persist, while negative emotions would consist of a perception of a theme and a desire that it end. This construes valence as extrinsic to the perception, which might strike some theorists as a disadvantage, but it is not critical, for present purposes, to determine whether valence is intrinsic or extrinsic to the state.

Since Prinz argues that emotions are bodily perceptions, and gives an account of the valence of emotions, it makes sense to interpret this as an account of the valence of those bodily perceptions. After all, bodily perceptions are a type of state that we, independent of any theory of emotions, take to have valence. However, this interpretation leads to a number of problems. First, Prinz's theory requires that the valence of bodily perceptions in emotional cases motivates us to prolong or end themes. However, in non-emotional cases it seems to motivate us to prolong or

end the peripheral states that the perceptions represent. When we perceive that our skin is getting burned, the negative quality of that perception motivates us to prevent further damage, and when we perceive the pressure of a skillful massage, the positive valence of that perception motivates us to prolong that pressure. Our initial presumption should be that the valence of bodily perceptions should play a similar role in both emotional and non-emotional cases, motivating us to prolong or end their objects.

Prinz seems to agree, since he argues that bodily perceptions represent themes, and that the valence of those perceptions motivate us to prolong or end those themes. However, as I argued in the previous section, bodily perceptions represent emotions rather than themes. It follows that the valence of those perceptions should motivate us to prolong or end those emotions, rather than themes. This seems to be consistent with our experience. We experience our emotions via bodily perception, and when those perceptions feel bad we do not want the emotion to continue. It is bad to be in danger, but we also dislike feeling the tightness in our chest that comes with fear. The direct effect of the valence of bodily perceptions is to motivate us to either prolong or end our current emotion. The valence of the emotion, however, should influence our response to the theme itself.

Prinz could argue that these perceptions motivate us to prolong/end themes as well as emotional states. For an acrophobe, the suffering experienced during an episode of intense fear is reason enough to avoid rock climbing, so the negative valence of the fear does motivate the avoidance of danger under certain conditions. However, in such cases we avoid danger as a means of avoiding fear. This is consistent with the view that the valence of bodily perceptions directly motivates us to prolong or end emotions rather than themes. We also need to explain

cases of so-called ‘self-medication’, where people drink or take drugs in order to escape from negative emotions. One natural interpretation is that the negative valence of bodily perceptions motivates us to escape from the emotions that we perceive. We can do that by preventing the types of circumstances that normally elicit or maintain those emotions, but another strategy is to chemically alter the peripheral state.

In conclusion: if the valence of emotions motivates us to prolong or end themes rather than emotions, then this valence must be distinct from the valence of bodily perceptions, which motivate us to prolong or end emotions themselves. And if the valence of emotions is not the valence of bodily perceptions, that undermines the view that emotions are identical to bodily perceptions. In fact, saying that emotional valence motivates us to prolong or end themes misses some important nuances. This view correctly entails that anger motivates us to avoid offenses and fear to avoid dangers. But it misses how anger makes us less risk-averse in general, and fear more risk-averse (Lerner and Keltner 2001). These emotions do not only influence our motivations with regard to their corresponding themes. It is consistent with Prinz’s view that disgust, a negatively valenced state, would make our moral judgments more severe (Schnall et al. 2008), since immoral behavior can be perceived as source of social contamination. However, his view would not predict that sadness would lessen the severity of moral judgments. We need a model of how emotions influence our motivations that can account for all this data.

There is a considerable empirical literature on the motivational power of different types of emotions. I have mentioned that anger motivates us take risk and fear to avoid them, disgust to judge harshly and sadness to judge laxly. Fear has been shown to strengthen sexual attraction (Dutton and Aron 1974), and there is also plenty of evidence that positive emotions contribute to

helping behavior (Isen 1970). This research most likely only scratches the surface. Each emotion type seems to have its own motivational profile. I will call these effects *generic impulses*, since each type of emotion drives us in a different way, but different tokens of the same emotion type drive us in the same way.

It is not entirely clear how generic impulses operate. One model, which I currently favor, is that each type of emotion temporarily strengthens or weakens certain classes of desires. Perhaps one difference between fear and anger is that fear temporarily strengthens any desires concerned with avoiding harm, while anger weakens them. That is only a partial characterization. Anger also seems to strengthen the desire to act aggressively, and fear, as mentioned above, strengthens sexual desires. These changes in the strength of desires are proportionate to the strength of the emotion, and last as long as the emotion lasts.

My characterization of generic impulses requires two clarifications. First, the same generic impulse can have radically different effects on our behavior, depending on external and internal circumstances. Nico Frijda (1986), among the most influential emotion researchers, says that emotions produce distinctive action tendencies. This view has trouble accounting for the same type of emotion having divergent behavioral effects. For example, fear might in one case drive me to clam up in conversation with a dominating person, and in another it might drive me to run out of a burning building. Identifying action tendencies with each emotion type seems hopeless. However, this problem does not arise for the generic impulse view. Fear strengthens our desire to avoid risk. A natural way to avoid risk in conversation is through silence, while in a burning building the most salient strategy is to flee. The common thread between these divergent

behaviors is risk aversion, so the generic impulse view can explain the contribution that fear makes in both cases.

Second, it is important to emphasize that we sometimes resist these generic impulses. Resisting disgust, we clean up after a sick child; resisting anger, we de-escalate a confrontation. Just because a certain desire is strengthened does not mean that it will ultimately drive behavior in a particular case. But even in cases where we resist these generic impulses, they often make their presence felt. An angry person might avoid overtly hostile behavior, but make a passive-aggressive comment. The strain of suppressing a generic impulse may produce fatigue or frustration. So neither the heteronomy of emotional responses, or the possibility of our not producing an overt emotional response, are incompatible with the generic impulse view.

Emotions do more than just produce generic impulses. I will attempt to give a fuller account of their role later in the paper. For present purposes, however, it is important that producing generic impulses is a conspicuous and important role that emotions play, and we should expect a theory of emotions to explain, or at least have the potential to explain, how emotions play that role.

Can Prinz's view explain why emotions produce generic impulses? If emotions are bodily perceptions, then they can influence our motivations in two ways: through their content and through their valence. I have already argued that the valence of bodily perceptions motivates us to prolong or end emotional states rather than themes. But even if Prinz was right, that would not explain generic motivations. Why should a motivation to end offenses (the negative valence of anger) make us generally more risk-taking, or a motivation to end danger (the negative valence

of fear) strengthen sexual attraction? Valence, as Prinz characterizes it, is unsuited to playing that role.

The content of bodily perceptions is not any more helpful. I have already argued that bodily perceptions represent emotions, rather than themes, but the problem is deeper than that. Perceptions have an assertoric mental attitude - they have the function of mentally asserting that something is the case. Generic impulses, I propose, alter the strengths of classes of desires. We need an explanation for why mental asserting that I am angry, or even mentally asserting that I have been wronged, would systematically make me more willing to tolerate risks. Prinz's view seems to lack the resources for explaining this critical feature.

1.3.4 Emotions as Peripheral States

After considering the challenges that Prinz faces in explaining Lange's Premise, it is only fair to consider Lange's own theory. Though Lange's view does not prompt the exact same objections, I will argue that it does not ultimately fare any better. I will address two objections to his view, both of which are based on the fact that bodily perceptions mediate the effects of peripheral states on the central states.⁵

Both James and Lange agree that peripheral states produce bodily perceptions, which in turn produce the motivational and behavioral features that are characteristic of different types of emotions. This is a direct result of accepting Lange's Premise. Unlike James, Lange identifies the emotion with the peripheral state. Since the effects of peripheral states on motivation are

⁵ I will refer to brain states as 'central states'. This highlights the contrast between brain states, which are governed by the central nervous system, and peripheral states, which are governed by the peripheral nervous system.

mediated by bodily perceptions, we need a reason to assign credit for the motivational effect to the peripheral state. This issue becomes more acute when there is a disconnect between the peripheral state and the bodily perception which mediates its effects on central states. If, for example, we directly manipulate the bodily perception, those perceptions would produce the same generic impulses as they would if they were produced by peripheral states. The result of that manipulation produces characteristic emotional effects, so why should we deny that it is a genuine emotion?

According to Lange's view, bodily perceptions represent emotions. If you manipulate a bodily perception to produce the state which normally occurs during fear, it will on the peripheral view represent the peripheral state that constitutes fear. Since in this case the individual is not actually in that peripheral state, Lange's view would characterize the bodily perception as a hallucination. Specifically, the individual hallucinates being in a state of fear. However, the individual manifests the generic impulses characteristic of fear. Just as a cognitivist needs to explain why the results of peripheral manipulation are not real emotions, a defender of the peripheral view needs to explain why manipulation of bodily perceptions does not produce real emotions.

The second objection is also based on the fact that perceptions mediate the effects of peripheral states. If, as I have argued, bodily perceptions have the wrong sort of content to cause generic impulses, how can peripheral states whose effects are mediated by those same bodily perceptions do any better? It seems as though my critiques of Prinz's view are at least as harmful to Lange's.

I argued in the previous section that the content and valence of bodily perceptions is not suited to producing generic impulses. The content of a bodily perception is that the individual is in a certain type of emotional state, and the valence of that perception motivates the individual to prolong or end that emotion. During emotional episodes we are often aware of our emotions, and we do often have preferences regarding whether those emotions persist, but those awarenesses and preferences do not explain why emotions produce generic impulses.

This argument is damaging to Prinz's view for obvious reasons, but it actually does as much damage to Lange's view. Peripheral states cannot directly influence brain states such as desires. If they produce generic impulses, that production must be mediated by central states which receive peripheral input. Bodily perceptions are the obvious candidates to play that role. If, as I have argued, bodily perceptions are not suited to producing generic impulses, then peripheral states seem to lack a suitable mediator, and may therefore not be qualified themselves.

There are two approaches to solving this problem. First, one could argue that some other kind of state mediates peripheral input, and unlike bodily perception it is suited to producing generic impulses. Making this work would require completing a series of steps: 1) distinguishing between bodily perceptions and non-perceptual central states which receive peripheral input, 2) arguing that there are non-perceptual central states which receive input from the relevant peripheral states, and 3) arguing that these states are better suited than bodily perceptions to producing generic impulses.

It may be possible to accomplish the first two steps, but the third looks like a serious obstacle. Those central states would only be relevantly different from perceptions if they did not represent peripheral states, but given that they receive input from the periphery, it is not clear

what else they would do. One could argue that they have the teleosemantic profile that Prinz attributes to bodily perceptions, and therefore represent themes. However, this strategy raises its own set of problems. First, we would need evidence that these states do play this teleofunctional role, and second, I have already argued that a state which represents themes would not be suited to producing generic impulses, so this would not provide any clear advantage.

The other general approach to dealing with this problem is to argue that bodily perceptions are, contrary to the argument above, well-suited to facilitating generic impulses. This requires some sophisticated triangulation. In order to defend Lange's view one must show that bodily perceptions can facilitate the production of generic impulses, but that simultaneously undermines the primary argument against Prinz's view. In summary, both Prinz's and Lange's views are in trouble unless bodily perceptions can generate generic motivations, and if bodily perceptions can play that role, the arguments from the previous section seem to favor Prinz's view over Lange's. Lange's view seems to lose either way.

1.4 The Inter-systemic View

The prospects are bleak for Lange's view as he presented it. However, in this section I will present a variant, which I will call the inter-systemic view, that is not vulnerable to the objections that I have raised. This view inherits from Lange the claim that peripheral states are emotional constituents, but unlike the original, the inter-systemic view includes central states as constituents as well. The integration of states of the central and peripheral nervous systems (hence: 'inter-systemic') is critical for explaining Lange's Premise.

1.4.1 Presenting the View

In order to make sense of the inter-systemic view it is necessary to explain the respective roles of the central and peripheral constituents, as well as how they relate to each other.

According to my proposal, the peripheral constituents have two direct responsibilities: preparing the body for certain types of actions, and signalling the current emotional state to conspecifics.

The central constituents are directly responsible for producing generic impulses, and perhaps other psychological effects as well. All of the central and peripheral constituents are additionally responsible for maintaining functional unity. If the peripheral constituents are currently suited for aggressive action and signalling preparedness to act aggressively, they need to make sure that the central constituents are producing a generic impulse which favors aggressive action as well.

Likewise, when the central constituents produce a generic impulse which favors aggressive action, they need to make sure that there are peripheral constituents which will facilitate such action and signal an aggressive disposition.

I have said little about physical capabilities or signalling, since those functions have not figured in the arguments for the perceptual or peripheral views. Defenders of cognitivist and appraisal theories can happily accept that peripheral states change our physical capabilities and signal our emotional status. They are less happy to accept that peripheral changes are responsible for motivational changes, since that is a role that these theorists reserve for purely central processes. However, a motivational change without a corresponding change in physical

capabilities is potentially disastrous, and emotion theorists have recently emphasized the role of emotions in interpersonal negotiation, which requires accurate signalling.⁶

In addition, the three emotional functions that I have mentioned reinforce each other. Returning to the example above, when you truly are motivated to act aggressively, and currently have the physical capabilities to act aggressively, it is useful to make this status known. However, it is dangerous to send that signal if you currently lack the capabilities or motivation required to act accordingly, so features that figure in signaling must stay in sync with features that figure in motivation and the modulation of physical capabilities.

Setting emotions aside, armchair physiology suggests that effective functioning of the organism requires inter-systemic integration, and the empirical literatures confirms that we do in fact have this beneficial feature. The controversial move, then, is not proposing that we have inter-systemic states, but identifying these inter-systemic states with emotions. I will argue that this identification is justified because it solves a number of problems in the emotions literature.

1.4.2 Reply to Objections

So far I have explained the inter-systemic view by appealing to the need for central and peripheral states to be coordinated. I'll now argue that this need for coordination is the key to addressing the objections that Lange's own view faces.

I'll begin with the second objection. In response to Prinz, I argued that bodily perceptions are not suited to producing generic impulses, since they represent emotions, and it is not clear why representing the presence of an emotion would produce generic impulses. Since the effects

⁶ See Griffiths and Scarantino 2009 and Morris and Keltner 2000.

of peripheral states on central states such as generic impulses are mediated by bodily perceptions, that means that peripheral states cannot produce generic impulses either.

This problem is solved by conceiving of emotions as constituted by multiple coordinated components. At any time, each component needs to make sure that the others are acting in harmony. We might expect that the central elements would trump, so if central and peripheral constituents fell out of sync for any reason the peripheral elements would need to fall in line. However, this assumption is belied by the anecdotal and empirical evidence that peripheral manipulations produce central changes. It turns out that if bodily perception represents that the peripheral components are currently in a state that normally (partially) constitutes happiness, the central elements will follow along and generate a complementary generic impulse. On the inter-systemic view, a representation with the content that the periphery is currently (partially) constituting a given emotion is exactly the right sort of state to produce a generic impulse.

The solution to the first objection is equally simple. If the question is whether to identify emotions solely with bodily perceptions or peripheral states, bodily perceptions have a built-in advantage. Emotions must be constituted by states that produce generic impulses, and when it comes to producing generic impulses there is a distinct asymmetry: bodily perceptions can produce them without the help of peripheral states, but peripheral states cannot produce them without the help of bodily perceptions. If we start with the assumption that one of those two types of states is primarily responsible for generic impulses, it is very difficult to assign credit to the peripheral state.

The story becomes very different if emotions are constituted by both central and peripheral states. We can accommodate the presupposition that emotions are constituted by states

responsible for generic impulses by including the immediate central causes of generic impulses among the constituents. They are also constituted by the peripheral states responsible for signalling and mobilizing physical capabilities. Peripheral states are also responsible for generic impulses. However, that is a consequence of the need for the central and peripheral elements to operate in harmony.

What should we say if bodily perceptions are directly manipulated, producing a generic impulse that is out of step with the peripheral state? Perhaps that bodily perception represents the peripheral state which normally occurs during fear, leading the individual to become more risk averse (among other psychological changes). At the same time, the individual is in a peripheral state which normally occurs during happiness. Is that individual happy, as the peripheral view predicts, or afraid, as the perceptual view predicts?

