

7-18-2008

## Is Core Affect a Natural Kind?

Brandie Martinez Bedard

Follow this and additional works at: [http://scholarworks.gsu.edu/philosophy\\_theses](http://scholarworks.gsu.edu/philosophy_theses)

---

### Recommended Citation

Martinez Bedard, Brandie, "Is Core Affect a Natural Kind?." Thesis, Georgia State University, 2008.  
[http://scholarworks.gsu.edu/philosophy\\_theses/42](http://scholarworks.gsu.edu/philosophy_theses/42)

This Thesis is brought to you for free and open access by the Department of Philosophy at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Philosophy Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact [scholarworks@gsu.edu](mailto:scholarworks@gsu.edu).

# IS CORE AFFECT A NATURAL KIND?

By

BRANDIE MARTINEZ BEDARD

Under the Direction of Dr. Andrea Scarantino

## ABSTRACT

In the scientific study of the emotions the goal is to find natural kinds. That is, to find categories about which interesting scientific generalizations and predictions can be formed. Core affect is dimensional approach to the emotions which claims that emotions emerge from the more basic psychological processes of valence (pleasant/unpleasant) and arousal (activation/deactivation). Lisa Feldman Barrett (2006b) has recently argued that the discrete emotion approach has failed to find natural kinds and thus should be dismissed as a failed paradigm. She offers core affect as an alternative theory that will better capture natural kinds in emotionally salient phenomena. In this thesis I evaluate Barrett's claim on the basis of a philosophically robust understanding of natural kinds and a careful assessment of the empirical evidence. I argue that while core affect is not a natural kind, subsets of core affect space may be natural kinds.

INDEX WORDS: Emotion theory, Emotions, Natural Kinds, Core Affect

IS CORE AFFECT A NATURAL KIND?

By

BRANDIE MARTINEZ BEDARD

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2008

Copyright by  
Brandie Martinez Bedard  
2008

IS CORE AFFECT A NATURAL KIND?

By

BRANDIE MARTINEZ BEDARD

Committee Chair: Andrea Scarantino

Committee: Eddy Nahmias  
George Rainbolt

Electronic Version Approved:

Office of Graduate Studies  
College of Arts and Sciences  
Georgia State University  
July 2008

## DEDICATION

I would like to dedicate this thesis to Dr. Sandra Dwyer whose support and encouragement throughout my time at Georgia State University has been integral to my success. I can never fully express how grateful and lucky I have been to have her in my life.

## ACKNOWLEDGEMENTS

I would like to acknowledge several people. First and foremost my parents who make all my dreams possible and have always been there for me. Second, I would like to acknowledge several colleagues and friends whose comments on drafts of this thesis were helpful: Ben Sheredos, Jenn Hudgens, Raleigh Miller, Dan Burnston, J. Aaron Brown, and Ron Myers.

## Table of Contents

ACKNOWLEDGMENTS	v
LIST OF TABLE	vii
LIST OF FIGURES	viii
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: CORE AFFECT AND ITS HISTORICAL ROOTS	3
2.1- Theoretical Frameworks in Emotion Theory	4
2.2- Core Affect Introduced	8
2.3- Are Valence and Arousal Well-Defined?	11
CHAPTER 3: THE CASE FOR CORE AFFECT	15
3.1- What is a Natural Kind?	16
3.2- Barrett's Case Against Discrete Approaches to Emotion	20
3.2.1- Do the Discrete Emotions Share Properties?	21
3.2.1.1- Facial Expressions (P <sub>1</sub> )	22
3.2.1.1.1- Automaticity	24
3.2.1.1.2- Panculturality	24
3.2.1.1.3- Primitivity	26
3.2.1.2- Semantic Structure (P <sub>2</sub> )	28
3.2.1.3- Physical Distinctness (P <sub>3</sub> )	29
3.2.1.4- Behaviors (P <sub>4</sub> )	31
3.2.2- Do the Discrete Emotions Share Causal Mechanisms?	32
3.3- Alleged Theoretical Advantages of Core Affect	34
3.3.1- Do Instances of Core Affect Share Properties?	36
3.3.1.1- Facial Expressions (P <sub>1</sub> )	36
3.3.1.1.1- Automaticity	37
3.3.1.1.2- Panculturality	38
3.3.1.1.3- Primitivity	39
3.3.1.2- Semantic Structure (P <sub>2</sub> )	40
3.3.1.3- Physical Distinctness (P <sub>3</sub> )	42
3.3.1.4- Behaviors (P <sub>4</sub> )	43
3.3.2- Do Instances of Core Affect Share a Causal Mechanism?	44
CHAPTER 4: ASSESSING CORE AFFECT	45
4.1- The Status of Core Affect as a Natural Kind	46
4.1.2- Is Valence a Natural Kind?	46
4.1.2.1- Shared Properties?	47
4.1.2.2- Shared Causal Mechanism?	51
4.1.3- Is Arousal a Natural Kind?	53
4.1.3.1- Shared Properties?	55
4.1.3.2- Shared Causal Mechanism?	56
4.1.4- Is Core Affect a Natural Kind?	58
4.1.5- Are Subsets of Core Affect Natural Kinds?	59
CHAPTER 5: CONCLUSION	62

LIST OF TABLE

Table 1

60

## LIST OF FIGURES

Figure 1: Core Affect Circumplex	10
Figure 2: Semantic-Based Circumplex	41

## CHAPTER 1: INTRODUCTION

Emotions are a central part of life. Emotional reactions represent the loss or gain of things we value, such as the thrill of victory or the disappointment of failure or grief at the loss of a parent. Emotions have long been thought of as passions defined in opposition to rationality (e.g. Plato *Republic*. 439d). At the same time emotions have also been proposed to explain a wide variety of phenomena that are important to rationality, such as memory (e.g. Damasio 1994). Emotions draw our attention to objects and we tend to have prominent memories about emotionally salient events. Emotions are seen as causes of diverse behaviors and psychological processes, as well as psychological disorders, such as anxiety disorders. Emotions are thus important psychological processes that are in need of explanation. Additionally, some argue that a scientific understanding of emotions is integral to any science of the mind or any explanation of consciousness (e.g. Barrett 2006b; Russell 2003; Chalmers 1999).

In recent years, there has been a significant amount of interest in the nature of the emotions from researchers in a wide variety of disciplines. Some of these include philosophers (e.g. Nussbaum 2001; Griffiths 1997; Prinz 2004b; Charland 2005), neuroscientists (e.g. Damasio 1999; Panksepp 1998), psychologists (e.g. Fridja 1986; Ekman 1999a), and more generally cognitive scientists (e.g. LeDoux 1996; Bradley and Lang 2000). One would expect that the result of this explosion of theoretical interest would be an accumulating body of knowledge and consensus over the nature of the emotions, but that has not been the case. Instead, the study of emotion has been hampered by disagreements about what the emotions are, how best to study them, and how best to interpret the existing empirical evidence.

Recently Lisa Feldman Barrett has argued that emotion theory has been hindered not so much by disagreement, but by the wide acceptance of the notion that discrete emotions, such as

*anger, fear, and sadness*, are natural kinds (2006b).<sup>1</sup> Natural kinds are the holy grail of scientific investigation because they allow scientists to form generalizations and predictions about the given category. But Barrett claims that the empirical evidence does not warrant the assumption that the discrete emotions are categories that are apt for generalizations and predictions. Instead she argues that the scientific study of emotion should focus on more fundamental processes, rather than these broad categories. Barrett presents core affect as a theory that focuses on these fundamental processes. Core affect is a dimensional, neuropsychological category corresponding to levels of valence (pleasant /unpleasant) and arousal (activation /deactivation). Core affect theorists, such as Lisa Feldman Barrett and James A. Russell claim that emotions emerge from this neuropsychological category and Barrett further argues that “the empirical case supporting the hypothesis that core affect is a natural kind is suggestive” (2006b, 48).<sup>2</sup>

In this paper I will evaluate the empirical case for Barrett’s hypothesis on the basis of a more philosophical understanding of natural kinds and a careful assessment of the evidence. Paul Griffiths has pointed out that “questions about the nature of the emotions cannot be answered in the armchair alone but must be sought in part by empirical investigation of emotional phenomena” (1997, 1). I will do just that. I will argue that the empirical evidence supports what Carl Craver (2004) has called “kind splitting.”<sup>3</sup> In other words, splitting the proposed kind (core affect) is more likely to yield the kinds about which inductive generalizations and predictions can

---

<sup>1</sup> Emotion terms that refer to linguistic forms will be indicated by quotations, while emotion categories or the discrete emotions will be indicated by italics. For example, “joy” refers to the word itself as used by the folk and *joy* refers to the discrete emotion.

<sup>2</sup> Being a good scientist, Barrett emphasizes that “the crucial experiments have yet to be done” and that we should withhold judgment until further empirical evidence is collected (48).

<sup>3</sup> Craver defines kind splitting as “a common (but rarely explicit) methodological assumption in neuroscience and elsewhere that discovering a kind to be dissociably realized [cases where a proposed kind is found to be realized by more than one distinct and independent realizer] mandates splitting the kind into as many as there are dissociable realizers” (960). The realizers vary according to the explanatory role they play in the theory. In the case of neuroscience and, arguably, cognitive science the realizers are often thought of as brain areas and/or scientifically relevant properties.

be formed. I will attempt to show that four potential candidates for natural kindhood can be located in subsets of the core affective category. Natural kinds are the primary subject of scientific research as well as much philosophical analysis of the emotions. Therefore, the discovery of four potential natural kinds in the core affective category contributes to future scientific research and philosophical analysis of the emotions.

In chapter two, I will discuss the major theoretical frameworks at play in current emotion theory, introduce core affect theory, and examine whether or not the dimensions of valence and arousal are well-defined. In chapter three, I will begin by explaining Richard Boyd's theory of natural kinds (1991; 1999), which Barrett assumes for her analysis (2006b). I will then reconstruct Barrett's argument that discrete emotions are not natural kinds and examine her argument in favor of the natural kind status of core affect. In chapter four I will assess the evidence to see if, in fact, core affect is a natural kind. I will argue, *contra* Barrett, that the evidence in favor of core affect being a natural kind is not suggestive. This being said, there may be a sense in which four subsets of the broad core affective category may be natural kinds. I will conclude with what I think core affect can add to emotion theory and a general lesson for emotion theorists looking for natural kinds: splitting proposed kinds is the best avenue for discovering natural kinds.

## **CHAPTER 2: CORE AFFECT AND ITS HISTORICAL ROOTS**

J. Russell defines core affect as “a neurophysiological state that is consciously accessible as a simple nonreflective feeling that is an integral blend of hedonic (pleasure-displeasure) and arousal (sleepy-activated) values” (2003, 147). He argues that valence and arousal have played a

fundamental role in theories of emotion, going back to the beginning of psychology.<sup>4</sup> Core affect is proposed as an alternate theoretical paradigm to the dominant paradigm in current emotion theory, the discrete emotion approach. Since core affect is hailed by Barrett and Russell as a better theory of the emotions, it is important to begin by discussing what a good theory of the emotions should look like.

Roughly speaking, a good theory of the emotions should have explanatory and predictive power. It should tell us why a given emotion is present, predict when an emotion will be present, and do so better than alternative theories. The theory should be applicable to emotions in all individuals that experience them and apply to all cultures in which emotions are exhibited. In this chapter, I will begin with a general outline of the theoretical frameworks in emotion theory and explain core affect theory in more detail. This will prepare us to compare and contrast different approaches to the study of emotions in terms of their explanatory and predictive power.

## **2.1- Theoretical Frameworks in Emotion Theory**

A fundamental question in emotion theory is whether emotions are best studied as broad dimensions or as discrete units (such as the *surprise* unit, the *sadness* unit, etc.). The former approach finds its origins in Wilhelm Wundt (1873; 1897), while the latter finds its origins in William James (1884).<sup>5</sup> Ever since, there has been a tension between dimensional and discrete approaches over which theoretical framework best captures the emotions.<sup>6</sup> Core affect theorists adopt a dimensional framework and claim that their approach is scientifically more fruitful than the discrete emotions framework. Discrete emotion theorists disagree and argue that emotions

---

<sup>4</sup> According to Russell, early psychologists like Wilhelm Wundt, Carl Stumpf, and Edward Bradford Titchner all posited something similar to core affect. In particular Wundt “identified dimensions of pleasant-unpleasant, tension-relaxation, and excitement-calm as the basis of feeling and emotion” (Russell 2003, 153).

<sup>5</sup> Wundt and James are considered the fathers of modern psychology.

<sup>6</sup> See Izard and Azkerman (2000) for a discussion of the historical debate and tensions that have arisen between dimensional and discrete emotion theorists.

should be studied as discrete units. In what follows I will briefly describe these two approaches in their contemporary forms.

The discrete emotions approach starts with folk terms for the emotions, such as “happiness,” “disgust,” “contempt,” and so on, and attempts to construct theories about what the folk emotion terms refer to in the world. Discrete emotion theorists claim that there are a small number of folk emotions (for example, “fear,” “anger,” “joy,” and “sadness”) that can be scientifically and theoretically identified and distinguished from one another. The underlying assumption is that some folk emotion terms refer to real distinctions in the world, or at least they partially refer to them, and can be precisified so as to be useful in scientific theories.<sup>7</sup> For example, the folk term “sadness” refers to a discrete emotion, *sadness*, because there are unique properties associated with it, such as distinct contractions of facial muscles and signature bodily responses. Moreover, the instantiation of these properties may be due to a common cause, assumed to be in the brain, such as activity in the subcallosal cingulate. According to this approach, the discrete emotions are the proper focus of scientific studies.

Discrete emotion theorists argue over the number of discrete emotions and postulate different theories about how to theoretically and scientifically identify these emotions. Three examples of discrete emotion theories are cognitivism (e.g. Arnold, 1960; Solomon, 2004), neo-Jamesianism (e.g. Damasio, 1999; Prinz 2004c), and basic emotions or affect program theory (e.g. Ekman, 1999a; Griffiths, 1997). It is beyond the scope of the present thesis to elaborate on the nuances of these theories. What follows is a brief sketch of these three discrete emotion theories. The goal is to provide the reader with a preliminary understanding of the discrete emotion approach.