According to the Neo-Langian view, neither. Unless there is further interference the central and peripheral constituents will eventually harmonize, settling into some coherent emotional state. But for the moment, the individual is in a mixture of happiness and fear. This is a different sort of mixture than is said to occur when we enter into two compatible emotions, such as anger and disgust. In those cases, often called emotion blends, the central and peripheral components each feature properties that are characteristic of each emotion. Here, the emotions are distinctly unblended. It is a strange situation, but we should expect strangeness given the setup. The important thing is that the inter-systemic view has a coherent and plausible explanation for why the peripheral state and not the bodily perception is a constituent of the emotion. The peripheral state performs some of the core emotional functions, and the perceptual state is just utilized to harmonize components of the emotion.

1.4.3 Central Constituents

I characterized the direct function of the central emotional constituents as the production of generic impulses. With that function in mind, I identified the central constituents as whatever central state is immediately responsible for producing generic impulses.⁷ Further work is required to characterize those states in a more satisfying manner.

I'm currently uncommitted between two alternative views, which are influenced by the debate over moods. The central elements of inter-systemic states are very similar to what most emotion theorists would identify as moods: they are central states that produce characteristically emotional motivation effects, but do so in a diffuse manner. This diffuseness is puzzling for theorist who consider emotionality to require direction towards an intentional object, and there are a number of proposed explanations in the literature. One strategy is to maintain that they do have objects that are either general in scope (Solomon 1976) or specific but obscure to the subject (Crane 1998).

The other strategy, adopted by Griffiths (1997) and Sizer (2000), is to deny that moods have intentional objects.

If I adopt a variant of the first strategy, I would say that the central constituents of emotions are either a special class of desire or aversion, or have a mental attitude that is similar to that of desire or aversion. Fear is, or resembles, an aversion to harm in general, sadness is, or resembles, an aversion to loss. To understand this proposal, imagine that Ben sees a spider. This

⁷ In chapter 3 I will argue that emotions do not motivate us in a manner that is specific to any particular object. For this reason, generic impulses are the only motivation influence that I consider here.

produces an aversion to being harmed by the spider, which in turn produces an aversion to harm in general. This general aversion then initiates a process to produce a certain sort of peripheral state. This process may not be successful, since other desires or aversions may initiate processes to produce other sorts of peripheral states, and the periphery may also be influenced by other factors entirely.

One unsatisfactory aspect of this proposal is that it is not as effective at explaining the episodic aspect of emotions, and relatedly, the results of peripheral manipulation. It is a familiar observation that disgust can easily coexist with anger, but not with joy. When we are in a sort of emotional state, that emotion has a comprehensive, though not entirely dominating, influence on our psychology for as long as it lasts. Disgust can coexist with anger but not with joy because the comprehensive, psychological profile of disgust is much closer to anger than joy. Relatedly, when disgust is elicited by peripheral manipulation, the resulting generic impulse will be relatively compatible with the pre-existing generic impulse from a prior state of anger, but incompatible with any generic impulse from a prior state of joy. The mind is not big enough for both disgust and joy. However, it is not clear why this would be true on this proposal, since desires and aversions do not seem to feature this sort of comprehensiveness.

One way to account for this is to deny that the central constituents have intentional content, adapting arguments from those that Sizer develops in reference to moods. She follows Pylyshyn (1980) in distinguishing states at the functional level of organization from states at the representational level of organization. Sizer says that,

Changes in operations at this [the functional] level have global effects on representation level states and processes but in a way that is independent from the semantic contents of those states. The changes are bottom-up and therefore influence the creation and

performance of all affected representations and representational processes. This is what explains the generalized, content independent effects of moods (p. 763).

While Sizer identifies moods with functional level states, my second proposal is to do the same with the central constituents of emotions. The central constituents produce generic impulses by temporarily altering the parameters of classes of desires. As an additional benefit, this strategy may be helpful for explaining the comprehensiveness of emotional states. The central constituents, on this view, are not just another representation within the representational system, but a state of the system that moderates all representations within that system. Incompatible emotions may feature (as partial constituents) incompatible states of that system.

Both of my proposals for the identity of the central constituents would benefit from further development. My goal in exploring these options is to indicate that there are no obvious obstacles for fully developing the inter-systemic theory.

1.5 Conclusion

There is both anecdotal and empirical support for the notion that peripheral manipulation can alter emotional experience in a manner that has genuine emotional consequences. Whenever this phenomenon has drawn the attention of theorists in the philosophy of emotions, it has been used to support theorists in the James-Lange tradition against whichever class of theories currently prevail. My arguments show that this phenomenon is also a powerful tool for settling disputes between theories within this tradition. Bodily perceptions do have a critical emotional role: to help integrate the central and peripheral constituents of the emotion.

Chapter 2: Inter-systemic States

2.1 Introduction

In the previous chapter I noted that none of the existing theories of emotion elegantly explain both of the following features: that emotions can be elicited or altered by manipulating the bodily periphery, and that emotions are responsible for generic impulses. The inter-systemic view was developed specifically to fill that gap. However, there are plenty of other features that a theory of emotions should explain. In this chapter I'll begin to address some of the most pressing concerns that the inter-systemic theory raises.

The first major class of concerns is based on the unconventional constitution of inter-systemic states. The inter-systemic view identifies emotions with integrated central and peripheral states, and it may seem as though these elements do not collectively constitute any sort of natural kind, or that the inter-systemic view identifies the wrong constituents. I will argue that inter-systemic states are natural, that they include the right constituents, and that even skeptics of embodied cognition in general should tolerate embodiment in the case of emotions.

The second class of concerns centers on a pair of features, intentionality and consciousness, that are each considered by some to be the mark of the mental. If emotions are mental, as the folk and philosophers alike commonly take them to be, and if one of these features does mark the mental, then we should expect emotions to have that feature. Additionally, mentality aside, it is widely accepted that emotions are intentional states and are frequently if not

always conscious. It is not clear that inter-systemic states have either of these features, since they include peripheral states as constituents. I will argue that inter-systemic states probably do lack these features, but I will deny that this gives us reason for rejecting the view.

2.2 The Constitution of Emotions

In this section of the chapter I will address four related concerns. First, in order for emotions to be identical to inter-systemic states, inter-systemic states need to be natural kinds. For those who are dubious about the special status of natural kinds, I will argue that inter-systemic states deserve inclusion in scientific theories. Second, one may accept most of the account I have presented but identify the emotion with only the central constituent of the inter-systemic state. I will argue that there is no basis for privileging the central constituent in this manner. Third, the inter-systemic theory resembles a number of hybrid theories in philosophy and psychology that include peripheral states as constituents. I will explain why the inter-systemic theory is stronger than others in this class. Finally, I will discuss how the inter-systemic theory relates to embodied cognition as a general research program.

2.2.1 Natural Kinds

There are many competing views about the nature of natural kinds. In this section I will only argue that emotions, as I describe them, would count as natural according to Richard Boyd's (1991) theory of natural kinds as homeostatic property clusters. On his view, a set of properties

collectively make up a natural kind if the instantiation of some of the properties increases the likelihood that each of the others will also be instantiated.

Inter-systemic states clearly count as natural kinds on Boyd's view. There are recurring types of situations that produce predictable physiological, communicatory, and psychological demands. In order to effectively address these situations, our central and peripheral nervous systems need to respond in a coordinated fashion. Otherwise, our current physical capabilities will not be suited to the actions that we are currently motivated to make, and we will not be motivated to act in the manner that our current physical capabilities would most effectively facilitate. Emotions align our physiology and motivation in preparation for coping with different kinds of challenges and opportunities, while signifying our emotional status to others who might help or thwart us.

Theorists who do not accept the metaphysical commitments implicit in theories of natural kinds may have a related concern about inter-systemic states. We have good reason for denying that states of toe-nails partially constitute our perceptions. In the case of emotions, I have argued that some, and only some, peripheral states are partial emotional constituents. The central and peripheral constituents are unified by (A) solving a common problem, and (B) feedback loops between the various constituents, due to the need for coordination between the constituents in solving that problem. That functional and causal unity justifies treating the elements as jointly constituting one thing. The rest of my arguments aim to show that it is the right thing to identify with emotions.

I have just argued that inter-systemic states are natural kinds, but there are actually two prominent arguments in the literature for denying that emotions form a natural kind. First, Lisa

Feldman Barrett (2006) denies that there is empirical evidence for distinct emotional states along folk-psychological lines. Instead, she argues along with James Russell (2003) that emotion can be decomposed into core affect, which varies in dimensions of arousal and valence, and conceptual thought that analyzes this affect using folk-psychological categories.

I will make two points in reply to Barrett's argument. First, there is significant evidence from physiology that I presented in chapter 1, and social psychology that I will present in chapter 3, that arousal and valence do not exhaust intrinsic emotional variation. Second, throughout this dissertation the emotions in my examples are identified by folk psychological kinds, and specifically tend to be drawn from the relatively small class of basic emotions that some theorists consider to be innate and universal.⁸ However, the inter-systemic view is not committed to preserving the folk-psychological taxonomy of emotional variation. The basic elements of the theory could accommodate a revisionary class of basic emotions, or even a more complex version of Barrett and Russell's dimensional taxonomy.

Unlike Barrett, Griffiths (1997, 2004) accepts that the basic emotions form a natural kind. However, he argues that there are other emotions such as guilt, shame, and jealousy, that do not belong in that class, and may not even form a uniform class with each other. He argues that they all diverge from the basic emotions by essentially involving some form of advanced cognition.⁹ I suspect that in these cases the cognitive processes do not actually play an essential role. They are involved in eliciting the emotion, and we use the conditions of elicitation to label the emotion, but the emotion itself may be no different than an emotion elicited in another manner. However,

⁸ See Ekman and Friesen 1971.

⁹ Psychologists have also explored these phenomena. See Tangney et al 1995.

for the purpose of this dissertation I will remain neutral about that question, and focus only on cases of what are commonly called basic emotions.

To summarize the conclusions from this section: I have argued that inter-systemic states are a natural kind, and I propose to identify them with the basic emotions, which I also consider to be a natural kind.

2.2.2 A Non-embodied Variant

One might agree with the general thrust of the inter-systemic view, but interpret it in a non-embodied manner. According to this variant, the central element alone is the emotion, and the peripheral elements just play a supporting role. However, there is no empirical reason for giving priority to the central over the peripheral elements. Peripheral elements are elicited and influenced by central elements, but central elements are also elicited and influenced by peripheral elements. They seem to operate as peers, with a shared purpose, and that shared purpose is the purpose that we would identify as the purpose of the emotion: to effectively cope with certain kinds of challenges and opportunities. Effective coping involves coordinated activity in the center and periphery, and we should identify emotions as that coordinated activity.

One argument for the non-embodied variant is based on brain-in-vat cases. On the inter-systemic view, the brain includes only part of the emotion, so an envatted brain would have only partial emotions. This may seem intuitively false, since the experience of an envatted brain is indistinguishable from our own. I could claim that the envatted individual would in fact have true emotions by arguing, following Chalmers (2005), that in such cases the emotion would be partially constituted by elements of the simulation, or perhaps by part of the machinery that

implements the simulation. However, I will not commit to any interpretation of such cases, because I deny that we should lend any credence to the intuition that envatted individuals really would have emotions. On the inter-systemic view, the envatted individual would, given the right setup, have states that are epistemically indistinguishable from emotions. However, the best way to interpret that type of scenario (if we must) is to start with the best available theory of emotions, and considering brains in vats will not get us there.

2.2.3 Evaluative Judgments

The inter-systemic theory bears, at first glance, a resemblance to a number of hybrid theories in philosophy and psychology that include peripheral states as constituents.¹⁰ It is worth considering whether it significantly departs from them, and if so, whether it is superior. This comparison would be a very lengthy process if there were not a simple feature that distinguishes the inter-systemic theory from the recent popular hybrids: while they all include peripheral states as constituents, the inter-systemic theory is the only one that does not include evaluative judgements.

The case for including evaluative judgments is simple. In prototypical cases, an evaluative judgment is responsible for producing the central and peripheral states that the inter-systemic view identifies with the emotion. Will sees a charging bear, evaluates the bear as an acute danger, and the rest follows from there. Further, once Will becomes afraid, his fear is (we would ordinarily say) fear of the bear. The constituents posited by the inter-systemic theory are

¹⁰ See Scherer 2009 for a prominent example from psychology, and Blackman 2013 for a discussion of relevant philosophical theories.

not targeted: the motivational constituent is explicitly generic, and the physiological changes and social signals do not distinguish between specific dangers. The only way to explain why Will's fear is fear of the bear is to include the initial evaluative judgment.

I will not address the issue of intentionality here, since I will do so later in this chapter as well as in the next. However, I will provide a pair of arguments against including evaluative judgments. First of all, a theory of emotions needs to explain the emotional effects of manipulating the bodily periphery. The hybrid theories that include evaluative judgments are faced with a dilemma: either the peripheral manipulation produces an evaluative judgment (how?), or the resulting state is not a true emotion (why not?). That problem does not arise if evaluative judgments are frequently the immediate cause of emotions, but not counted as constituents.

Second, I justified the inter-systemic unity by noting that the constituents are both functionally and causally unified. For evaluative judgments to merit inclusion, peripheral manipulation would need to influence them in a targeted manner. For example, imagine that Alice is confronted by two obnoxious individuals, Charles and John. Charles says something obnoxious that provokes Alice to form a negative evaluative judgement, which subsequently elicits a state of anger. This state makes her temporarily less tolerant of either individual, since angry people have less patience for obnoxious behavior.¹¹ In order to avoid making a scene Alice then takes a series of deep breaths.

¹¹ I will argue in chapter 3 that the emotion influences her responses in a manner which does not discriminate between two seemingly obnoxious individuals, but for the purposes of this discussion we can remain neutral on that topic. It is uncontroversial to note that the emotion would have at least some effect on her response to any obnoxious behavior.

If the original evaluative judgment about Charles belongs in the emotional unity, then this peripheral manipulation should increase her tolerance in a manner that is specific to Charles.

While I would be interested in empirical evidence that bears on that question, I find that extremely implausible. Instead, it is much more likely that the peripheral manipulation makes her more tolerant of any obnoxious behavior, whatever the source. In other words, it directly influences generic impulses, which then influence any relevant evaluative judgments. The particular judgment that elicited the emotion has no special status, so it does not belong in the homeostatic cluster.

2.2.3 Embodied Cognition

Since the inter-systemic theory includes peripheral states as emotional constituents, it may seem to require a firm commitment to embodied cognition. Supporters of an embodied approach to the cognitive sciences in general often lean on evidence for emotional embodiment of the sort that I discuss in chapter 1.¹² By arguing for an embodied view I may seem to lend support to their camp, for better or worse.

In fact, there is no reason why a skeptic about embodied cognition in general needs to oppose the inter-systemic view.¹³ I argued that peripheral states are emotional constituents because the function of emotions is to ensure that the central and peripheral nervous systems operate in harmony to support the task at hand, whatever that might be. Even if I am mistaken in

¹² See e.g. Niedenthal et al. 2005.

¹³ See Goldman and de Vignemont (2009) for a largely skeptical take on embodied cognition in the social domain.

identifying emotions with those inter-systemic states, that sort of integration clearly must exist for organisms such as us to function effectively. Those inter-systemic states, whatever we call them, are indisputably embodied. Since any reasonable theorist is implicitly committed to embodiment of this sort, identifying emotions with these inter-systemic states requires no additional concession. If the inter-systemic view is correct then emotion are embodied, but the arguments I have made for emotional embodiment do not apply automatically to any other class of mental states.

2.3 Marks of the Mental

In the previous section I addressed the concern that inter-systemic states are not mental because they are not states at all. I will now take for granted that inter-systemic states are natural kinds, or at least kinds that are of scientific interest, and that I have properly identified their constituents. In this section I will consider whether there is any reasonable way to characterize the mark of the mental such that it applies to emotions but not to inter-systemic states.