---

<sup>7</sup> It is important to keep in mind that discrete emotions do not equate to folk emotion terms. The discrete emotion approach begins with these terms, but may need to eliminate certain properties associated with them in order to find a precise scientific term.

Cognitivism claims that discrete emotions can be theoretically and scientifically identified and distinguished from each other by the way an object is judged or appraised. The basic slogan of the research program is: “Emotions are appraisals or value judgments” (Nussbaum 2001, 4). The psychologist Magda Arnold, for instance, claimed that the differences between discrete emotions can be explained by an organism’s direct, intuitive, and immediate appraisal of environmental changes that affect their well-being (1960).<sup>8</sup> For example, *fear* may be identified by the appraisal that an object, like a tiger, is dangerous. *Grief* may be identified by the appraisal that an important object, like a spouse, has been lost, and *anger* by the appraisal that someone has been mistreated.

Neo-Jamesianism is a second example of a discrete emotion theory. Neo-Jamesians argue that discrete emotions are scientifically and theoretically identified and distinguished by complex changes in the body in reaction to appraisals of emotionally relevant stimuli. According to Jesse Prinz, “emotions are perceptions (conscious or unconscious) of patterned changes in the body (construed inclusively)” (2004c, 45). These bodily changes are correlated with activations in brain areas associated with bodily perception, such as the somatosensory cortex.<sup>9</sup> For example, *disgust* may be identified by the sensation that one’s “skin is crawling” and nausea. *Fear* may be identified by a rise in body temperature, resulting in sweating, and the “freezing” of one’s limbs, and *anger* may be identified by an increase in heart rate and contractions in the brow area.

---

<sup>8</sup> Arnold argued that every organism is constantly appraising the environment around them in order to identify things that are helpful or harmful, easy or hard to avoid or approach, and which objects are present or absent. However, organisms do differ in their level and ability of appraisal. While Arnold’s theory posited discrete emotions as the entities that best explain emotional phenomena, many of her intellectual predecessors posited dimensions instead. Most current appraisal theories, at least in the natural sciences, assume a dimensional approach, although some, such as Lazarus (1991), have argued for a hybrid theory. See Ellsworth and Scherer (2003) for a detailed discussion of appraisal theories.

<sup>9</sup> The somatosensory cortex is an area of the brain that receives information about general sensations, such as touch, pain, pressure, temperature, and proprioception (the sensation of joint and muscle position).

A third influential discrete emotion theory is affect program theory (or basic emotions theory), arguably the dominant non-dimensional approach in contemporary emotion theory. The fundamental claim is that discrete emotions are distinguished and identified by a set of complex responses to the emoter's environment that are modular (Fodor, 1983)<sup>10</sup> and evolutionarily determined.<sup>11</sup> According to Ekman, "the primary function of emotion is to mobilize the organism to deal quickly with important interpersonal encounters, prepared to do so by what types of activity have been adaptive in the past [as a species and as individuals]" (1999b, 46). A few of the responses associated with the discrete emotions by affect program theorists are pan-culturally recognized facial expressions, distinctive physiology, quick onset, and presence in non-humans, such as primates.<sup>12</sup> Ekman has argued that at least *fear*, *anger*, *sadness*, *disgust*, *joy*, and *surprise* are affect programs or basic emotions in this sense. For example, *surprise* may be identified by a distinct facial expression that involves the raising of the eyebrows, increased heart rate, and rapid onset, and that these responses are present in non-humans.

What these three discrete emotion theories share are two assumptions about how scientists should study the emotions. First, the assumption that emotions are best captured in terms of discrete units (e.g. the *fear* unit, the *anger* unit, and so on) and second, the assumption that folk terms are the natural starting point for any scientific investigation of the emotions. What differentiates these theories, on the other hand, is how to *identify* scientifically fruitful discrete

---

<sup>10</sup> A module on Fodor's account is "domain specific, innately specified, hardwired, autonomous, and not assembled" (1983, 37). One of the main points of a modular is that they are informationally encapsulated (i.e. the module cannot access any mental process or information outside of itself, only information in the specific module can be used, and this information can contradict information stored in other modules). Griffiths has argued that this is the most important aspect of the basic emotion theory (1997).

<sup>11</sup> There are a variety of different versions of the basic emotion theory (e.g. Ekman, 1999b; Fridja, 1986; Izard, 1977; Plutchik, 1980; Panksepp, 1998), but most agree with this fundamental claim. These accounts diverge on issues such as the number of basic emotions and which of the responses are needed to pick out particular basic emotions. See Ortony and Turner (1990) for a detailed discussion of the differences between these theories. I will focus on Ekman's version here because, arguably, he was the first to propose it and provide empirical support for it (1971; 1992; Ekman et al., 1983).

<sup>12</sup> For a list of the characteristic responses see table 3.1 in Ekman (1999b, 56).

emotions. For example, cognitivists claim that *fear* is identified by the judgment that an object is dangerous. Neo-Jamesians claim that *fear* should be identified by the perception of complex bodily changes, such as sweat and increased heart rate. Affect program theorists or basic emotion theorists claim that *fear* should be identified by a complex set of reactions to the emoter's environment that are evolutionarily determined, such as involuntary physiological reactions, the production of a suite of distinct facial expressions, characteristic behavioral responses, and so on. This completes my introduction of the discrete emotions approaches in emotion theory. Let us now turn to the dimensional approach, of which core affect theory is the most prominent example.

In the current scientific study of emotions, dimensional theorists believe that we should focus on the underlying *constituents* of the discrete emotions. All dimensional theories share one assumption, namely that all discrete emotions share more basic psychological properties. It is argued that emotions emerge from these more basic properties and that focusing on these properties is scientifically more fruitful than focusing on discrete emotions. According to dimensional theorists, discrete emotion categories do not represent independent phenomena to be explained and identified one by one. Rather, emotion categories are higher-order phenomena which emerge from changes along continuous dimensions of lower-order properties. These properties are commonly defined by two dimensions, although a dimensional theory can posit as many dimensions as is needed in order to achieve explanatory and predictive power. Core affect posits valence and arousal as the key dimensions. In the next section I will elaborate on these.

## **2.2- Core Affect Introduced**

As I previously mentioned, core affect is a neuropsychological category corresponding to combinations of valence and arousal levels. Valence is broadly understood as hedonic tone

(positive/negative or pleasant/unpleasant) and arousal is broadly understood as activation/deactivation of the mind and body. For example, “joy” is positively valenced and highly arousing and “disgust” is negatively valenced and mildly arousing. Core affect is not just a theory about the emotions, but about all affective states.<sup>13</sup> As Russell (2003, 146) puts it:

The scope of the proposed framework is broader than emotion (including states such as comfort, serenity, drowsiness, and lethargy). Gone is the assumption that all events called *emotion* or *fear* or *anger* can be accounted for in the same way. These concepts are not abandoned but are put in their proper place as folk rather than scientific concepts, and their role limited to whatever role folk concepts actually play in emotion (and in perception of emotion in others).<sup>14</sup>

Core affect theorists claim that emotions “emerge from” affective states and that affective states are best described as combinations of valence and arousal values (Barrett 2007a). These states of valence and arousal endow the emoters with a kind of “core knowledge about whether objects or events are helpful or harmful, rewarding or threatening, calling for acceptance or rejection”

(Barrett 2007a, 377). Here is Barrett’s (2007a, 377) definition of core affect:

Information about the external world... indicates whether an object or situation is helpful or harmful, rewarding or threatening, requiring approach or withdrawal. With awareness, core affect is experienced as feelings of pleasure or displeasure that are to some extent arousing or quieting.

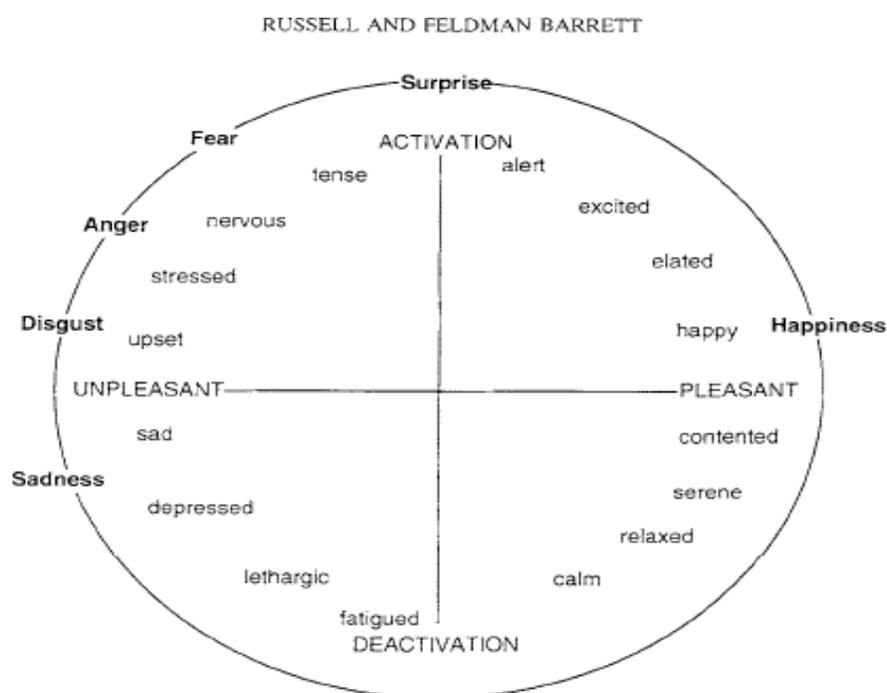
The clearest way to conceive of the core affective category is in terms of the so-called “core affect circumplex.” Broadly speaking, a circumplex is a visual representation of certain aspects of a theoretically defined psychological space, aiming to capture certain psychological

---

<sup>13</sup> Nico H. Fridja (2000) gives the best explanation of the term “affective” that I have seen. He explains that salient events (events associated with strong feelings) often intrude upon goal-directed behavior and thought and at times even interrupt it. Salient events can also elicit unplanned thought and behavior. These sort of salient feelings or events “affect” the person. According to Fridja, the predecessors to the notion of affect are the Greek *pathema*, the Latin *affectus*, and the early French and English “passion.” These terms all indicate a form of behavior that is in some sense passive, which differs from “action,” and the terms contrast affect and passion. Salient events intrude upon goal-directed behavior and may elicit desires, thoughts, plans, and behaviors that persist over time. Some of these behaviors may even be performed regardless of the cost, moral objections, and external obstacles. According to Fridja, this is the more modern sense of passion- the behaviors, desires, and thoughts that suggest urges with considerable force. Although, core affect is a broad category that encompasses all affective states, in this paper I will typically focus on the emotions. The generality of core affect will be discussed where relevant.

<sup>14</sup> Emphasis in original.

phenomena. In the case of core affect, the psychological space is represented by two axes: valence and arousal. Valence is plotted along the horizontal axis and arousal is plotted along the vertical axis. Below is the core affect circumplex:



*Figure 1.* The inner circle shows a schematic map of core affect. The outer circle shows where several prototypical emotional episodes typically fall. Modified from Feldman Barrett and Russell (1998). Copyright 1998 by the American Psychological Association.

In order to map the affective space portrayed above, core affect theorists ask individuals to make judgments about the relationships between valence and arousal for various affective terms using surveys and self-reports. For instance, they ask subjects a number of questions about the levels of valence and arousal for terms like “fatigued,” “contented,” “happy,” and “stressed” and then chart how individuals group the concepts of valence and arousal. Just to give an example, subjects consistently rate “depressed” and “fatigued” as low activation and negative valence. The theorists then take the information from the surveys and using statistical techniques form the circumplex and plot the affective space (see Barrett 1996). I will return to the evidence that supports the core affective circumplex in section 3.3.1.2.

Since core affect theorists claim that emotions emerge from these more basic psychological states of valence and arousal, the prototypical emotions are supposed to map onto the circumplex, as shown above. For example, “fear” and “anger” are commonly rated as both being highly arousing and negatively valenced, so they will be placed in the top left quadrant. “Happiness” is commonly rated as both being positively valenced and highly arousing, so it is placed in the top right quadrant. It is important to keep the circumplex in mind as it is a visual representation of the core affect category. Since core affect is defined as combinations of valence and arousal, it is important to figure out what core affect theorists mean by these terms. This is the task for the next section.

### **2.3- Are Valence and Arousal Well-Defined?**

As it turns out, there are multiple definitions put forth for both valence and arousal by core affect theorists, some of which may not be compatible with each other. I do not argue that these definitions cannot be made compatible, but rather suggest that doing so would require clarifications of deep methodological issues. A thorough discussion of this issue, however, lies outside the scope of this thesis. What I need is a clear, yet admittedly provisional understanding of valence and arousal, which I will then adopt in the subsequent chapters.

The first dimension I will examine is arousal, which is characterized as a continuum from a state of maximal activation to a state of maximal deactivation (visually represented by the vertical axis of the circumplex). “Activation is a general mobilization in preparation for vigorous action” and “the feeling [of arousal] is one’s sense of mobilization and energy” (Russell 2003, 150 and 148). The intuition behind positing arousal as a dimension is that all emotions seem to *vary* in arousal. For example, “anger” can range from mild irritation (low arousal) to “rage” (high arousal). It is hard to imagine an emotion that does not vary in arousal and so it appears

that arousal is central to emotion. Additionally, arousal appears to be an integral component in explaining an emoter's behavioral reactions. Organisms appear to mobilize resources for actions according to the level of intensity of an occurrent emotion. "Rage" mobilizes a high amount of energy so that an individual is ready to fight and refocuses the subject's attention almost exclusively on the eliciting stimuli. Mild irritation, on the other hand, mobilizes a low amount of energy that does not necessarily prepare an individual to fight, and does not refocus attention exclusively.