2.3.1 Intentionality

The most famous conception of the mental in contemporary philosophy is Brentano's (1995) view that intentionality is the mark of the mental. To be a mental state is to have an intentional object — for that state to be about something. Accepting Brentano's view for the sake of the argument, we can now state an objection:

P1i: emotions are intentional states

P2i: inter-systemic states are not intentional states

∴: emotions are not inter-systemic states

Given this argument, the question becomes whether emotions really are intentional, and whether inter-systemic states really are not.

However, it is possible to refine this argument further. Following Anthony Kenny (1963), it is common to distinguish between the formal and the particular intentional objects of an emotion. If I am disgusted by tasting some mayonnaise, then the mayonnaise itself is the particular object. The formal object determines what can possibly be a particular object of a certain type of emotion. For disgust, the formal object might be contamination, in which case I cannot be disgusted by mayonnaise unless I construe it as a contaminant. If Brentano is right, then emotions are not mental unless they have objects, and if Kenny is right, an adequate theory of emotions needs to explain both formal and particular objects. With Kenny's distinction in mind, we now have the following argument:

P1i': emotions have particular and formal objects

P2i': inter-systemic states do not have particular and formal objects

∴: emotions are not inter-systemic states

In chapter 3 I argue that emotions do not have particular objects, because once elicited, emotions will influence judgment in a manner that does not discriminate between objects of the same type.¹⁴ I may enter a state of disgust as a result of inadvertently eating a sandwich with mayo on it, but once I am disgusted, this state will make me more sensitive to any apparent

¹⁴ My arguments in chapter 3 do not presuppose the inter-systemic view, so this approach is not circular.

source of contamination. Having ruled out emotions having particular objects, I leave open two possibilities: that emotions are directed at the world in general, or that emotions are not intentional at all.

It is difficult to evaluate whether inter-systemic states are intentional, since their constituents are so dissimilar. Anger might be constituted by features such as flared nostrils, elevated blood pressure and heart rate, diminished aversion to risk, and elevated aversion to goal-obstruction. The altered blood pressure and heart rate do not, in isolation, seem to be intentional. The flared nostrils (and other visible features) could be construed as intentional since they serve a communicative role, with content like ‘I am angry’. I presented two accounts of the central constituents in chapter 1, and in one but not the other they are intentional states representing general features of the environment.

Given such diverse constituents, it is not clear how to characterize the intentionality of the states that they constitute. Blackman (2013, p. 85) gives an example of a statue in honor of fallen soldiers built out of biographies of Andy Warhol. This nicely illustrates that even when an object’s constituents have a uniform intentional object, the whole may be about something entirely different. Blackman himself defends a hybrid account of emotions in which they include evaluative representations, and argues that the content of the whole is identical to the content of that evaluative representation because, “one cannot have a change in the intentional object of the emotion without a change in the intentional object of the evaluative representation” (p. 86). However, I deny that emotions have particular objects, so I also deny that evaluative representations determine their content.

The best method for analyzing the intentionality of composites such as inter-systemic states may be to consider the basis of their unity. The inter-systemic view unifies emotional constituents on the basis of their shared task: effective coping with a type of situation. In that respect, inter-systemic states are analogous to the integrated immune responses that occur in response to an infection, or the various organs involved in digestion. The responses are integrated by the need to solve a shared problem, but that does not mean that the whole is intentionally directed at that problem. The inter-systemic case may differ if the central constituent is itself intentional, but there is no reason to think that the constituents are unified in a manner that features representation.

In conclusion, I have argued that inter-systemic states are not intentional, and that there is reason, inter-systemic view aside, to doubt that emotions are intentional either.

2.3.2 Consciousness

Another view of the mental, typically associated with Descartes, is that mental states are necessarily conscious. This sort of strong Cartesianism is not very popular today, but John Searle's (1992) weaker view, that mental states must be at least potentially conscious, is more palatable. Perhaps when we consider it obvious that emotions are mental, that impression is driven by the view that emotions are potentially conscious states. We can state this argument in the following way:

P1c: emotions are potentially conscious

P2c: inter-systemic states are not potentially conscious

∴ emotions are not inter-systemic states

As was the case in the previous section, it is necessary to address the two premises individually.

In order to evaluate whether emotions themselves are potentially conscious, it is useful to first describe the distinctive phenomenological features of emotional episodes, since it is not obvious that emotions themselves are responsible for all of these features. James (1922) and Lange (1922) argue that conscious bodily perception is the only phenomenological feature that distinguishes emotional episodes from episodes of dispassionate intellectual activity. Peter Goldie (2002, p. 241) argues that emotional experience also includes what he called ‘feeling toward,’ which he describes as “unreflective emotional engagement with the world beyond the body; it is not a consciousness of oneself, either of one’s bodily condition or of oneself *as* experiencing an emotion.” Uriah Kriegel (2011) does not make a commitment about the exact constituents of emotional phenomenology, but he additionally considers cognitive, conative and some sort of *sui generis* affective phenomenology as potential candidates.

There is consensus that the phenomenology of emotional episodes typically includes conscious bodily perceptions. It is not entirely clear what Goldie’s feeling towards is supposed to be, but during emotional episodes we do typically have some motivational state directed at what we take to be the object of the emotion. While I do not accept that there is cognitive or conative phenomenology, I also accept that there are often conscious thoughts and desires during emotional episodes. The only element that I reject altogether from this list is *sui generis* emotional phenomenology, since the other features mentioned seem exhaustive.

Having listed all of these sources of phenomenology, we can turn to the question of which features, if any, are due to a conscious emotion. Unfortunately, methodological hurdles

immediately appear. First of all, it is not clear that we have any reliable method of subtracting the bodily perceptions from experiences during emotional episodes, as James asks us to do. At the same time, it is not obvious that Goldie's feeling toward is distinctively emotional. In chapters 3 and 4 I discuss ways in which it is difficult to distinguish between emotions and desires, and that problem arises here as well. An aggressive barking dog may elicit a state of fear, but it will also certainly elicit a strong desire to avoid the dog. Is my attitude toward the dog due to emotion or desire?

I argue in chapter 3 that the impression that emotions are directed at particular objects does not derive from direct conscious awareness, but is instead an inference that we make on the basis of situational features. However, this claim is not based on direct phenomenological analysis, but on an analysis of various theoretical considerations. Phenomenology alone simply does not tell us what emotions themselves contribute to emotional experience, and what contributions are due to other states that tend to co-occur with emotions. It just is not pre-theoretically clear whether emotions are ever conscious states.

We can now turn to the second premise of the argument, and ask whether inter-systemic states are ever conscious. As was the case in the section on intentionality, the mixed constitution of inter-systemic states makes them difficult to analyze. The peripheral constituents, seen in isolation, do not seem to ever become conscious. They contribute to phenomenology only through being consciously perceived.

The central constituents are more difficult to analyze. They produce generic impulses, so their presence might be felt indirectly through those effects. However, it is not clear whether, additionally, the central states themselves become conscious. I mentioned in the discussion of

intentionality that they may resemble desires, since their functional role includes strengthening and weakening classes of desires and aversions. If so they may potentially become conscious states, though when conscious they would not manifest any of Goldie's feeling toward.

In summary, on the inter-systemic view we feel the peripheral constituents, the central constituents are manifest through influencing our desires and perhaps become conscious themselves, and our awareness of how these changes were elicited leads us to posit that we are in a certain type of emotional state with a certain particular object.¹⁵ However, I see no reason to think that the inter-systemic states, considered as wholes, would thereby become conscious as well. This is not, however, a challenge to the inter-systemic view in the absence of a compelling reason for thinking that emotions themselves become conscious states rather than states that we feel.

2.4 Conclusion

Over the course of this chapter I have discussed a number of peculiar features of inter-systemic states. There is no straightforward method for deciding whether they are intentional, or whether they are potentially conscious. I have argued, however, that this does not rule out the possibility that they are the emotions.

I have not directly stated whether emotions are, on the inter-system view, mental states. I argued in this chapter that inter-systemic belong in our scientific theories, and in the

¹⁵ I discuss the reasons we have for mistakenly positing particular objects in greater depth in chapter 3.

previous chapter I argued that they explain the effects of peripheral manipulation and the production of generic impulses. They are compatible with a satisfactory, though somewhat counterintuitive, account of emotional intentionality and consciousness. Once all that has been said, it is unclear what more is to be gained by talking about mentality itself.

Chapter 3: Emotions without Objects

3.1 Introduction

The near consensus among theorists working on emotion, in both philosophy and psychology, is that emotions have objects. When we are angry, we are angry at someone or about something. When we are frightened, we are frightened of something. We can call this the Token Object Theory (TOKEN), though the defenders of this view are so numerous and diverse that they agree on little else. Adherents of TOKEN rarely treat it as a theoretical position in need of defense, with empirically falsifiable implications.

In fact, some relevant empirical work has been done. I will argue that this research supports an alternative model, the Type Influence Theory (TYPE), according to which emotions lack particular objects. Fear, for example, influences judgments about anything that seems dangerous, without any special status for whatever the subject is “really” afraid of. TYPE proposes that emotions are, with respect to intentionality, no different from moods.¹⁶ The conclusion of the arguments in this chapter will remain neutral about whether emotions are directed at general features of the world, such as danger or offensiveness, or whether they lack direction altogether. They are therefore consistent with the conclusion in the previous chapter

¹⁶ It may still be useful to distinguish emotions from moods based on other features, such as time-course.

that emotions are not in fact intentional states, but I will not discuss the relations between these arguments until the conclusion.

My proposal is so far outside the mainstream that readers familiar with the emotions literature might wonder how it could possibly be right. After all, much of the philosophical discussion of emotions, including the treatment of key features such as normativity and motivation, is based on the assumption that emotions have objects. I cannot hope to produce here a full alternative to the existing emotions literature. However, I will briefly reply to a number of challenges that may initially appear to be fatal. I hope to show that TYPE presents a robust alternative to prevailing views.

Part 2 will introduce a criterion for determining whether emotions have objects. Parts 3 and 4 will discuss in detail the two psychological studies that form the basis for my argument against TOKEN. In the remainder of the paper I will consider a variety of objections: in part 5 I'll reply to objections concerning my interpretation of the studies, in part 6 I'll address our ordinary sense that emotions are about something, and in part 7 I'll discuss some additional objections drawn from the philosophy of emotions.

3.2 The Empirical Case

3.2.1 The Criterion

Before discussing empirical studies it is useful to introduce a criterion that will guide their interpretation. Here is my proposal: if emotions have particular objects, then they will influence experimental participants' behavior in ways that are specific to those objects. On the

contrary, if emotions influence participants in ways that are the same for all objects, or all objects of the same type, then those states do not have particular objects.¹⁷ This criterion is derived from a conventional argument in favor of taking any mental states to have objects (Fodor 1987). On a daily basis we attribute beliefs and desires to our peers. But merely knowing that Jane believes tells us nothing about what Jane will do. We need to know *what* Jane believes. Knowing that Fred desires tells us nothing about what Fred will do. We need to know *what* Fred desires.¹⁸ The same goes for emotions. If they have objects, predicting the influence of an emotion on an individual's behavior should require correctly attributing an object to that emotion.

It is not important for these purposes whether the purported intentional object is a particular, a property of a particular, an event, a state of affairs, or a proposition. If Will is afraid of a bear, and the object of his fear is a particular bear, then Will's fear should influence him in ways that are specific to that bear, rather than other particulars. If the object of his fear is the fierceness of the bear, then it should influence him in ways that pertain specifically to that property of that bear. If, instead, it is more accurate to say that he is afraid that the bear will harm him, where the object of the fear is the proposition 'that the bear will harm him', then the state of fear should influence Will's behavior in ways that depend on the specific content of that proposition. That includes, at a minimum, influencing Will's response to that bear. It is unwieldy to repeatedly address these different variations of TOKEN, so in places I will write as though I

¹⁷ I do not address the vast literature comparing the strengths and weaknesses of various theories of intentionality. However, given the lack of consensus or even significant convergence among theorists of mental content, it would be unhelpful to anchor my discussion to any particular theory. Instead, my strategy is to focus on a datum that any viable theory of intentionality would need to accommodate.

¹⁸ 'Desire' is sometimes used as an emotion term, referring to sexual arousal and other approach-related emotions. However, what is meant here by 'desire' is the type of mental state that we typically report via expressions such as "I want a cup of coffee" or "I want to take a walk."

am targetting one or another. However, my arguments should apply equally to theories that attribute any sort of intentional object.

Of course in our ordinary lives we do often attribute objects to our emotions and to the emotions of others. Typically we take an emotion to be directed at an object when a perception, thought, or imaginative act with that intentional object was responsible for eliciting the emotion. But it is one thing to say that a person made you angry, and another to say that your anger is in any meaningful sense directed at that person. Again, a comparison with belief should clarify this point. Suppose Sam believes that there is a pen on her desk. This belief is about that pen not merely because it was elicited by a perception of the pen, but because it will play a special role in pen-related thought or behavior she may engage in. If the belief influenced her in ways that were perfectly neutral between pen and pencil-related thought or behavior, we would have no basis for taking her to believe that there is a pen, rather than a writing implement or some other more general type of object. The conventional view seems to be that emotions, like beliefs and desires, play a special role in determining how individuals in such states will relate to their intentional object. I will argue that emotions do not have this sort of special role.

This criterion might seem flawed for requiring that emotions manifest their object-directedness in behavior. If an emotion influences other mental states in an a way that is token-specific but not apparent from observing the individual's behavior, that is clearly enough reason to consider the emotion to have an object. Any token-specific effects provide evidence that a mental state is object-directed. However, if emotions do have token-specific effects on other mental states, then at least some of the time this should be observable through subsequent influence on behavior. If we cannot find any evidence that emotions have any token-specific

influence on behavior, then we have no reason to assert that they have any token-specific influence on other mental states as well.

Some critics might claim that behavior aside, criteria based on phenomenology or normativity are sufficient in themselves for positing intentional states. Behavioral criteria for intentionality play an essential role in my argument, so I'll make two points in response to this challenge before proceeding. First, behavioral criteria for intentionality make a lot of sense if you assume naturalism, and you assume that our faculty for entering into intentional states arose via natural selection.¹⁹ If intentionality is an adaptation then we should expect it to influence behavior, so a behavioral criterion is appropriate.²⁰

Second, criteria for intentionality aside, prototypical intentional states such as beliefs and desires do tend to influence thought and behavior in ways that are specific to their objects. Otherwise, folk-psychological explanations would not be effective for predicting behavior. If it turns out that emotions are not like beliefs and desires in that respect, that would itself be a surprising discovery. Even if we deny that this discovery has any implications for the intentionality of emotions, it would be interesting in its own right. It would certainly put pressure

¹⁹ Though teleosemantic theories of mental content make explicit reference to natural selection, other naturalistic theories of mental content are also perfectly compatible with intentionality being an adaptation. Teleosemantic theories diverge from other naturalistic theories in claiming that evolutionary history plays a role in determining intentional contents, and my criterion is neutral on that issue. I would like to thank a referee for suggesting that I discuss teleosemantic theories in this context.

²⁰ I would like to thank a reviewer for pointing out that adaptations need not currently be beneficial, though they will tend to have some current behavioral influence.

on views that identify emotions as judgments or perceptions, since judgments and perceptions do normally have a systematic, object-driven influence on behavior.²¹

3.2.2 Affect as Information

The psychological work on misattributed emotions is fairly well-known among psychologists and among philosophers interested in psychology. Using the criterion presented above, I will investigate how this literature bears on the question of whether emotions have objects. This section will begin with a classic paper which helps to raise the relevant questions, before moving on to a recent paper which is better suited to resolving them. I make no effort to provide a systematic review of the misattribution literature. I discuss these two papers because they are the best available for investigating the relevant issues. The rest of the literature is compatible with TYPE, just less well suited for distinguishing between TYPE and TOKEN.