Arousal can be explained in at least two ways, as physiological arousal and felt arousal. Physiological arousal essentially involves the body and is measured by reactions in the autonomic nervous system.<sup>15</sup> The body is physiologically aroused when running, for example, and deactivated when sleeping. Felt arousal is similar, but does not necessarily involve the body although it often does. A person may feel aroused when they are anxious, but that feeling may or may not correspond to something that can be physically measured.

Barrett seems to combine these two definitions and defines arousal as "feeling as if the mind or body is active, as in aroused, attentive, or wound-up, versus feeling that the mind or body is still, as in quiet, still, or sleepy. *Felt activation is typically related to, but does not have a one-to-one correspondence with, actual physiologic activity*" (2007a, 379).<sup>16</sup> Clearly one component of what is "felt" in felt arousal is the bodily changes involved in physiological arousal, but "felt" also indicates feelings that the mind is aroused. To summarize, the dimension

---

<sup>15</sup> The autonomic nervous system consists of two contrasting neuron pathways: the parasympathetic (controls the body when at rest) and the sympathetic (controls the body in "flight-or-fight" situations). The autonomic nervous system is not under voluntary control (and so is thought to be a good measure of the effects of the emotions since it eliminates socially learned reactions), is controlled by the hypothalamus, and effects contractions of smooth muscles and glands. Measurements of the autonomic nervous system include heart rate, blood pressure and flow, respiration, and movement of the eyes. The scientific assumption is that correlating autonomic nervous system reactions with categories provides evidence for an underlying causal mechanism. See Barrett 2006b for more on this assumption.

<sup>16</sup> Emphasis added.

of arousal is defined both as felt arousal and physiological arousal with an emphasis placed on felt arousal.

There are numerous philosophical and scientific problems with defining arousal in terms of felt arousal. On the scientific side, it is difficult to see how a scientist would measure felt arousal. According to Barrett this can be accomplished through “cross-sectional studies examining how...participants self-report their experiences using common... emotion words [and these self-reports] often give evidence of arousal-based content” (2007a, 379). Barrett has argued that self-reports do tell us about the phenomenal experience of subjects and has codified a method to access the phenomenological feel of core affect (1996; 1997; 2004a; 2007a). But it is questionable whether self-reports can capture the phenomenology of subjective experience. It is beyond the scope of the present thesis to discuss the issue in detail, but one of the central objections to such a method is that self-reports essentially involve introspection on the part of the subject, which may alter the phenomena in question.<sup>17</sup> In the subsequent chapters I will adopt the physiological notion of arousal and leave behind these problematic claims about felt arousal. I will return to my provisional notion of arousal at the end of this section.

The second dimension of core affect is valence. Valence is characterized as a continuum from a state of maximal pleasantness to a state of maximal unpleasantness (visually represented by the horizontal axis of the circumplex). “Pleasure-displeasure (or valence) is a dimension of experience that refers to hedonic tone” (Barrett 1999b, 10). Intuitively all emotions seem to be either pleasant or unpleasant. For example, “fear” seems unpleasant whereas “joy” seems pleasant and “sadness” seems unpleasant whereas “happiness” seems pleasant.

---

<sup>17</sup> For a more detailed philosophical discussion of the problems and virtues of self-reports see Chalmers 1998 and 2004.

The following is Barrett's more detailed explanation of valence: "the valenced aspect of core affect has been called many things- hedonic tone, utility, good-bad, mood, pleasure-pain, approach-withdrawal, rewarding-punishing, appetitive-aversive, positive-negative- but the similarity is clear" (2006d, 40). Philosophers have argued that the similarities are not clear and that the term "valence" is ambiguous (e.g. Prinz, 2004a; Charland, 2005). To explore this alleged ambiguity I will look at two of these notions: positive-negative and approach-withdrawal. The positive-negative notion of valence is often explained in terms of pleasant and unpleasant *feelings*. For example, "joy" is a positive emotion in which pleasant *feelings* are involved and "fear" is a negative emotion in which unpleasant *feelings* are involved. On the other hand, the approach-withdrawal notion of valence is often explained in terms of approach and withdrawal *behaviors*. For example, "love" elicits the tendency to approach the person that invokes the positive emotion and "disgust" elicits the tendency to withdrawal or stay away from the person who invokes the negative emotion. The first notion fundamentally involves a feeling aspect, while the second can be explained in terms of behaviors.

Defining valence as fundamentally involving feelings or behaviors radically changes the types of experiments that can be done to study valence and the scope of the conclusions that can be drawn. The feeling aspect limits the subjects involved in the experiment to those who possess language, while defining valence in terms of behaviors widens the range of potential experimental subjects and thus the conclusions that can be drawn. For example, if a researcher uses the notion of pleasant and unpleasant feelings, then the method will, in all likelihood, be to use self-reports to try and access these feelings. This methodology excludes infants and non-humans. On the other hand, if a researcher uses the notion of approach-withdrawal, then a number of different methods could be used and some methods will be unacceptable if used on

humans. For example, cocaine could be used to elicit approach behaviors and electric shocks to elicit withdrawal behaviors. In this case valence includes non-humans and there may be interesting homologous structures of the mammalian brain involved in valence. So how the experiment is conducted and the scope of the conclusions are affected by the notion of valence that is adopted.<sup>18</sup> Although I will use the terms “positive valence” and “negative valence” in subsequent chapters, I will not adopt the feeling notion of valence. For the purposes of this thesis, “valence” is to be understood in the behavior sense described above.

To summarize, both valence and arousal are mongrel concepts. I have argued that arousal is defined both in terms of feelings and physiological responses and valence is defined both in terms of feelings and behaviors. I have tried to show above that the adoption of a phenomenological notion of valence and arousal are problematic. The core affect category would be much clearer, more useful and less objectionable if these phenomenological claims were left behind. At least until there are advances in measuring phenomenological data and the methodological issues are satisfactorily addressed. I do not mean to revert back to behaviorism and I acknowledge that an important aspect of the emotions is that they feel a certain way, but given the issues involved in recording subjective experience this data cannot be included in the assessment of core affect as a natural kind. Thus, in the subsequent chapters I will focus solely on the physiological and behavioral notions. For the purposes of my thesis, I will understand valence and arousal, and consequently, core affect, as essentially behavioral and physiological.

### **CHAPTER 3: THE CASE FOR CORE AFFECT**

Now that we have at least a preliminary understanding of core affect, we are in a position to assess Barrett’s arguments against discrete emotions and in favor of core affect (2006b). Barrett

---

<sup>18</sup> I am not alone in thinking that this is problematic. See Panksepp (2007) for a further discussion of this problem.

argues that the empirical evidence does not support the claim that there exist discrete emotion categories about which scientists could formulate scientifically interesting inductive generalizations. To put it briefly, she thinks that discrete emotion categories are not natural kinds. On the other hand, she argues that the empirical evidence suggests that core affect may be a natural kind. In the first section of this chapter I will explain Richard Boyd's notion of natural kinds to ground my discussion of Barrett's arguments in a philosophically robust notion of natural kinds. In the second and third sections I will reconstruct Barrett's arguments in some detail.

### **3.1- What is a Natural Kind?**

There is a long-standing debate in philosophy about natural kinds (for a sample of recent discussions on natural kinds, see Hacking 1991; Boyd 1999; Millikan 1999; LaPorte 2003; Griffiths 2004b). The central idea is that natural kinds "carve nature at its joints" (Plato *Statesman* 287c) and thus provide us with fundamental ways to classify objects and study their properties. The notion of a natural kind is essentially tied up with the problem of induction because a natural kind singles out a category about which scientific generalizations can be formed (Machery, 2005; Mill, 1843; Quine, 1969; Boyd, 1991, 1999; Hacking, 1991). Natural kinds are contrasted with arbitrary groupings of things for which only a few or no generalizations can be formed. For example, species such as "whales" and chemical elements such as "oxygen" are natural kinds, but collections of jewels and things that weigh more than 50 pounds are arbitrary groups. Members of a natural kind must share, non-accidentally, a number of theoretically and scientifically important properties suitable to produce generalizations about them. These shared properties mark natural kinds as "the building blocks of scientific generalizations" (Machery 2005, 446). Since scientific generalizations are a goal of most

empirical sciences, scientists aim to develop empirical theories that identify natural kinds in their domain.

Given this general notion of a natural kind, philosophers have proposed many different ways to characterize them (e.g. Locke, 1690; Venn, 1886; Russell, 1948; Mill, 1843; Kripke, 1980; Boyd, 1991). The most influential recent notion of natural kinds comes from Richard Boyd (1991; 1999). Boyd advocates a *causal* notion of natural kinds where a kind is natural insofar as it constitutes a maximal class whose instances share non-accidentally many, but not all, scientifically relevant properties.<sup>19</sup> He proposes that this notion of natural kinds is captured by what he terms a “causal homeostatic property cluster.” Boyd writes:

I argue that there are a number of scientifically important kinds (properties, relations, etc.) whose natural definitions are very much like the property cluster definitions postulated by ordinary-language philosophers except that the unity of properties in the defining cluster is mainly causal rather than conceptual. The natural definition of one of these *homeostatic property cluster kinds* is determined by the members of a cluster of often co-occurring properties and by the (“homeostatic”) mechanism that brings about their co-occurrence.<sup>20</sup>

The members of a homeostatic property cluster kind do not have to share *all* the same properties. Instead, Boyd proposes that what matters is that the imperfectly shared properties are shared non-accidentally due to a common causal mechanism. On the basis of (i) the imperfectly shared cluster of properties and (ii) a common causal (homeostatic) mechanism that underlies the properties, we can, respectively, identify instances of a natural kind, and form inductive generalizations about the kind.

---

<sup>19</sup> Boyd claims that his theory of natural kinds is independent of any general metaphysical realist claims (1991). He argues that both realists and empiricists should accept his theory. Realists will accept that natural kind terms are projectable (provide explanations and predictions) because these terms refer to real causal structures in the world. Boyd also believes that empiricists should accept that natural kind terms are projectable because there are empirically adequate background theories that allow us to construct projectable terms. It is beyond the scope of the present paper to analyze this claim, but see Griffiths (1997) for more on this argument.

<sup>20</sup> Boyd (1991), 141. Emphasis in original

According to Boyd, the paradigmatic example of a natural kind is “species” because species specify “categories apt for induction and explanation in [biological] science” (1991, 142). The natural kind “species” constitutes a homeostatic property cluster because the fact that species members (imperfectly) share morphological, physiological, and behavioral properties is caused by the mechanism of natural selection (the exchange of genetic materials) underlying the property cluster.<sup>21</sup> In other words, members of a species do not have to share necessary and sufficient properties for being “species” members. In fact, Boyd believes that there will necessarily be indeterminacy in the extension of species because speciation depends on populations that are intermediate between the parent species and the emerging new species. Boyd (1991, 142) writes:

Any ‘refinement’ of classification which artificially eliminated the resulting indeterminacy in classification would obscure the central fact about heritable variations in phenotype upon which biological evolution depends and would be scientifically inappropriate and misleading.

Since “species” is a natural kind, it seems that some natural kinds may be counterexamples to “essentialist” theories that require necessary and sufficient conditions of membership for natural kinds (e.g., Kripke 1980). Additionally, the species example demonstrates that the notion of a homeostatic property cluster captures the way at least some paradigmatic natural kinds are used.

This being said, Ruth Millikan (1999) has offered a different notion of natural kinds. According to Millikan, at least some “kinds are defined by reference to historical relations among members, not, in the first instance, by reference to properties” (54). On her view, natural kinds are best characterized by the historical relations between members of the kind and not the properties that the members share due to a causal mechanism. Although Millikan’s notion is different than Boyd’s, it is not clear if Millikan’s view is a critique of Boyd’s theory. She admits

---

<sup>21</sup> Although there are debates, both in philosophy and biology, about how to define a species, most agree that the exchange of genetic material is essential to speciation.

this in a footnote: “I am not clear whether the next few paragraphs are best read as exegesis of Boyd or as criticism” (*Ibid.*, 63). Additionally, Boyd argues that Millikan’s ‘historical kinds’ form a “sub-set of the HPC kinds” (Boyd 1999, 84). In other words, Boyd believes that Millikan’s notion of natural kinds is reducible to his theory. I will not get into the details of this debate and assume Boyd’s notion of natural kinds in what follows.

There are four reasons why I will assume Boyd’s notion of natural kinds. First, Boyd’s notion of natural kinds is widely accepted by many philosophers of emotion (e.g. Griffiths, 1997; Prinz, 2004a; Charland, 2002) and emotion scientists (e.g. Barrett, 2006b; Ekman, 1999b; Izard, 2007). Second, Griffiths (1997) has argued that defining a natural kind by a homeostatic property cluster is ideal for the blurry boundaries of the emotions. “Emotion” is a notoriously difficult term to define in terms of necessary and sufficient conditions. On Boyd’s account this need not preclude research on the emotions- it is even a live possibility that “emotion,” like “species” *must* exhibit indeterminacy. Third, most emotion scientists, including Barrett, are concerned with classifying emotions in terms of homologues (shared ancestry) rather than analogues (similar function). Since the concern is with finding shared ancestry rather than similar function, members will not necessarily need to share all the properties.

Fourth, adopting Boyd’s theory of natural kinds allows scientists to form testable hypotheses and evaluate the evidence for these hypotheses. In other words, scientists can form a hypothesis that X is a natural kind and test this hypothesis. To test the hypothesis, researchers look for a cluster of observable properties with a common underlying causal mechanism. If X meets these requirements, then X is a natural kind about which scientifically interesting inductive generalizations can be made. The ability to form testable hypotheses, arguably, is what has made Boyd’s notion of a natural kind so attractive to scientists, like Barrett who explicitly cites Boyd.

Barrett's arguments are best understood in light of Boyd's theory because her arguments involve a number of scientifically relevant properties and an underlying causal mechanism. These are the criteria that she presents for natural kindhood.