Schwarz and Clore (1983) called random University of Illinois students and asked them to rate their happiness and satisfaction. The calls varied in two respects. Participants were called on either sunny or rainy days, and in some cases the caller described the survey as an investigation of the effect of weather on mood, while in other cases they left out this information. The participants who were called on sunny days reported greater well-being than those called on rainy days, but the effect did not hold when the caller described the survey as related to the weather.

²¹ See e.g. Solomon (1976) or Nussbaum (2001) for theories of emotions as judgments, and Prinz (2004) for a theory of emotions as perceptions.

On the basis of this experiment and others like it, Clore et al. (2001) developed the Affect-as-Information hypothesis (INFORMATION).²² According to INFORMATION, “affective feelings influence judgments when they are experienced as reactions to what is being judged (p. 29).” They use the term ‘affective’, rather than ‘emotional’, to remain neutral between emotions, which they consider to have objects, and moods, which they think have diffuse objects or lack objects entirely (p. 32). When we are in an affective state we take that state to be informative about the world. Anger alerts us to offense, disgust to contamination, fear to danger, etc. However, we are somewhat selective about how we use that information. We only take an affective state to be informative about something that we encounter if that state occurred in reaction to that object. To put things differently, we only take affective states to be informative about their apparent intentional objects.

INFORMATION neatly explains the data from the weather study. Participants called on sunny days were more likely to be happy, and those called on rainy days were more likely to be sad. When they were asked about their lives in general, positive affect alerted them to benefit, negative affect to harm. However, among the participants who were told that the study was about the influence of weather on affect, many realized that the weather was responsible for their current highs or lows. As a result, INFORMATION predicts that they would ignore their current affect when forming a judgment about their lives in general, and the data bears this out.

In this study you can see the rudiments of the argument against TOKEN. Affective states do influence our judgments about some objects and not others. However, according to

²² A number of other studies have been done in the wake of Schwarz and Clore (1983) that are based on INFORMATION. For recent examples see Wheatley and Haidt (2005) and Schnall et al. (2008).

INFORMATION, the scope of an affective state's influence is not determined by any features of the state itself. What matters is the *apparent* object of the affect. The same affective state either will or will not influence judgments about a particular, depending on whether the subject attributes the right intentional content to that state. Actual, as opposed to apparent, intentional objects of affective states serve no explanatory function in INFORMATION.

One might object that attributing an intentional object to an affective state is sufficient for it to have an object. However, this does not respect the distinction between appearance and reality with respect to the objects of intentional states. I might take myself to desire that *p*, when I actually desire that *q* (where *q* is not identical to *p*). My attributing to myself the desire that *p* may influence my thought and behavior in ways that are specific to *p*, since that attribution is itself an intentional state that concerns *p*. However, we should distinguish the effects of that attribution from the effects of my actual desire, since they are distinct mental states. Even if I am entirely unaware of my desire that *q*, that desire should influence me in ways that concern *q*. Any state that is about an object will influence thought and behavior in ways that concern that object, regardless of whether the individual attributes the object to that state. This is just another way of expressing the criterion.

Based on this reasoning, the study gives us reason to doubt that affective states have objects. INFORMATION, as applied to this data, suggests that affective objects are fictions. It is as fictions that they are theoretically indispensable. Ironically, the originators of INFORMATION are themselves defenders of TOKEN. Clore et al. (2001) explicitly distinguishes between affective states that do and those that do not have objects. It is not clear

how they reconcile INFORMATION with TOKEN, since the evidence for INFORMATION seems to be evidence against TOKEN.

That said, it would be premature to reject TOKEN at this point. First of all, it is necessary to address the distinction between emotions and moods. In Schwarz and Clore (1984) the relevant affective states were elicited by good or bad weather. It is natural to suppose that weather would put participants in good or bad moods, and that moods, rather than emotions, were responsible for influencing participants' judgments in that study. A critic might argue that INFORMATION is a good theory if you want to explain moods. If moods, as some suppose, do not have intentional objects, it is not necessary to posit such objects to explain their effect on judgment. Emotions, on the other hand, may have object-specific effects which are not captured by INFORMATION. This study is therefore unable to settle the issue of whether emotions have intentional objects. A successful critique of TOKEN would require a similar study where, given the way in which the affect was elicited, critics could not reasonably claim that moods, rather than emotions, were involved.

Aside from that issue, INFORMATION has additional shortcomings as an explanation of the misattribution data. In the next section I will discuss a study that highlights these shortcomings and suggests an alternative theory.

3.2.3 Discounting or Correcting

DeSteno et al. (2000) contains a number of studies, but this discussion will focus on study 4, where they compare INFORMATION with another similar theory. According to INFORMATION, when participants who are in an affective state consider an object, they

consider whether it is the intentional object of their affect. If they decide it is not, then they will ignore or discount their affect when forming judgments about that object. DeSteno et al. propose an alternative theory which involves correction, rather than discounting. According to their proposal (CORRECTION), when participants make judgments about a given object, affective states automatically influence that judgment. Participants then consider whether the object they are judging is the intentional object of their affect. If not, they adjust their judgment about the object to correct for the influence of the affect, which in such cases is seen as a harmful bias.

The end results of INFORMATION and CORRECTION are similar. When 1) participants take their affective states to be about token objects, and 2) participants are making judgments about those same objects, then and only then will those affective states influence their considered judgments. A considered judgement, for these purposes, is the judgement that results after any discounting or correcting has been done. The difference between the theories concerns how participants reach those considered judgments. According to CORRECTION, affective states have an automatic influence on judgment, and that influence occurs prior to participants attributing an object to their affect. INFORMATION, on the contrary, holds that affective states have no influence on judgment unless participants decide they should, and it only occurs after they attribute objects to those affective states.

DeSteno et al. (p. 408) designed their experiment based on the realization that emotion regulation requires work, and INFORMATION and CORRECTION differ over what sort of work individuals need to do in order to regulate their emotions. There is a personality scale, Need for Cognition (NFC), which measures an individual's willingness or unwillingness to do

cognitive work. By observing the difference between participants with high (cognitively industrious) and low (cognitively lazy) levels of NFC, they can test the two models.

First, the experimenters induced anger, sadness or happiness by having participants describe events in their lives when they had strong emotions of that type. Note that contrary to Schwarz and Clore (1983), there is every reason to believe that this procedure would, at least initially elicit an emotion rather than a mood. Second, participants were given a questionnaire to test the effectiveness of the emotion-elicitation. Finally, they were asked how likely certain types of events were. For example, they might be asked, ‘Of the 50,000 students currently enrolled at Ohio State, how many will experience the death of a loved one (e.g., close relative, close friend) within the next year?’ or ‘Of the 20,000 people who will drive on U.S. highways today, how many will be sharply cutoff by another driver?’ (p. 415). Depending on how participants correct or discount, we would expect sadness to increase estimates of the likelihood of losses, anger of offenses, and happiness of benefits. These are considered ‘emotion-congruent’ effects: effects in which emotions (e.g. sadness) inflate estimates of the type of event that tends to elicit that kind of emotion (e.g. losses). All of the event-types from the third stage of the experiment involved losses, offenses, or benefits, so it was possible to compare the prevalence of these effects in high and low NFC participants.

Both INFORMATION and CORRECTION, as described above, predict that emotions will only influence considered judgments when the participants think that their emotions are about the subject matter of their judgments. Via discounting or correcting, participants should avoid emotion-congruent effects when their emotion is clearly not directed at the object of their task. The stages of the experiment were sequenced in such a way as to make it perfectly clear to

participants that they were made angry, sad or happy by thinking about the autobiographical events that they described, and that these emotions were unrelated to the estimates in the third stage of the study, so a naive version of either theory would predict that no emotion-congruent effects would be observed.

This is where cognitive effort becomes relevant. DeSteno and colleagues (p. 408) predicted, on the basis of CORRECTION, that participants with low NFC would demonstrate emotion-congruent effects. CORRECTION states that emotions automatically bias judgements, which are then corrected afterwards when necessary. Lazy, low NFC participants will not correct consistently and thoroughly, so their judgments should still reflect emotion-congruent biases. That is exactly what they found

This alone would also be compatible with INFORMATION. Participants who are cognitively lazy might have a policy of assuming that emotions are directed at whatever they are currently considering. They would rarely bother to discount, and as a result, demonstrate emotion-congruent effects, as observed.

The more striking results concern participants with high NFC. In general, the cognitively industrious actually over-corrected. They gave lower, rather than higher, estimates of the likelihood of emotion-congruent events (p. 410). It is perfectly clear why, on the CORRECTION view, this would happen. First of all, according to CORRECTION, emotions influence judgments of emotion-congruent events automatically, and do so prior to any regulation by the participants. When participants decide that this influence is a harmful bias and decide to correct for it, there is no precise way for them to tell how strong a correction is required. Apparently

high NFC participants are prone to over-correct in that type of situation. This contrasts with low NFC participants, who tend to under-correct or fail to correct at all.

While defenders of INFORMATION can easily explain emotion-congruent biases in low NFC participants, they do not seem to have any way to account for emotion-incongruent biases in high NFC participants. According to INFORMATION, the way that you avoid bias is by discounting affect when making judgments. Perfect bias avoidance in this task would be reflected in a lack of any correlation between a participant's current emotion and their performance in the task. Good but imperfect bias avoidance would be reflected in a small emotion-congruent bias. Only CORRECTION can explain emotion-incongruent biases.

CORRECTION supports the Type Influence Theory (TYPE). According to TYPE, different types of emotions influence our responses to different types of objects. This influence has two critical features: a) it does not discriminate between different objects of the same type, though it does discriminate between different types of objects, and b) it occurs automatically and involuntarily. In this study emotions inflated estimates of the likelihood of the types of events which elicit that emotion. That is certainly compatible with TYPE, but TYPE would be compatible with other type-specific influences as well.

We can see CORRECTION as a stronger version of TYPE. In addition to the features that are distinction of TYPE, CORRECTION adds a few of its own: c) unless the emotion influences judgment about a token object which is believed to be the object of the emotion, that influence will be considered a harmful bias, d) participants respond to apparent harmful biases via correction, and e) the degree of correction is in part a function of participants' willingness to apply cognitive effort. This study supports the full version of CORRECTION, which includes

TYPE along with these additional commitments. The goal of this paper is to develop a case for TYPE, but the case for CORRECTION seems very strong as well.

In this section I discussed two studies from the emotion misattribution literature. The conventional interpretation of these studies is that participants are easily confused about the objects of their emotions. However, I argued that they actually demonstrate that emotions do not have objects.

3.2.4 Empirical Objections

In this section I will consider some objections that concern my interpretation of the empirical research. First of all, let us return to the distinction between emotions and moods. In part 3 I criticized TOKEN by appealing to Schwarz and Clore (1983). As was noted, in that study the weather may have produced either emotions or moods. We often talk about bad or good weather putting one in a bad or good mood. If only moods were involved, then it should be no surprise that the participants' affective states do not seem to be directed at token objects.

However, that critique is not plausible when applied to DeSteno et al. (2000). In the study from that paper that I have discussed, affect is elicited by asking participants to recall episodes when they were sad, angry or happy. It is reasonable to assume that performing this task would cause participants to develop emotions of those kinds. Unless we have reason to suppose otherwise, we should also assume that it was in virtue of entering those emotions that the participants' judgments were altered in the manner that was observed.

Some critics might acknowledge that participants in the study developed emotions, not moods. They might also accept that emotions have type-specific effects, as this study

demonstrates. However, these critics might claim that this is only a side-effect. Emotions, the argument goes, primarily have token-specific effects, but also have some type-specific side-effects. In this study only the side-effects are manifest, but that may be due to the peculiarity of the design.

This objection may be framed in a number of ways. In another version, the emotion itself is directed at the object, and therefore has token-specific effects, but it is accompanied by a mood, which is undirected. My mistake, on this view, is focusing on the indiscriminate effects of the mood, and losing sight of the emotion itself. I consider this to be equivalent to the primary effect/side-effect objection, since it posits distinct mental processes responsible for type-specific and token-specific effects.

This may be the most persuasive defense of TOKEN, and it is difficult to decisively refute. My critics and I agree that there is explanatory value in positing mental processes responsible for type-specific effects on judgment and behavior. In my view these processes are emotions, while my critics either view them as moods or as aspects of emotions. The question is whether we have any reason to posit an emotional process that produces token-specific effects, in addition to the afore-mentioned type-specific effects. I have argued that the type-specific effects, plus correction processes which are due to distinct states, do all of the explanatory work. With such a good side-effect, who needs a primary-effect?

Of course, my argument so far has focused exclusively on two psychological experiments. Perhaps we do need to posit additional token-specific processes to explain our emotional responses outside of the laboratory. The goal of the rest of the paper is to argue that my view does in fact have the power to explain normal emotional behavior. It may seem as

though we need objects to explain emotions, but that is because so few have tried to do the explanatory work without them.

Before leaving the laboratory, I will address one more alternative to my interpretation of the data. There is a theory of emotions which goes back to Schachter and Singer (1962), according to which emotions have two elements. One element is undifferentiated arousal, and the other is an attribution, in which participants take themselves to be in one sort of emotion or another. Since the arousal is undifferentiated, the attribution is wholly responsible for distinguishing between types of emotions, and for causing an emotion to be directed at its object.

More recently, social constructivists such as Russell (2003) and Barrett (2006) have argued for something quite similar. Russell and Barrett refer to the first element as core affect, and claim that it only varies in valence and degree of arousal. The other element, as with Schachter's theory, involves labeling the affect, and is also responsible for the emotion having an intentional object. Russell and Barrett are social constructivists because they argue that the labels we use for types of emotions are a function of our cultural experiences. The only innate elements are arousal, valence, and the conceptual resources we use to label them. TYPE has some similarities to Russell's and Barrett's views. What TYPE calls the emotion is similar to what they call core affect, and the process from CORRECTION in which participants attribute an object to their emotion is like the conceptual element in the Russell/Barrett theory.

However, TYPE (and therefore CORRECTION) differs from the social constructivist views in two main respects. First, TYPE denies that emotions only vary with respect to valence and arousal. Fear and anger, for example, are both high-arousal negative-valence emotions, and

they have different effects on judgement (Lerner & Keltner 2001). This difference between fear and anger seems to be independent of any attribution process.²³

Second, Russell and Barrett consider the attribution (the conceptual element) to be part of the emotion, while TYPE considers it to be distinct. If you include the attribution in the emotion then TYPE would be false, since the correction, which is based on the attribution, leads to token-specific effects. However, there is no good reason to individuate emotions in this way. The emotion (which is roughly equivalent in the constructivist model to core-affect) and the attribution operate independently of one another, and often even in opposition. They are initiated at different times, and have different causal powers. According to CORRECTION, we use attributions help us determine whether we should correct for the effects of an emotion. If you consider the attribution to be part of the emotion, then what we are doing is using part of the emotion to decide whether to correct for the other part of the emotion. There is no principled way of individuating mental states which would include the attribution within the emotion.

²³ I will not make any commitment, at this point, about how we should individuate emotions. The two main options in the psychological literature are to add additional dimensions in the manner of appraisal theories (see Elsworth & Scherer 2003) or to posit some number of basic emotions in the manner of Paul Ekman (1992).

3.3 Philosophical Responses

3.3.1 Explaining Appearances

In the preceding sections I have presented empirical reasons for denying TOKEN. However, in our daily lives we frequently attribute objects to our emotions. If I cannot explain why we make this error, critics could fairly suggest that it is not an error after all. I will argue that we ordinarily attribute emotional objects for a wide range of reasons. In developing my explanation I will appeal to four contributing factors: the conflation of occurrent and dispositional states, the salience of emotion-elicitors, the need to comply with social norms, and the utility of selective suppression more generally.