In the next sections of this chapter, I will reconstruct Barrett's argument that discrete emotions are not natural kinds. Then I will lay out her argument that the empirical evidence for core affect as a natural kind is suggestive. In order to make the arguments as clear as possible I propose a compact summary of Boyd's core idea and requirements for natural kindhood:

*Core Idea:* A kind is natural if and only if its members tend to share a family of co-occurring properties and a causal (homeostatic) mechanism that brings about their co-occurrence non-accidentally.

In light of this account, there are two requirements a kind K must satisfy in order to qualify as natural:

*Requirement 1:* There is a family of often co-occurring properties  $P_1 \dots P_n$  associated with instances of K.

*Requirement 2:* There is a causal (homeostatic) mechanism that brings about the family of often co-occurring properties seen in instances of K.

Are discrete emotions natural kinds under this view? Is core affect a natural kind? Barrett provides a negative answer to the first questions and a positive answer to the second question. I will argue in chapter four that core affect is not a natural kind, but that subsets of core affective space may be natural kinds. Let us consider her arguments in some detail.

### **3.2- Barrett's Case Against Discrete Approaches to Emotions**

Some discrete emotion theorists, for example Ekman (1999a; 1999b) and Panksepp (1998; 2000), have argued that discrete emotions are natural kinds *sensu* Boyd. The idea is that *fear, anger, joy, sadness*, and so on are natural kinds. Barrett is convinced that the empirical evidence does not support the claim that discrete emotions are natural kinds. Barrett concludes that the discrete emotions approach has failed to find natural kinds and therefore should be

abandoned. She presents core affect as an alternative paradigm that will be more successful in finding natural kinds. In what follows I will reconstruct Barrett's argument that the discrete emotions do not fulfill requirements 1 and 2. In evaluating Barrett's arguments, I will focus solely on the disconfirming evidence that she presents and not on the arguments given by discrete emotion theorists.

### 3.2.1- Do the Discrete Emotions Share Properties?

There are four shared properties commonly mentioned by discrete emotion theorists (especially affect program theorists) as substantiating the view that discrete emotions are natural kinds. These are (P<sub>1</sub>) automatic,<sup>22</sup> pan-cultural,<sup>23</sup> and primitive<sup>24</sup> signature facial expressions for each emotion, (P<sub>2</sub>) the semantic structure of discrete emotion terms as indicated by self-reports, (P<sub>3</sub>) physiological distinctness, and (P<sub>4</sub>) behavioral distinctiveness. Barrett argues that there is no compelling evidence that instances of discrete emotion categories share these properties. The consequence is that Barrett thinks, *requirement 1*, is not met: There is no family of often co-occurring properties P<sub>1</sub>...P<sub>n</sub> associated with any discrete emotion. I now turn to a discussion of each property.

---

<sup>22</sup> In the scientific literature, the term "automatic" is often used to indicate that a given emotion is associated with distinct autonomic nervous system reactions. In this thesis I am referring to this as "physical distinctiveness." By "automatic" I mean that the quick on-set of the production of facial expressions and automatic activations of attitudes. That is, the accessibility of information in memory that can be called upon immediately in response to an emotionally salient stimuli. See Fazio (1986) and Bargh (1992) for more on the topic of the automatic activation of attitudes.

<sup>23</sup> In the literature what I am calling pan-cultural is often termed "universal." If something is universal or pan-cultural, then it is reliably recognized in all cultures.

<sup>24</sup> In the scientific literature if something is primitive, then it appears in infants and is homologous in other mammals.

### 3.2.1.1- Facial Expressions (P<sub>1</sub>)

One of the most frequently used methods to study the emotions is to correlate empirical data with emotionally salient facial expressions.<sup>25</sup> I will begin this section with a general discussion of the methodology and goals in the scientific study of facial expressions of emotion, which will be relevant throughout the thesis. I will then reconstruct Barrett's argument that there are no signature facial expressions for discrete emotions.

In general, there are two methodologies in the study of facial expressions of emotions. The first, and most commonly used, Barrett calls *perception-based* studies of the face (2006b, 37).<sup>26</sup> In these studies subjects are presented with pictures of facial expressions of discrete emotions and asked to judge which discrete emotion is being portrayed.<sup>27</sup> While subjects are engaged in this task, certain physiological measurements, such as autonomic nervous system reactions<sup>28</sup> as well as neural activations, are recorded. The assumption is that in order for subjects to exhibit above-chance accuracy at judging the correct discrete emotion portrayed in the facial expression they must be able to extract or decode information about the specific discrete emotion based on the facial expression.

---

<sup>25</sup> Historically, the scientific desire to find the facial signals of emotions can be traced back to Darwin's book *The Expression of the Emotions in Man and Animals* (1859), but there has recently been an explosion of empirical work specifically on facial signals. This is due to studies conducted in the 1960's and 1970's by Ekman (1971) and, independently, Izard (1977) where they presented subjects with pictures of emotional facial expressions and asked them to identify the corresponding emotion. They found that subjects reliably recognized the facial expressions across cultures and even in preliterate cultures with little contact with Western cultures. These findings lead scientists, like Ekman and Izard, to claim that these expressions are pancultural, primitive, and automatic. Ekman (1976) further mapped out the facial muscles involved in these expressions and formed the Facial Action Coding Scheme (FACS), which allows scientists to describe and recognize any facial movement.

<sup>26</sup> This type of study is often termed the "recognition of emotional expressions."

<sup>27</sup> There are databases of pictures that researchers use that have been tested to show that they are reliably judged by subjects as depicting the given emotion. Ekman's (1976) Facial Action Coding Scheme (FACS) has produced a set of pictures of facial expressions of discrete emotions that researchers use. There is also an alternative set of pictures that researchers use created by Lang, Bradley, and Cuthbert (1999) called the International Affective Picture System (IAPS). These pictures are not of facial expressions per se, but more generally show emotionally evocative events. The pictures from the IAPS database are used more frequently in studies that assume a dimensional approach to emotions.

<sup>28</sup> See footnote 16 for an explanation of the autonomic nervous system. The central point to keep in mind is that the autonomic nervous system is generally not under conscious control and so is thought to provide good evidence for an underlying causal mechanism.

The second methodology used in studies of facial expressions Barrett calls *production-based* studies of the face (2006b, 37). In these sorts of studies, researchers measure the facial expressions of the subjects and other physiological responses (such as heart rate) during an emotionally evocative event. In order to evoke the emotion, researchers ask subjects to remember an emotionally salient event in their lives (for example, the death of a loved one) or present subjects with pictures that are emotionally evocative, such as pictures of dead children. Throughout the paper I will distinguish between perception-based studies and production-based studies to clarify an understanding of the scientific research I am about to present.

In production-based studies the goal is to map out the specific set of facial muscle movements for each discrete emotion and find bodily signatures for them. In perception-based studies the goal is to distinguish which discrete emotions are recognized universally across cultures (what I will call *panculturality*) and the physiological responses associated with the discrete emotion.<sup>29</sup> The general goal of both methodologies is to find facial expressions of emotion that are automatic (what I will call *automaticity*)<sup>30</sup> and primitive (what I will call *primitivity*).<sup>31</sup> Fulfilling these three criteria -automaticity, panculturality, and primitivity- ensures that one's theory of the emotions meets the rough requirements for a good theory of the emotions as introduced at the beginning of chapter two. Panculturality is assumed to demonstrate that the emotion is not culturally relative or individualistic. An ideal theory of the emotions would not just tell us about an individual's emotion, but about any individual in any culture. Automaticity and primitivity are thought to provide explanations of the emotions wherever they appear, in

---

<sup>29</sup> In other words, in perception-based studies subjects from a wide variety of cultures must be able to accurately judge the emotion in the picture and this is assumed to demonstrate that the emotion has an underlying causal mechanism.