The first source of error is that we often conflate occurrent and dispositional states of all kinds, including emotions. Dispositional fear is not, strictly speaking, fear. It is just a disposition to enter into a state of fear. The systems that elicit emotions in different people respond to different things. My fear-eliciting system is highly sensitive to the perception or vivid imagination of spiders. To speak more colloquially, I am afraid of spiders. It is natural to describe dispositional fear in object-directed terms. However, this does not mean that fear itself is an object-directed state. My occurrent state of fear, whether it is elicited by spiders or heights, will have the same effects on my subsequent responses.²⁴

²⁴ Demian Whiting (2011, p.10) gives a similar response as part of his own argument that emotions do not have objects. Otherwise his arguments are quite different than the ones in this paper, relying primarily on his interpretation of emotional phenomenology.

It may seem that I am walking into a familiar trap. Anthony Kenny (1963, p. 72) argued that the object of an emotion is not always its cause. After all, sometimes we seem to be afraid of future events, which cannot cause anything in the present. Deonna and Scherer (2010, p. 45) provide a different type of example: “My fear of the elephant may have been caused by rumors concerning its having escaped the zoo. Here the object of my fear is still the elephant, but the cause is the rumors.” I argued that one source of TOKEN is that we notice the eliciting conditions of emotions, and take the emotions to be about, in an intentional sense, the elicitors. But this strategy seems unable to deal with cases where the apparent object of the emotion is different from its cause.

My solution is actually the same one that Kenny (p. 72) attributes to Descartes. We do not take emotions to be about their causes. Instead, we take emotions to be about the objects of the mental states that elicit them. We notice that spider-representations elicit fear, so we take ourselves to be afraid of spiders. This takes care of emotions that are apparently about the future, since in those cases the emotions are elicited by representations concerning the future. As for the elephant case, the fear seems to be directed at the elephant because it was caused by imagining a runaway elephant rampaging through town. Kenny (p. 73), in rejecting this solution, said, “If the relation between an emotion and its object were one of effect to cause, then it would be only by induction and tentative hypothesis that one knew on any particular occasion *what* one was afraid of or excited about. But this is sometimes obviously untrue.” The empirical evidence that I have presented suggests that we do, in fact, attribute objects to emotions via induction and tentative hypothesis, and that our confidence in these attributions is misplaced.

Second, the object that elicits an emotion is in some cases the only present object of a type that is salient to that emotion. Imagine that David is typing away at his desk when he notices a giant hairy spider. Seeing the spider will elicit fear, which will make him more sensitive to any apparent danger. Since the spider is the only dangerous-seeming thing in his environment, the effect of his emotion will only be manifest in his response to the spider. It will seem as though his fear influences him in a way that is specific to the spider, which is also the object of the perception which elicited that fear. This sort of case reinforces Kenny's intuition that we immediately know the objects of our emotions. In fact, the evidence I have presented suggests that David's fear would have increased his sensitivity to any other apparent dangers, if he were exposed to any.

Since I deny that emotions have targeted effects, it may seem as though I cannot explain targeted motivational behavior. We need to explain why David would flee the spider if his fear were not directed at that spider. However, we should not overlook the motivational role played by desires. Setting fear aside, most of us desire to avoid bodily harm. If I am driving a car at a moderate speed, I will avoid wrenching the wheel over in a way that would send me crashing into a tree or into oncoming traffic. I avoid doing this even when I am in a state of deep relaxation, listening to music as I drive. When I am in a state of fear, like David in the spider example, my desire to avoid any apparent dangers will be magnified, but the only kind of token-specific motivational state we need to appeal to is desire. Fear magnifies my desire not to be harmed, and that desire, along with my beliefs about which actions will lead to avoiding harm, explains why I perform one action rather than another.

We can clarify this point by considering an analogy between emotions and drives such as hunger and thirst. We sometimes say things like “I’m hungry for a big bowl of ice cream,” which on the face of it implies that states of hunger can be directed at types of food. However, we do not take these claims at face value. Hunger is just a mechanism that regulates our general desire to eat. Instead, when we say that we are hungry for ice cream, we really mean that we desire ice cream, and that desire is strengthened by (among other factors) a global state of hunger. Anyone who has ever attempted to maintain a diet is well aware that our desire to eat appealing foods can persist in the absence of actual hunger. Likewise, fear globally strengthens our desire to avoid danger, but that desire persists in the absence of fear.

Third, attributing objects to our emotions helps us comply with social norms. Susan may have a contentious argument with Janet, and then encounter Sue, who makes a mildly obnoxious comment. Anger causes all offensive things to seem more offensive, so Susan will find it difficult to be as tolerant of Sue as she normally would be. It may seem perfectly appropriate at that moment to respond tersely or with a sarcastic remark, but if she does so, the hostility of her behavior will be out of proportion to Sue’s offense. Showing a bit of anger now and then is seen as normal, reflecting strength and passion. However, people who let their anger at one person color their responses to others just end up looking like jerks. In order to avoid looking like a jerk without suppressing our anger entirely, we need to attribute objects to our states of anger, and suppress our responses accordingly.

It would be natural to describe this norm by saying that if you are angry at one person, you should not take out that anger on someone else. This description presumes that states of anger are, at least in some cases, directed at people. However, it is possible to describe this norm

by appealing to the etiology of the anger rather than its object: if a state of anger is elicited in response to a thought about p , then it is inappropriate to act towards q in a way that is influenced by that anger, unless q is suitably related to p . This description even preserves the force of the norm. If someone makes you angry,²⁵ then given human nature, they should expect you to respond forcefully, perhaps out of proportion to the original offense. The acceptability of that response will depend on the context. However, if you are angry, and you are interacting with someone who was not responsible for making you angry, then it would be unfair for them to deal with your belligerent behavior under almost any context.

This point becomes more clear when we consider the norms governing moods. When we encounter someone who is in a bad mood, the norms are much the same as encountering someone who is angry, in cases where we played no relevant role in the etiology of that anger. Out of prudence and sensitivity we step lightly around such people. At the same time, we consider it inappropriate when they do lash out at us without our doing anything to deserve it. This shows that these norms do not rely in any critical way on emotions having intentional objects.

Fourth, attributions can help us identify when to suppress an emotional response that will interfere with the task at hand. Emotions can be beneficial because they get us in the right mindset to address certain kinds of problems or opportunities. However, a mindset that is appropriate for addressing one situation may not be helpful when dealing with another. If Joanne loses a job she will become sad, which will make her reflective and risk-averse (Lerner and Keltner 2001; and Bodenhausen et al. 1994). This response is appropriate when dealing with the

²⁵ To be precise: if representations of them make you angry...

aspect of her life in which he suffered the loss, but general moroseness might lead her to pass up opportunities for growth or pleasure. Joanne, by attributing her sadness to the loss of her job, sets the groundwork for suppressing the symptoms of her sadness when dealing with other issues. We attribute objects to our emotions because they do not really have them. Our emotional responses are indiscriminate, so we need to discriminate.

3.3.2 Objections from Philosophical Theory

I will finish by considering a pair of additional objections that would likely occur to philosophers working on emotion. As in the last section, my replies are quite brief, and further moves could be made on both sides. Consider this a sketch, rather than a full treatment, of how TYPE can address issues that are found in the literature.

First, one might argue that emotions have intentional objects by appealing to the semantics of emotion ascriptions. As Graham Forbes (1997) puts it, ‘Lex Luthor is afraid of Superman’ is true, but ‘Lex Luthor is afraid of Clark Kent’ is false. Intentional state terms like ‘believes’ and ‘desires’ create opaque contexts. Emotion terms also create opaque contexts, so it is natural to suppose that emotions are themselves intentional states.

Graeme Forbes begins his explanation of opacity in transitive verbs by suggesting that, ‘Luthor fears Superman, so-labelled, but not Clark, so-labelled’ (Forbes 1997, p. 10). This analysis seems to be correct, as long as Luthor’s fear of Superman is understood to be a dispositional emotion. When we talk about emotions in a dispositional sense, we are interested in their elicitation conditions, so the elicitation conditions of an emotion determine the truth values of sentences that describe them. Since it is possible for ‘Superman’-representations to

elicit fear in an individual for whom ‘Clark Kent’-representations do not, ‘is afraid of’ can create an opaque context. Occurrent states of fear, however they are elicited, do not have objects.

Second, Deonna and Scherer (2010 p. 48) argue that emotions have objects by appealing to utility:

“Why should this important mechanism be so strongly affected by transient noise? Would it not seem more reasonable to assume that emotions are elicited and differentiated by the judgment or appraisal of what philosophers have called the “intentional object” and that conscious feeling reflects the content of appraisal as well as its mental and bodily consequences, in particular, adaptive action tendencies?”

I have two responses to this objection. First, we have empirical evidence showing that emotions *really are* strongly affected by transient noise, and no evidence that adaptive action tendencies are directed exclusively at intentional objects. A satisfactory theory will not only explain the utility of emotions, but also the respects in which that utility is limited.

Second, states with only type-specific effects can still do a great deal of good. DeSteno et al. (2000) argue persuasively that we use correction processes to limit useless or harmful effects. As long as we are willing to devote the necessary cognitive resources, we can have the next best thing to truly token-specific effects. But even unfiltered type-specific effects can be useful. Emotions are often fleeting, and they are often elicited by perceptions of people or things in our immediate proximity, or thoughts about people or things that are not present, but which require immediate attention. If we assume that emotions are most likely to be useful when they influence our response to p , where p is the intentional object of the state which elicited that emotion, we

should expect emotions to be useful much of the time. The effects that are documented in the misattribution literature are of great theoretical interest, but they are produced under highly contrived circumstances, so it is easy to exaggerate the scope of their practical importance.

3.4 Conclusion

The goal of this chapter was to argue that emotions do not have particular objects, and to do so without relying on the inter-systemic view. This may lead to some confusion, so I will conclude by connecting the terminology that I develop in different chapters.

In this chapter I discussed the effects of emotions on judgment, since those were the effects measured by the most relevant psychological research. My conclusion was that emotions influence broad types of judgements, rather than judgments directed at some particular object. This is, at least superficially, distinct from the generic impulses that I discussed in the previous chapters, since those influence broad types of desires.

My hypothesis is that these are two sides of the same coin. Anger changes judgments about offenses because it temporarily increases our aversion to them. Fear changes our judgments about dangers because it temporarily makes us less willing to tolerate them. Similarly, when we correct for the effects of emotion on our judgment, which I have argued we sometimes do, we are indirectly correcting for the effects on our desires. The generic impulse influenced our desire which in turn influenced our judgment, and if we think that this generic impulse was not relevant to the subject matter of our judgment, we will consider that influence to be a harmful bias.

On one level, this proposal requires a great deal more explanation. Ultimately I hope to explore further how desires mediate this influence on judgment. However, in the meantime the arguments in this paper are useful for supporting the inter-systemic view in a more general manner. If emotions did, through their influence on judgment, manifest that they have particular objects, then that would directly contradict the inter-systemic view. In the meantime there is further work to be done to integrate all of the evidence, but no smoking gun.

Chapter 4: Looking for Emotions in the Brain

4.1 Introduction

It is taken for granted in the literature on consciousness that while many brain areas are active during episodes of conscious experience, only some of these should be identified as the regions where conscious experience is actually located. There is a long-standing, sophisticated, and interdisciplinary debate over which regions are the neural correlates of consciousness.²⁶ In the literature on emotions and their neural basis there is not yet a comparable debate.

Casual familiarity with contemporary research might give the impression that we already know where to find emotions: areas such as the amygdala, ventromedial prefrontal cortex (VMPFC), anterior cingulate cortex (ACC) or insula. There is plenty of evidence linking these brain areas to emotions, just as there is plenty of evidence linking a wide range of brain areas to consciousness. Identifying these links does not settle the debate about where to locate emotions any more than it does in the case of consciousness. It only provides us with the material we need to begin the debate in earnest.

In this chapter I will identify two theoretical challenges for locating emotions in brain regions. First, we need to develop a method for distinguishing between affective states such

²⁶ See e.g. Crick and Koch (2003), Block (2007), and Lau and Rosenthal (2011).

as emotions and moods, on the one hand, and conative states such as desires and aversions, on the other. Affective and conative states are both implicated in motivating action, and they seem to be causally interrelated. We cannot use empirical evidence to locate emotions unless we know how to empirically distinguish emotions from desires.

Second, we often identify emotions by their bodily expressions. Neuroscientists in particular tend to use this method, since rats and monkeys are frequently their research subjects. If an animal undergoes certain bodily changes, it is reasonably safe to assume that it has entered an emotional state, so it is natural to infer that brain areas that are critical for producing those bodily changes have some relation to emotions. However, in order to assign a more specific role to such a brain area, it is necessary to develop a theory of the role that bodily changes play in emotional episodes. A number of such theories have been proposed, and they have radically different implications.

The goal of this chapter is to clearly characterize these challenges, and propose general strategies for addressing them. For that reason, I am going to focus almost exclusively on two brain areas, the VMPFC and amygdala. I chose these areas because they are frequently discussed as emotion-areas, especially in the well-publicized research familiar to philosophers, and some of the research on these areas highlights the theoretical challenges that I will discuss. Similar challenges arise for interpreting the role of other brain regions, but I will not address them here.

In section 2 I will set up a framework for distinguishing affective and conative states, which I will connect with the reinforcement learning framework that is pervasive in affective neuroscience. In section 3 I will begin to address some of the most prominent findings, primarily those that associate activity in the ventromedial prefrontal cortex (VMPFC) with emotions, but

also the anterior cingulate cortex (ACC). I will argue that these findings supply the wrong sort of evidence for locating emotions in the brain, and that they may tell us more about desires than they do about emotions. In section 4 I will turn to the amygdala. After reviewing some neuroscience research, I will explore the ways in which interpretation of those findings depends on theories of how emotions relate to bodily changes. I will argue that the evidence is compatible with the view that emotions are constituted, at least in part, by peripheral states.

4.2 From Folk Psychology to Affective Neuroscience

Joseph LeDoux (2012, p. 653) asks in a recent critique of the affect neuroscience literature,

If we don't have an agreed-upon definition of emotion that allows us to say what emotion is, and how emotion differs from other psychological states, how can we study emotion in animals or humans, and how can we make comparisons between species?

His answer, "that we fake it," is blunt but not unfair. For example, a review of the neuroscience of emotion by Rudolf Cardinal and colleagues (2002) begins by discussing the difficulty of defining emotions, and distinguishing them from other mental processes. The authors conclude, "Consequently, it is useful to consider under the umbrella of emotion those neural processes by which an animal judges and represents the value of something in the world, and responds accordingly (Cardinal et al. 2002, p. 322). This exhibits just the sort of error that LeDoux criticizes, defining emotions in a way that includes diverse mental phenomena that require different theoretical treatment.

Specifically, their approach risks conflating emotion with two other classes of mental states: purely cognitive evaluative judgment²⁷ (PCEJ) and desire. By ‘PCEJ’ I mean an ordinary judgment in virtue of which we attribute positive or negative value, moral or otherwise. Looking at a work of art, I might consider it beautiful without being moved; reading about a crime in the newspaper, I might consider it horrible without being horrified; observing someone perform an act of kindness, I might consider the act admirable without admiring the actor for doing it. Judgements in which I apply value-laden predicates such as ‘beautiful’, ‘horrible’, or ‘admirable’ count as evaluative. However, the only evaluative feature of these judgments is that they apply a certain class of predicates, so we should distinguish them from the affective and conative states that are frequently elicited by beautiful, horrible, or admirable things. It follows, of course, that we should distinguish their neural bases as well.

I will not discuss PCEJ’s any further in this paper, to better focus on the distinction between emotions and desires. ‘Desire’ can refer to a number of different things. Sometimes it refers to a state of lust or appetite, which might be classified as an emotion or a drive. However, in what follows I will not use ‘desire’ in those ways. In my usage, to desire is simply to want something, as when we want to eat pretzels, want to get a good grade on an assignment, or want our hair to be arranged in a certain way. In addition to positive desires, as in the examples above,

²⁷ I would like to thank a reviewer for suggesting that I discuss evaluative judgments in this context.

there are also negative desires, which are sometimes called aversions. One might be averse to smelling sulfur, being laughed at, or paying high ATM fees.²⁸

We all have countless idiosyncratic desires that make up our personal taste, but there are also whole classes of desires which tend to be very similar between individuals. For example, we desire to avoid bodily harm or contamination. There is debate over whether desires essentially motivate action, or whether they have some other essential feature, such as the production of pleasure or the facilitation of learning.²⁹ However, whatever one's view about the essential nature of desire, the most direct behavioral evidence about an individual's desires will come from what they strive to get and what they strive to avoid. In general, people strive to avoid bodily harm or contamination, which provides evidence about certain classes of desires.

However, there is significant overlap between the behavioral profiles of certain types of emotions and desires. Again, there are disputes about the essential features of emotions. Emotions are characterized by various theorists as judgment-like propositional attitudes, structured appraisals, or bodily perceptions. Despite this diversity of views, few theorists would deny that occurrent states of fear motivate us to avoid harm, and occurrent states of disgust motivate us to avoid contamination. It would be possible to characterize the motivational profile of each type of emotion. The problem is that the motivational profile of each emotion-type also

²⁸ One view is that aversions are just desires for something not to be the case. An aversion to paying high ATM fees, on this view, is just a desire not to pay high ATM fees. Schroeder (2004) and Sinhababu (2009) both defend the contrary view, that aversion is a distinct kind of mental state. It is not important, for present purposes, to settle this issue. I will typically use 'desire' for both positive and negative desires, remaining neutral about whether they are different in any significant way.

²⁹ See Schroeder 2004 for a discussion of different theories of desire.

seems to fit a class of desires. If we are going to interpret empirical research that bears on motivation, we need a strategy distinguishing between desires and emotions.

4.2.1 Methodology

Before presenting my preferred strategy for solving that problem, I will briefly describe my general approach to applying folk-psychological terminology. Empirical investigation tells us about the different sorts of states that influence the production of action, and how they interact with each other. Pretend for the sake of simplicity that the sciences provide a unified catalog of motivational states, with their functional roles specified but without folk-psychological labels. Given this catalog, a theorist might wonder which, if any, of these states we should label ‘desire’ or ‘emotion’.

My approach is to focus on the influence that folk-psychological terminology has on scientific theory. We are accustomed to applying these terms according to implicit folk-psychological rules, so when we use them as labels for states in our scientific theory, there will tend to be spillover. Unless the folk terms that we apply seem arbitrary, like ‘charm’ for a variety of quark, we will be tempted to make inferences about the states based on their functional role in folk psychology. This is not entirely inevitable. We can, in our scientific theory, explicitly capture divergences between the inferences about emotions that are sanctioned by folk psychology and those that are sanctioned by our revisionary scientific theory. However, given the natural tendency for inferential spillover, it is beneficial to use folk-psychological terms for states of our scientific theory when there is significant similarity between the functional roles posited by folk psychology and our scientific theory.

For this reason, I will reserve terms like ‘desire’ and ‘emotion’ for whichever states will maximize the benefits, and minimize the harms that result from this spillover. We do not, of course, have available fully specified characterizations of folk psychology, or a fully developed, empirically based functional model of motivation. Given the tools available, I will introduce an example that illustrates two important features of how folk-psychology distinguishes between emotions and desires. In part three of this chapter, I will argue that attention to these features produces an unorthodox interpretation of neuroscience findings.

4.2.2 Desires and Emotions

Jane is calmly driving to work. She steers her car in such a way that it stays within the lanes, never heading off of the road or into oncoming traffic. Why does she do this? Fear does not play any ineliminable role, since I can stipulate that Jane is not in a state of fear throughout her drive. It is true that she is in some sense afraid of crashing her car. However, this is a dispositional, rather than an occurrent fear. I am in an occurrent state of fear if I am, right now, afraid. Dispositional fear is just a disposition to enter a state of fear. Jane’s having this dispositional fear just amounts to the fact that she would likely become afraid if she thought that she were in danger of crashing. Being an experienced driver, she rarely has the impression that she is in any danger when she is at the wheel, so this disposition is irrelevant most of the time.

The folk explain Jane’s behavior by positing that she has a strong, occurrent desire to avoid crashing. Desires, like emotions, come in occurrent and dispositional varieties. It is plausible to suppose that most of us dispositionally desire to avoid crashing a car, since whenever we are aware that crashing a car is a relevant possibility we desire not to do so. Jane

has this desire when she drives, since crashing is a relevant possibility when driving, and this desire explains her danger-avoidance. In her case, this desire is so effective at motivating safe behavior that she does not become occurrently afraid. Therefore in my story she has an occurrent desire to avoid crashing but does not enter an occurrent state of fear.³⁰

Extrapolating from this example, we can develop a general picture of when related emotions and desires are elicited. I'll begin with harm-aversion and fear, before generalizing further. If subject *s* believes occurrently or dispositionally that event *e* would constitute harm, then *s* has a dispositional desire to avoid *e*, and a dispositional fear of *e*. If the subjective likelihood of *e* passes a threshold *a*, then *s* develops an occurrent desire to avoid *e*. If the subjective likelihood of *e* passes threshold *b*, where *b* is higher than *a*, then *s* will enter an occurrent state of fear. It is plausible to suppose that when the subjective likelihood of *e* passes threshold *b*, the desire that was elicited at threshold *a* plays a role in eliciting the emotion.

A similar pattern may hold for other emotion-types, and the desires that are involved in their elicitation. A low subjective likelihood of contamination, public failure, or insult would elicit desires to avoid these events, and if the subjective likelihood rises sufficiently, those desires would contribute to the elicitation of disgust, shame, or anger respectively. In my calm-driving example the desire to avoid harm is effective enough that fear is never elicited. Similarly, we can

³⁰ It may seem odd to posit a desire to avoid crashing to explain Jane's behavior, since that would intuitively have a distinctive mental quality (Mikko Salmela, personal communication). On my view, which I will not defend here, desires do not intrinsically have mental qualities. It often seems as though we feel our desire because the satisfaction or frustration of that desire, or the mere anticipation of the desire's satisfaction or frustration, elicits an emotion. The emotion, in those cases, is what we feel, but since Jane's desire does not elicit an emotion, she has no such experience.

all think of cases where we successfully avoid contamination, public failure or insult because of various kinds of effective desires.

It is important to note that the subjective likelihood of event e does not need to be consistent with s 's reflective judgment about the probability of the event. As Tamar Gendler (2008) has emphasized, there is often a contradiction between reflective judgment and other assertoric states, and in those cases the emotion will often be elicited by those other states.

The calm-driving example illustrates that there are cases of harm avoidance where the folk posit a desire, rather than a state of fear. It is reasonable to take a revisionary approach if it turns out that these theoretical posits fail. I will argue that neuroscientists are frequently interpreting their results in a manner that deviates from folk usage, and that in some cases this deviation results from confusion over how to use terms such as 'emotion', rather than from a deliberate revisionary strategy.

To summarize: there is no type of motivation, either to approach or avoid, that by itself counts as evidence of an occurrent emotion. Instead, we ordinarily identify an occurrent emotion by observing a temporary, general change in the strength of certain types of motivations. Other emotions work just as well as fear to illustrate this point. Punishing a transgression, by itself, merely demonstrates a desire to punish that transgression. However, if Joe demonstrates a temporary increase in his desire to punish transgressions, and that effect holds across unrelated transgressions, then we have reason to believe that Joe is angry.

4.2.3 Reward and Punishment

One of the problems that arises in theorizing about the neuroscience of emotion is that the experiments are difficult to interpret using concepts that are typically deployed in philosophical theories. This literature relies heavily on Pavlovian and operant conditioning paradigms, which are primarily familiar to philosophers from the bad old days of behaviorism.

Theories of reinforcement learning have progressed a great deal since Skinner. Computational neuroscientists now match reinforcement learning models to activity in the brain, so this is not a behaviorist research program.³¹ However, the core concepts that researchers use in designing and interpreting this research include reward, punishment, association, and reinforcement. Neuroscientists use this research to explain emotions, but this type of interpretation is plainly theory-laden. Philosophers do not uncritically accept theoretical moves made by other philosophers, so it is not clear why we should be less critical of neuroscientists. In the next section I will begin to examine this research and its interpretation in detail, but first I will consider how one could even begin to give a philosophical interpretation of this type of research.

I have argued that we cannot study the neuroscience of emotion without considering how emotion relates to desire. This may seem like a burden, but it actually ends up as a benefit, as it is easier to understand how emotions relate to reward through the lens of desire. Timothy

³¹ See Schultz et al. 1997.

³² Schroeder expands upon a prior theory developed in Dretske (1988).

Schroeder developed a theory of desire and aversion in terms of reward and punishment.³² I will not endorse his theory as is, but it is a useful starting point for developing my own approach.

Before discussing his theory, it is necessary to explain some terminology. Psychologists and neuroscientists consider a reward to be something that we will perform arbitrary work to get, and a punishment to be something that we will perform arbitrary work to avoid (Rolls 2005, p. 2). If when presented with buttons A and B, you will push whichever button you think will dispense a can of soda, then getting soda is a reward for you. If you will press whichever button you think will prevent you from receiving an electric shock, then receiving an electric shock is a punishment for you. Rewards and punishments are considered reinforcers, since they reinforce the performance of certain behaviors and discourage the performance of others. According to theories of reinforcement learning, we are constantly monitoring the conditions under which reinforcers are present or absent, which are known as reinforcement contingencies. We associate reinforcers with the actions and stimuli that predict their presence, and we use the occurrence of these associated phenomena to predict future presence or absence of reinforcers.

According to Schroeder's reward theory of desire, a desires that p if representations of p function as reward representations in a's system for learning reward contingencies, and a is averse to p if representations of p function as punishment representations in that system (Schroeder 2004, p. 131). Angela desires that the Portland Trailblazers will win if she is sensitive to any factors that seem to bear on whether Portland will win, including any factors that may be under her control. Specifically, she will monitor these factors in a manner that is consistent with

the models developed by computational neuroscientists. Schroeder takes conformity to these models of reinforcement learning to be the only necessary feature of desiring that Portland win.

If Schroeder is right, we can use computational theories of reinforcement learning to locate desires in the brain. Whatever region of the brain is responsible for the fact that *p* is a reward for *a* is also responsible for the fact that *a* desires that *p*. Similarly, whatever region makes *p* a powerful reward for *a* makes *a*'s desire for *p* a strong desire. There is a great deal of research on the neural systems that subserve the processing of reward and punishment, so Schroeder's theory makes locating desires in the brain fairly straightforward.

Of course this only works if you accept his theory. There are multiple alternative theories of desire in the literature, but non-specialists are likely to think of desires as states that motivate action.³³ Schroeder thinks that desires typically do motivate action, but he thinks that this is a contingent feature. The only necessary feature of desires, on his view, is that we calculate in a certain manner the contingencies under which they will be satisfied. It may seem counter-intuitive that prediction, rather than motivation, is the essential feature of desires. Schroeder has replies to both of these critiques. However, for present purposes it is unnecessary to evaluate their effectiveness. What we need is a way of interpreting reinforcement learning data in terms of either emotions or desire. Schroeder's reward theory of desire provides a way to do this, but a weakened, more easily defended version of his theory would also serve that purpose well enough.

Recall that psychologists consider *p* to be a reward for *a* if *a* will do arbitrary work to get *p*. Critics of Schroeder's theory would be likely to accept the following: if *a* (positively) desires

³³ See, e.g. Smith 1987.

that p , and a believes that the best way to make it the case that p is to w , then *ceteris paribus*, a will be disposed to w . That is enough for p to count as a reward, given the psychological definition. Perhaps not all rewards are the objects of desires, but the objects of desires are all rewards. To be more precise, the weak reward theory accepts that if a desires that p then p is a reward for a , but does not assert the converse. Likewise, it holds that the objects of our aversions are all punishments, but does not claim that all punishments are objects of aversions.

Critics who insist that desires are motivational states should be satisfied, since the feature of desires that I appeal to in my argument is their motivational role. The worry that some reward processing is too primitive to count as desire is also no longer relevant. There may be states that fall short of desires that motivate a to make it the case that p , but the weak reward theory does not deny this. It is only committed to the claim that when a does desire that p , a will be disposed, *ceteris paribus*, to perform any action that will make it the case that p . Schroeder would not be entirely satisfied, since he thinks that the motivation of action is a contingent feature of desires. However, the systems that implement the type of learning that Schroeder identifies with desire do tend to motivate action.

This theory is almost as useful as Schroeder's strong reward theory for locating desires in the brain. We cannot say that any region in the brain which is responsible for p being a reward for a thereby is responsible for a 's desire that p , since this might be the location of reward-processing which falls short of desire. However, only brain regions that contribute to p being a reward for a are candidates for being the location of a 's desire that p . That narrows the search considerably. We can narrow the search for desires in the brain to regions that are responsible for reinforcement.

4.3 The Ventromedial Prefrontal Cortex

4.3.1 Damasio and Somatic Markers

With these theoretical tools available, I will now turn to the empirical evidence. The ventromedial prefrontal cortex (VMPFC) is considered to be a significant emotion-center in the brain. The VMPFC, as the name suggests, is located in the lower-middle part of the prefrontal cortex. The most famous research on the VMPFC and emotion was performed by Antoine Bechara and Antonio Damasio (Bechara et al. 1994). They observed that patients with VMPFC lesions had normal cognitive abilities as measured by available tests, but still demonstrated poor decision-making outside of the lab. In order to explain this discrepancy the researchers developed the Iowa Gambling Task. In this task, participants are given a choice of four decks of cards. In each round they must take a card from one of the decks, and depending on the card they gain or lose different amounts of money. The cards in two of the decks offer high gains or high losses, with losses that on average outweigh the gains, and the cards in the other decks offer lower gains or lower losses, with gains that on average outweigh the losses.

Participants without brain damage eventually figure out the differences between the decks, but even before they can articulate what those differences are they demonstrate a preference for the good decks. During this stage of beneficial behavior without an explicit understanding, normal participants demonstrate increased skin conductivity prior to selecting cards from the bad decks. Skin conductivity is a symptom of autonomic arousal, which in turn is associated with a variety of emotional states. Patients with VMPFC damage performed much

worse at the task. They persisted in picking cards from the high-risk, high-reward deck. Unlike the other participants, they never demonstrated increased skin conductivity in advance of selecting cards from either deck.

The interpretation that Bechara et al. (1994) developed, and which Damasio (1994) subsequently popularized, is that VMPFC patients suffer from an emotional deficit. In advance of choosing a card, normal participants make an implicit judgment of the outcome. Although this judgment is not available to consciousness, it has the power to elicit an emotion. Damasio, who is a follower of William James (1922), believes that during emotional episodes we undergo distinctive bodily changes, and he identifies emotional experience with the perception of those changes. When healthy participants consider selecting from the bad deck, they 1) implicitly judge that choosing from that deck would be bad, 2) undergo a negative emotion in response to considering that prospect, which they associate with that action, and then 3) that association discourages them from selecting cards from the bad deck. VMPFC damage prevents the elicitation of certain negative emotions, so VMPFC patients are not discouraged from selecting cards from the bad deck.

This story generalizes beyond the laboratory. According to Damasio, emotions play a critical role in helping us make decisions. When we consider a course of action we typically enter an emotional state which is a function of the anticipated consequences of that action. For example, if we anticipate harm we might enter into a state of fear, and if we anticipate a benefit we might enter into a state of joy. These emotional feelings are then associated with the possible action. He calls this the somatic marker theory, because alternatives are marked with associated somatic or bodily feelings.

The process of producing these somatic markers could be considered a full body loop, because it involves the brain affecting the body, after which a perceptual state in the brain responds to that bodily change. It would be relatively costly to perceptibly alter the body any time that an action is considered, so this is an obvious shortcoming of Damasio's theory. He dealt with this objection by suggesting that in some cases we do not enter into a full loop. Instead, we sometimes just imagine the feeling we would have if we did enter into the emotion that we associate with that consequence. He calls these imagined feelings 'as-if body loops', since we are only imagining the bodily change that would take place in a full body loop (1994, p. 184).

Damasio posits that VMPFC stores our associations between possible consequences and different types of emotions. When patients with VMPFC damage consider a given action they do not enter an emotional state or even an as-if loop. This means that they cannot use somatic markers to compare different possible actions, and they end up making worse decisions. It turns out that their decisions are much worse, which for Damasio is evidence that emotions are more important, and more beneficial, than we normally realize.

This theory has some intuitive appeal. For example, people often experience a sinking feeling in their stomach or a tightness in their chest when they consider doing something dangerous or potentially harmful. It is plausible that these feelings help deter us from making terrible mistakes. At the same time, there are empirical reasons to be dubious of this picture. The main evidence that Damasio appeals to are correlations between VMPFC damage and 1) a lack of emotional arousal in advance of selecting from the bad deck on the Iowa Gambling Task, 2) persistently selecting cards from the bad deck on that task, and 3) making harmful decisions

outside of the laboratory. Damasio's theory does connect the dots, but it is not clear that his is the best way to do it.

4.3.2 Rolls and Reinforcer Representation

Edmund Rolls (2005) has done extensive research on the VMPFC. Most prominently, he has recorded activity in individual VMPFC neurons in macaque monkeys as they respond to reinforcers. His research is best understood in the context of the neuroscience of perception, and our knowledge of the contribution to perception that are made by various brain regions. Different regions are involved for each sensory modality, but they follow the same general pattern. Earlier regions represent features of a stimulus. They respond differently to the same object depending on orientation and distance, because different features will be apparent. Further along are regions that represent objects in a non-perspectival manner. In vision the inferior temporal visual cortex plays this role. Still further along is a region that represents the reward value of stimuli.

Rolls presents extensive evidence that VMPFC represents reward value across various sensory modalities. In one type of experiment, researchers identified individual VMPFC neurons that respond to the smell of different types of appealing fruit juice. Initially one VMPFC neuron fired whenever the monkey smelled blackcurrant juice, and another whenever the monkey smelled mango juice. After the monkey had been fed a great deal of the blackcurrant juice, the former neuron would no longer fire in response to blackcurrant juice, while the latter still fired in response to smelling mango juice. Rolls' (p. 104) interpretation is that the first neuron did not fire in response to the presence of blackcurrant juice, it fired in response to the reward value of the juice.

In another paradigm, monkeys can choose to lick a square or a triangle. Initially, licking the square produces a reward in the form of fruit juice squirted into their mouths, and licking the triangle produces a punishing squirt of salt water. In the middle of the experiment the conditions are swapped, with the square producing a punishment and the triangle a reward. There was one neuron in VMPFC that fired in response to the sight of the square before, but not after the switch, and a different neuron that fired in response to the triangle after, but not before the switch. The behavior of the monkey - whether it licks the square or the triangle - is correlated with the firing of those neurons. Again, Rolls (pp. 109-110) interprets this as evidence that these neurons do not represent the identity of a given stimulus, but instead represent its reward value.

Rolls' argument does not rely entirely on evidence from monkeys. For example, activation in VMPFC is correlated with monetary reward in humans, and VMPFC patients do a bad job of adjusting when a given type of behavior switches from producing a monetary reward to a monetary punishment (Rolls pp. 133-4). His research with monkeys, which gives a finer-grained perspective on the processing of reward, clearly cannot be done with human subjects, but the research involving humans is compatible with theories derived from his research in monkeys.

According to Rolls' theory of emotions, emotions are defined as states that occur in response to reinforcers. This is similar to the view from the view in Cardinal et al. (2002) that emotions occur in response to perceived value. The VMPFC is a critical region for evaluating whether one is in the presence of a reinforcer, so it is critical on his view for producing emotions. Rolls agrees with Damasio that VMPFC produces bodily changes, and that these bodily changes typically occur during emotional episodes, but he denies that the absence of bodily changes is responsible for the deficits demonstrated by VMPFC patients. Instead, VMPFC damage

compromises ability to perform the task because the VMPFC represents the reinforcement value of objects, and the task requires participants to compare the reinforcement value of cards in the two different decks. Failure to undergo anticipatory bodily changes is correlated with failure to perform the task, but Rolls thinks that is only because both result from failure to correctly represent reinforcement values.

4.3.3 Desire in the VMPFC

Before attempting to adjudicate between these views, it is useful to step back and consider the general issues that this research raises. Bechara et al. (1994) shows us that VMPFC damage has behavioral and bodily effects. Rolls (2005) demonstrates correlations between activity in VMPFC and perceptual and behavioral responses to reinforcers. What does all of this have to do with emotion?

Answering this question requires that we pick up where part 2 left off. I argued that a neural system responsible for a's desire that p is thereby responsible for p's being a reinforcer for a. Desire-systems are reinforcement-systems. The converse is more controversial, and unlike Schroeder I do not defend it. However, any brain region which is responsible for p being a reinforcer for a is ipso facto a candidate to be the neural location of a's desire that p. Reward-systems are candidate desire-systems, pending further evidence. Rolls presents strong evidence that the VMPFC is responsible for reward values, so VMPFC is a candidate to be the neural location of desire.

This argument leaves us with two competing interpretations of the role of VMPFC: 1) it is the location of desire, and 2) it is the location of a different variety of reward processing, with

desires located elsewhere. In favor of the first interpretation, VMPFC is located in the prefrontal cortex, which is prime real estate for higher cognitive functioning. It is adjacent to and connected with the dorsolateral prefrontal cortex, which has been tied to executive functioning and working memory (Royall et al. 2002). According to folk-psychological models of decision-making, beliefs and desires operate in conjunction, so we should expect to find desires in a region such as VMPFC which is integrated with the main higher-cognitive areas.

Of course, armchair neuroscience has its limits. We would need more research to locate desires in VMPFC with any confidence. My argument so far is merely meant to establish that this proposal is consistent with all of the evidence, and is at least as strong as alternatives that take the VMPFC to be a specialized emotion-region. In the next stage of the argument I will consider how locating desires in VMPFC would influence the debate over the VMPFC's role in emotions.

As discussed in part 2, the satisfaction and frustration of desires may be responsible for eliciting different kinds of emotions. Rolls argues that patients with VMPFC damage suffer from failure to adjust their representations of reward in response to changing circumstances. When apparent rewards turn out to be punishers, or vice versa, they do not reliably adjust their behavior. This helps explain the deficits that VMPFC patients suffer. However, if the VMPFC is the neural location of desire, we could take this explanation further: VMPFC patients suffer from a failure to adjust their desires in response to changing circumstances.

I mentioned in passing that Rolls considers emotions to be states that occur in response to reinforcers. This position echoes Schroeder's and Sinhababu's accounts of the effects of positive and negative desires. Rolls claims that the delivery of a reward leads to ecstasy, elation, or

pleasure; the omission or termination of a reward leads to rage, anger, frustration, grief or sadness; the omission or termination of a punishment leads to relief; and the delivery of a reward leads to apprehension, fear, or terror. There are remarkable parallels between this account, in which emotions are elicited by reinforcement contingencies, and the views of Schroeder and Sinhababu, who take emotions to be elicited by desires. For example, Schroeder and Sinhababu both say that avoiding/escaping the objects of negative desires leads to relief, while Rolls says that relief results from the omission or termination of a punishment. Similarly, Schroeder and Sinhababu claim that frustrated positive desires lead to disappointment, which is similar to Rolls' claim that the omission of a reward can lead to frustration. It is striking that the parallels are so strong, given that Rolls frames his theory in terms of reinforcement, never mentioning desire.

This converging evidence suggests an interpretation of the role of VMPFC which differs from what either Damasio and Rolls offer. In healthy individuals, emotions are often elicited by the apparent satisfaction or frustration of desires, or, as I argued earlier, by the elevated subjective likelihood of the satisfaction or frustration of desires.³⁴ In order to cope with changing circumstances, our desires need to change as well. As our desires change, so will the conditions that lead to the elicitation of different emotions. The VMPFC is the neural location of desires, so it is the primary brain region responsible for making these changes and for eliciting the relevant emotions. As a result, patients with VMPFC damage will struggle to adjust to changing circumstances, and they will make bad choices. VMPFC damage alters emotions because VMPFC is the location of processes that elicit emotions.

³⁴ This is distinct from the claim made by Robert Gordon (1987) that emotions just are states that result from satisfied or frustrated desires. I allow for the possibility that emotions can be elicited in a manner that does not involve any cognitive or conative states.

This explanation is consistent with the evidence that Damasio and Rolls appeal to, and has the advantage of drawing an explicit connection between reinforcement and emotion. Emotions are often elicited by desires, and desire-processing is a variety of reinforcement processing, so one way that research on reinforcement can help us understand emotions is by explaining the conditions under which they are elicited. Damasio has no theory relating emotions and desires, and Rolls just says that emotions are states that are elicited by reinforcers. This may be true, but this claim does not distinguish between the theory that VMPFC is the location of emotions themselves and my proposal that it is responsible for eliciting emotions. As things stand, we do not have any good reason for actually locating emotions in VMPFC.

4.3.4 A Word on the Anterior Cingulate Cortex

This is perhaps the best time to mention the anterior cingulate cortex (ACC), which after the amygdala and VMPFC is probably the brain area most often linked to emotions. Etkins et al (2011, p. 85) described the current debate in the following manner,

Some investigators have described evaluative (appraisal) functions of the ACC and mPFC, such as representation of the value of stimuli or actions and the monitoring of somatic states. Others hold that the ACC is primarily a generator of physiological or behavioral responses. Still others have described a regulatory role for these regions, such as in the top-down modulation of limbic and endocrine systems for the purpose of emotion regulation.

To make things yet more complicated, abnormalities in this area have long been associated with mood disorders.³⁵

³⁵ See Drevets et al. (2008).

It is hard to know what to make of these findings. Rolls (2008, pp. 144-5) found that some neurons in the ACC seem to represent the reward values of tastes, and on that basis suggests that we should see all of this data through the lens of reward representation. This does on the face of it seem promising, since individuals with mood disorders seem to respond differently to rewards. Perhaps the ACC, like the VMPFC, houses desire or desire-like reward processing. We should not, after all, expect mental state types to be constituted entirely by activity within one brain region. However, activity in the ACC and VMPFC is not entirely of the same kind, and this difference may not be easily captured in our folk psychological vocabulary.

As with the VMPFC, we cannot determine whether emotions are located within the ACC until we have considered other explanations. The obvious alternative is that the ACC is involved in calculating the reward value of stimuli, though a full analysis is beyond the scope of this chapter.

4.4 The Amygdala

4.4.1 Sources of Bodily Change

Joseph LeDoux (1996) performed some of the most influential research on the role of the amygdala. His method was to selectively lesion parts of the brains of rats with the goal of discovering which areas are necessary for developing fear responses. This process led him to the amygdala, a pair of almond-shaped nubs which are located in the interior of the temporal lobes. Normally rats will enter a state of fear in response to hearing a sound that is associated with a punishment such as an electric shock. Depending on the location of the lesion, this process can

be disrupted in various ways. LeDoux traced the flow of information from auditory centers of the brain to the amygdala, and from the amygdala to brain areas which are responsible for effects such as freezing in place, elevated blood pressure, release of certain hormones, and startling.

Based on this research by LeDoux and others working in the same vein, we now have a fairly fine-grained picture of the roles played by different parts of the amygdala.³⁶ Two areas of the amygdala are particularly important: the central nucleus (CeA) and the basolateral nucleus (BLA). Providing a full explanation of their roles would require digging deep into the subtleties of conditioning methodology. The short version is that the BLA receives input about a stimulus, and if that stimulus is associated with a reinforcer it sends a signal to the CeA. The CeA, based largely on input from the BLA, sends signals to other brain areas which are responsible for producing different physiological changes. Although discussion of the amygdala has tended to focus on fear, there is evidence that it is not specialized in this way. In fact, a review by Lindquist et al. (2011) found no evidence tying specific emotion-types to different brain areas.

In general, evidence on the role of the amygdala in humans is extensive, though less extensive than our evidence in non-human animals. Bilateral damage to the amygdala in humans interferes with (but does not eliminate) the elicitation of fear, and activity in the amygdala is correlated with various learning processes that lead to the elicitation of fear (Phelps 2006). The amygdala also has an influence on perception and attention, specifically with regard to what are known as emotional stimuli. These are stimuli of the sort that tend to elicit strong emotions. This label is potentially misleading, because processing information about this sort of stimuli does not

³⁶ See Cardinal et al. 2002 for a detailed review.

entail that one is currently in an emotional state, or that this processing is itself directly related to any emotions.

The amygdala has significant similarities to the VMPFC. Bilateral damage to the amygdala reduces performance in the Iowa Gambling Task more severely than damage to the VMPFC (Bechara et al. 1999). Rolls argues that neurons in the amygdala, as in the VMPFC, represents reinforcement values. However, the amygdala is not as effective as the VMPFC at adjusting to changes in reinforcement contingencies, as when a stimulus initially is associated with a reward and then becomes associated with a punishment (Rolls p. 168). The VMPFC seems to play a critical role in inhibiting responses from the amygdala that are no longer appropriate (Cardinal et al. 2002, pp. 338-9).

4.4.2 Contested Priority

As the review above demonstrates, the amygdala has a number of features that lead theorists to associate it with emotions. Most prominently, it is involved in the production of some of the bodily changes that occur during emotional episodes. In this section I will consider how this role bears on the question of whether emotions are located in the amygdala. Since the VMPFC plays a similar role in eliciting bodily changes, much of the discussion below will apply to the VMPFC as well.

One view of fear is that it is an inner state that makes your heart beat faster, your palms sweat, your stomach clench, and your face turn white. I will call this the subsequentist view, since it holds that peripheral changes occur after the emotion. This is compatible with the views of cognitivists in philosophy and appraisal theorists in psychology who take emotions to be a

variety of judgment. We typically undergo peripheral changes during emotional episodes, but these theorists consider them to be consequences of emotions, which are themselves judgments.

For subsequentists, the roles that the amygdala and the VMPFC play in causing bodily changes are evidence for locating emotions in those regions. They still have to decide between the views that (a) emotions are located in one region but not the other, (b) one kind of emotion is located in one region and another kind in the other, or (c) emotions supervene on activity in both regions. Another pair of possibilities is that emotions occur prior to peripheral changes, but (d) they occur prior to activity in the VMPFC and amygdala, or (e) they occur after activity in the VMPFC and amygdala, but still prior to bodily changes. So adopting a subsequentist view would not come close to settling the issue of where emotions are located.

In contrast to subsequentism, perceptual theorists such as Damasio (1994) Prinz (2004) can be classified as prioritists, since on their view the peripheral state occurs before the emotion. If you are a prioritist, evidence that the amygdala or the VMPFC is responsible for eliciting peripheral changes is evidence against locating emotions there. My proposal that the VMPFC is the location of desire is compatible with prioritism. A prioritist might hold that bodily changes are elicited by desires in the VMPFC or by the reinforcement processing in the amygdala, and then identify emotions with perceptions of bodily changes. The insula, which is involved in bodily perception and is active during emotional episodes,³⁷ would on this view be a likely location for emotions.

A third view on the chronology of emotions is concurrentism, according to which neither emotions nor peripheral states have temporal priority. Lange's (1922) peripheral view is

³⁷ See Craig 2009.

concurrentist since it identifies emotions entirely with peripheral states. On this view affective neuroscience is only relevant for identifying the antecedents and effects of emotions.

The inter-systemic view, by contrast, combines elements of subsequentism, prioritism, and concurrentism, since it considers peripheral states to be emotional constituents, along with central states that may cause or be caused by the peripheral states. However, the central constituents on the inter-systemic view differ from those posited by prominent prioritist or subsequentist views. Unlike prioritist perceptual theories, bodily perceptions are not considered to be constituents, and unlike subsequentist cognitivist and appraisal theories, the emotion does not include targeted judgments.

Additionally, I have argued that the central constituents of emotions are intimately related to desires. Sufficient subjective likelihood of a desire's satisfaction or frustration will elicit emotions, and emotions are responsible for generic impulses, which are alterations in the strength of classes of desires. In the first chapter I proposed two ways of understanding the central constituents: as desires with general content or as states at the functional rather than the representation level of the brain's organization. Following the former proposal, the central constituents may be distributed in some combination of the VMPFC, amygdala and ACC, since all of those regions are involved in the processing of reward. On the latter proposal we should expect the central constituents to have connections to these areas, since those connections would be required for producing generic impulses, but it is not clear whether the constituents themselves would be located there or elsewhere.

Since I support the inter-systemic view, I interpret affective neuroscience in a manner that follows from that commitment. However, the aim of this section has been to show that

productive analysis of affective neuroscience requires a great deal of nuance, and attention to theoretical commitments. It is one thing to say that an area of the brain is active during emotional episodes, but quite another to say how that activity relates to emotions themselves.

4.5 Conclusion

The arguments in this chapter are not, for the most part, meant to undermine the claims of affective neuroscientists. Instead, they are meant to discourage readers of the affective neuroscience literature from drawing stronger conclusions than it can reasonably support.

I point to one general and two more specific difficulties that arise in developing a theory of the neural location of emotions. The general difficulty is that the neuroscience literature is thick with the language of reinforcement learning, and it is not clear how to translate between this language and the more familiar folk psychological language we use to discuss emotions. Specifically, it is difficult to 1) distinguish emotions from desires and other motivational states, and 2) determine the temporal and causal priority of emotions and peripheral bodily changes. We should be suspicious of anyone who claims to know where emotions are located but has nothing to say about these issues.

Conclusion

During the development of what eventually became the inter-systemic view, a number of themes have come to shape the way I understand emotions. I will conclude by describing these themes, and how they have shaped my thinking.

First, peripheral manipulation alters emotions. The existence of these effects is not exactly controversial, but they are treated by many theorists as marginal phenomena. The general structure of the inter-systemic view was developed, in large part, to unify these bottom-up cases with the top-down cases that are more frequently discussed. Peripheral changes alter our motivations for the same reason that motivational changes alter our bodily periphery: because the systems responsible for motivation, action-preparation and social display need to be integrated. Emotions just are states of integration across these systems.

Second, emotions are distinct from, but intimately related to desires. It is a platitude that emotions motivate, but they are typically compared to mental states with assertoric mental attitude such as judgments and perceptions. The key psychological effect of emotions, in my view, is to alter our desires, and emotions are typically elicited by the prospect of our desires being satisfied or frustrated. Additionally, there is no reason to suppose that emotions play an indispensable role in the motivation of action. The desire to avoid harm is often effective in the absence of fear. With this distinction in mind, I find that much “affective” phenomena is actually due to desires rather than emotions, including the subject matter of much affective neuroscience.

Third, emotional effects are general. In its modest form — that emotions have some general effects — this view is widely accepted, but it is also treated as a marginal phenomenon. I treat generality as a core distinctive feature of emotionality. Motivation without generality is indicative of desire rather than emotion. I have defended a more radical claim, that emotional effects are entirely general. This view is extremely controversial in philosophical circles, and I initially defended it in order to defend a view of emotions that I no longer accept, but I have not encountered any devastating objections. My argument for this view has remained almost entirely unchanged: that positing targeted emotional effects does not lend any additional explanatory power.

Finally, emotions are episodic. This may seem too obvious to mention, but this theme actually brings out the significance of the other three. The effects of peripheral manipulation suggest that emotions are states of inter-systemic integration. However, our bodily periphery is not as effective at multi-tasking as our brain, which helps explain why we need to have general motivational changes. Thinking of emotions as episodic and general helps relate them to desires, since emotion episodically alter desires in a general manner. Thinking about the episodic nature of emotions also encourages distinguishing between occurrent and dispositional emotions. That distinction is essential to my argument for emotions lacking particular intentional objects, since it is appropriate to characterize dispositional states in intentional terms.

Thinking about emotions as general, episodic states that are intimately related to desires and subject to peripheral manipulation has lead me to develop the inter-systemic view. In fact, this way of looking at emotions is still fundamental for me, and I support the inter-systemic view because it is the only theory I have found that encompasses all of these features.

Bibliography

- Adolphs, R., Damasio, H., Tranel, D., Cooper, G., and Damasio, A. R. (2000). A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *The Journal of Neuroscience*, 20(7), 2685-2690.
- Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science*, 1(1), 28-58.
- Bechara, A., Damasio, A. R., Damasio, H., and Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, 50, 7-15.
- Bechara, A., Damasio, H., Damasio, A. R., and Lee, G. P. (1999). Different contributions of the human amygdala and ventromedial prefrontal cortex to decision making. *Journal of Neuroscience*, 19, 5473-5481.
- Berntson, G. G., Norman, G. J., Bechara, A., Bruss, J., Tranel, D., and Cacioppo, J. T. (2011). The insula and evaluative processes. *Psychological Sciences*, 22(1), 80-86.
- Blackman, R. D. (2013). Intentionality and compound accounts of the emotions. *The Southern Journal of Philosophy*, 51(1), 67-90.
- Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Science*, 30, 411-499.
- Bodenhausen, G. V., Shepard, L., & Kramer, G. P. (1994). Negative affect and social judgment: The different impact of anger and sadness. *European Journal of Social Psychology*, 24: 45-62.
- Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies*, 61, 127-148.
- Brentano, F. C. (1995). *Psychology from an empirical standpoint*. New York: Routledge.
- Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poelmann, K. M., and Ito, T. A. (2000). The psychophysiology of emotion. In M. Lewis & J. M. Haviland (Ed.), *Handbook of Emotions* (173-191). New York: Guilford Press.
- Cannon, W. B. (1927). The James-Lange theory of emotion: A critical examination. *American Journal of Psychology*, 39, 106-124.
- Cardinal, R. N., Parkinson, J. A., Hall, J., and Everitt, B. J. (2002). Emotion and motivation: The role of the amygdala, ventral striatum, and prefrontal cortex. *Neuroscience and Biobehavioral Reviews*, 26, 321-352.
- Chalmers, D. (2005). The Matrix as metaphysics. In C. Grau (Ed.), *Philosophers Explore The Matrix*. New York: Oxford University Press.

- Chwalisz, K., Diener, E., & Gallagher, D. (1988). Autonomic arousal feedback and emotional experience: Evidence from the spinal cord injured. *Journal of Personality and Social Psychology*, 54, 820-828.
- Clore, G. L., Wyer, R. S., Jr., Dienes, B., Gaspar, K., Gohm, C., & Isbell, L. (2001), Affective feelings as feedback: Some cognitive consequences. In L. L. Martin & G. L. Clore (Eds.), *Theories of mood and cognition: A user's guidebook* (pp. 63-84). Mahway, NJ: Erlbaum.
- Craig, A. D. (2009). How do you feel — now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, 10, 59-70.
- Crane, T. (1998), Intentionality as the mark of the mental, in A. O'Hear (Ed.), *Contemporary Issues in the Philosophy of Mind* (229-251), Cambridge: Cambridge University Press.
- Crick, F., and Koch, C. (2003). A framework for consciousness. *Nature Neuroscience*, 6(2), 119-26.
- Dael, N., Mortillaro, M., and Scherer, K. R. (2011). Emotion expression in body action and posture. *Emotion*, doi:10.1037/a0025737
- Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Grosset/Putnam.
- Damasio, A. R. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. New York: Harcourt Brace & Company.
- Delancey, C. (2006), Basic moods, *Philosophical Psychology*, 19(4), 527-538.
- Deonna, J. A., and Scherer, K. R. (2010), The case of the disappearing intentional object: Constraints on a definition of emotion, *Emotion Review*, 2(1): 44-52.
- DeSteno, D., Petty, R. E., Wegener, D. T., & Rucker, D. D. (2000), Beyond valence in the perception of likelihood: The role of emotion specificity, *Journal of Personality and Social Psychology*, 69: 1052-1068.
- Dretske, F. (1988). *Explaining Behavior: Reasons in a World of Causes*. Cambridge, MA: MIT Press.
- Drevits, W. C., Savitz, J, and Trimble, M. (2008). The subgenual anterior cingulate cortex in mood disorders. *CNS Spectrums*, 13(8), 663-681.
- Dutton, D. G., and Aron, A. P. (1974). Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of Personality and Social Psychology*, 30(4), 510-517.
- Ekman, P. (1992), An argument for basic emotions, *Cognition and Emotion*, 6 (3/4): 169-200.
- Ekman, P. and Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17, 124-129.

- Ekman, P., Levenson, R. W., and Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, 221, 1208-1210.
- Ellsworth, P. C. & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, K. R. Scherer & H. H. Goldsmith (Ed.), *Handbook of Affective Sciences* (572-595). New York: Oxford University Press.
- Etkin, A., Egner, T., and Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends in Cognitive Science*, 15(2), 85-93.
- Fodor, J. A. (1987), *Psychosemantics: The problem of meaning in the philosophy of mind*, Cambridge: MIT Press.
- Folkow, B. (2000). Perspectives on the integrative functions of the 'symptho-adrenomedullary system'. *Autonomic Neuroscience: Basic and Clinical*, 83, 101-115.
- Forbes, G. (1997), How much substitutivity? *Analysis*, 57 (2): 109-113.
- Frijda, N. (1986). *The emotions*. New York: Cambridge University Press.
- Gendler, T. S. (2008) Alief and belief. *The Journal of Philosophy*, 105(10), 634-63.
- Goldie, P. (2002). *The emotions: A philosophical exploration*. Oxford: Oxford University Press.
- Goldman, A., and de Vignemont, F. (2009). Is social cognition embodied? *Trends in Cognitive Sciences*, 13(4), 154-9.
- Gordon, R. (1987). *The Structure of Emotion*. Cambridge, U.K.: Cambridge University Press.
- Griffiths, P. E. (1997), *What emotions really are*, Chicago: University of Chicago Press.
- Griffiths, P. E. (2004), Is emotion a natural kind? in R. C. Solomon (Ed.), *Thinking about feeling: Contemporary philosophers on emotion* (pp. 233-259), New York: Oxford University Press.
- Griffiths, P. E., and Scarantino, A. (2009). Emotions in the wild: The situated perspective on emotion. In P. Robbins & M. Aydede (Eds.), *Cambridge handbook of situated cognition* (pp. 437-453). New York: Cambridge University Press.
- Heims, H. C., Critchley, H. D., Dolan, R., Mathias, C. J., Cipolotti, L. (2004). Social and motivation functioning is not critically dependent on feedback of autonomic responses: neuropsychological evidence from patients with pure autonomic failure. *Neuropsychologia*, 42(14), 1979-1988.
- Hohmann, G. W. (1966). Some effects of spinal cord lesions on experienced emotional feelings. *Psychophysiology*, 3, 143-156.
- Isen, A. M. (1970). Success, failure, attention, and reaction to others: The warm glow of success. *Journal of Personality and Social Psychology*, 15(4), 294-301.

- James, W. (1922). What is an emotion? In K. Dunlap (Ed.), *The Emotions* (pp. 11-30). Baltimore: Waverly Press.
- Kenny, A. (1963), *Action, emotion and will*, London: Routledge & Kegan Paul.
- Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, 84(3), 394-422.
- Kriegel, U. (2011). Towards a new feeling theory of emotion. *European Journal of Philosophy*. doi: 10.1111/j.1468-0378.2011.00493.x
- Lange, C. G. (1922). The emotions. In K. Dunlap (Ed.), *The Emotions* (pp. 33-90). Baltimore: Waverly Press.
- Lau, H., and Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Science*, 15(8), 365-373.
- Lazarus, R. S. (1991). *Emotion and adaptation*. New York: Oxford University Press.
- LeDoux, J. E. (1996). *The Emotional Brain*. New York: Simon and Schuster.
- LeDoux, J. E. (2012). Rethinking the emotional brain. *Neuron*, 73(4), 653-676.
- Lerner, J. S., & Keltner, D. (2001), Fear, anger, and risk, *Journal of Personality and Social Psychology*, 81: 146-159.
- Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, 50(3), 249-258.
- Lindquist, K. A., Wager, T. D., Kober, H., Bliss-moreu, E., and Barrett, L. F. (2011). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences*, 17(4), 1-86.
- Millikan, R. G. (1993). *White Queen psychology and other essays for Alice*. Cambridge: MIT Press.
- Morris, M. W., & Keltner, D. (2000). How emotions work: The social functions of emotional expression in negotiations. *Research in Organizational Behavior*, 22, 1-50.
- Niedenthal, P. M., Barsalou, L. W., Wilkielman, P., Krauth-Gruber, S., and Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review*, 9(3), 184-211.
- Nussbaum, M. C. (2001), *Upheavals of thought: The intelligence of the emotions*, Cambridge, UK: Cambridge University Press.
- Philippot, P., Chappelle, G., & Blairy, S. (2002). Respiratory feedback in the generation of emotion. *Cognition & Emotion*, 16(5), 605-627.
- Prinz, J. J. (2004). *Gut reactions*. New York: Oxford University Press.

- Pylyshyn, Z. (1980). Computation and cognition: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences*, 3, 111-69.
- Rolls, E. T. (2005). *Emotion Explained*. New York: Oxford University Press.
- Royall, D. R., Lauterbach, E. C., Cummings, J. L., Reeve, A., Rummans, T. A., Kaufer, D. I., LaFrance, Jr., W. C., and Coffey, C. E. (2002). Executive control function: A review of its promise and challenges for clinical research. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 14, 377-405.
- Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of cross-cultural studies. *Psychological Bulletin*, 115, 102-141.
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110(1): 145-172.
- Sauter, D. A., Panattoni, C. and Happé, F. (2012). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology*, doi: 10.1111/j.2044-835X.2012.02081.x
- Schachter, S., & Singer, J. (1962). Cognitive, social and physiological determinants of emotional state. *Psychological Review*, 69(5): 379-399.
- Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion*, 23(7), 1307-1351.
- Schnall, S., Haidt, J., Clore, G.L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, 34 (8), 1096-1109.
- Schroeder, T. (2004). *Three Faces of Desire*, Oxford: Oxford University Press.
- Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275, 1593-1599.
- Schwarz, N., & Clore, G. L. (1983). Mood, misattribution and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, 45: 513-523.
- Searle, J. R. (1992). *The rediscovery of the mind*. Cambridge: MIT Press.
- Sequeira, H., Hot, P., Silvert, L., Delplanque, S. (2009). Electrical autonomic correlates of emotion. *International Journal of Psychophysiology*, 71(1), 50-56.
- Sinhababu, N. (2009). The Humean theory of motivation reformulated and defended., *Philosophical Review*, 118:4: 465-500.
- Sizer, L. (2000). Towards a computational theory of mood, *British Journal of the Philosophy of Science*, 51, 743-769.

- Smith, M. (1987). The Humean theory of motivation. *Mind*, 96, 36-61.
- Solomon, R. C. (1976), *The Passions*, New York: Doubleday.
- Solomon, R. C. (1988). On emotions as judgments. *American Philosophical Quarterly*, 25(2), 183-191.
- Soussignan, R. (2002). Duchenne smile, emotional experience, and autonomic reactivity: A test of the facial feedback hypothesis. *Emotion*, 2(1), 52-74.
- Stepper, S. & Strack, F. (1993). Proprioceptive determinants of affective and nonaffective feelings. *Journal of Personality and Social Psychology*, 64(2), 211-220.
- Strack, F., Martin, L.L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology*, 54(5), 768-777.
- Tangney, J. P., Miller, R. S., Flicker, L., and Barlow, D. H. (1996). Are shame, guilt, and embarrassment distinct emotions? *Journal of Personality and Social Psychology*, 70(6), 1256-1269.
- Wheatley, T., & Haidt, J. (2005), Hypnotically induced disgust make moral judgments more severe, *Psychological Science*, 16: 780-784.
- Whiting, D. (2011), The feeling theory of emotion and the object-directed emotions. *European Journal of Philosophy*, 19(2): 281-303.