<sup>30</sup> The extraction or decoding and production of the discrete emotion portrayed in the pictures of facial expressions by the subject is thought to take place compulsively and quickly.

<sup>31</sup> The ability to decode or extract this information is thought to be primitive or innate because children and other mammals may be able to recognize and/or perform facial expressions of discrete emotions. For example, a primate may display a facial expression that is equivalent to a human smile and thus an expression of happiness.

humans or non-humans, and may indicate that there are homologous structures. Meeting all three criteria provide reasons for hypothesizing that there is an underlying causal mechanism for the emotion and ensures that the theory meets the criteria for a good theory of the emotions.

Comparative studies are invaluable to ensure that the theory applies to all individuals that experience emotions, especially non-humans. I now turn to Barrett's argument that discrete emotions do not fulfill these criteria.

#### **3.2.1.1.1- Automaticity**

Production of a facial expression is automatic when it is not under rational control and has a quick on-set. For example, a facial expression of *fear* is automatic if it is produced by the subject within a short amount of time absent rational control. Discrete emotion theorists have argued that the production of facial expressions corresponding to the discrete emotions takes place within around 25 ms (Ekman 1999b). According to Barrett, evidence from animal communication research has failed to support this hypothesis. Seyfarth and Cheney (2003) showed clearly that nonhuman animals rarely produce involuntary, reflexive displays (Barrett 2006b, 39). It is not clear if this is really an objection to the claim that the facial expressions of discrete emotions have a quick on-set, but given this "evidence" Barrett concludes that production of discrete emotion facial expressions are not automatic. Thus, automaticity is not a property that can be attributed to discrete emotions as part of the property cluster.

#### **3.2.1.1.2- Panculturality**

In addition to the automaticity of facial expressions, discrete emotion theorists claim that facial expressions of discrete emotions are pan-culturally recognized (e.g. Izard 1977). Barrett rejects this claim. She argues that in perception-based studies of the face there are several factors that bias the results- i.e. factors that inflate subjects' accuracy in ratings of discrete emotion

faces. In a meta-review of the empirical literature Elfenbein and Ambady (2002) found that accuracy rates reported for each discrete emotion portrayed in the facial expression vary with the research team that conducted the study. Barrett attributes this variation to the employment of “methods that were more likely to produce large cross-cultural accuracy rates” (Barrett 2006b, 38). In other words, the results can be attributed to the methodology employed by the researchers, not the recognitional abilities of subjects. Three classes of biases artificially raise accuracy rates.

First, there are social biases. Take for example the phenomena known as “in-group advantage.” The argument is that “people are generally more accurate at judging emotional behaviors depicted by members of their own cultural group than at judging those depicted by members of a different cultural group” (2006b, 38). A study by Elfenbein and Ambady (2003), found that in-group advantage has significant impact on the accuracy of subjects’ judgments. The subjects in Elfenbein’s and Ambady’s study were more accurate at picking out the correct facial expression on average by 9.3% if the facial expression was portrayed by a member of their own culture. As another example of a social bias, Mesquita (2003) has argued that the task itself is highly Westernized task and may not be a valid task for members of non-Western cultures.

Second, there are technique biases. Discrete emotion theorists often use short lists of terms, creating what Russell calls a “forced-choice method” (1991). The method is “forced” because the subjects are not allowed to assign their own terms to the facial expressions, but are provided with terms to label the expressions. The method is further “forced” because the terms that subjects are given are discrete emotion terms that are already suspected to be pan-cultural. Another technique bias is using static, poised expressions that Barrett says “depict caricatures of emotion” rather than having subjects judge spontaneous facial movements (2006b, 38). She

believes that these caricatures do not reflect the prototypical expression (emoters' actual production of the expression in everyday interactions) and that it "departs from the central tendency of its category in a way that will make it maximally distinctive from other categories" (Barrett 2006b, 38). In other words, the pictures show exaggerated facial expressions and these are different from the expressions seen in everyday life. Thus, subjects' responses are not indicative of the discriminating abilities that we employ in everyday life.

For example, a prototypical facial expression of *anger* would depict the expression that is seen when an individual is experiencing *anger*, whereas a caricature facial expression of anger would exaggerate the prototypical expression so that it is easier to distinguish and classify from other emotions like *sadness*. "Caricatured stimuli are easier to categorize than prototypic stimuli when the categories in question are highly interrelated" (Barrett 2006b, 38). Barrett concludes that "the fact that caricatures give the clearest results in emotion-perception studies may be indirect evidence that the categories of *anger*, *sadness*, *fear*, and so on are themselves highly interrelated without firm boundaries" (Barrett 2006b, 38). In sum, given these issues the claim by discrete emotion theorists that the ability to distinguish between these expressions is pan-cultural is on poor empirical footing. Therefore, panculturality is not a property that can be attributed to discrete emotions as part of the property cluster.

### **3.2.1.1.3- Primitivity**

Discrete emotion theorists have also claimed that facial expressions of discrete emotions are primitive (e.g. Ekman 1999b). The ability to discriminate facial expressions is said to be primitive if it is exhibited early in development (e.g. by infants) and in non-humans. Barrett believes that evidence from developmental psychology calls into question this claim. First, she argues that the ability to recognize facial expressions itself (whether they are emotional or not) is

not primitive. Infants do not seem to recognize faces *as* faces. Although infants show a general preference for faces, this is arguably due to general perceptual preferences and not due to the meaning of the facial expression (for a review see Turati 2004). Second, children categorize faces on the basis of simple features of the face like open mouths, or the showing of teeth. In several studies, researchers have found that infants cannot distinguish between a smile, associated with *happiness*, that involves the showing of teeth and *angry* faces that also involve the showing of teeth (Caron et al., 1985). According to Barrett, this shows that infants use criteria such as teeth to categorize facial expressions rather than by using the emotional meaning associated with the face, such as *happiness* or *anger*.

Barrett acknowledges that this objection may be dismissed. It could be the case that the recognition of facial expressions of discrete emotions is primitive, but that this ability does not materialize at birth. Instead the ability may develop as certain neural systems develop. Even though this may be the case, Barrett believes we can at least conclude that infants and children do not have the conceptual knowledge needed to recognize facial expressions of discrete emotions. In fact, according to a study by Widen and Russell (2003) children do not possess the knowledge of emotion concepts until the age of five. Before the age of five children use a few emotional terms, like “happy” and “sad,” to describe all emotions and use larger categories like “positive” and “negative” to describe emotions in general (*Ibid.*). Barrett concludes that perception based studies of discrete emotion faces demonstrate that the ability to discriminate facial expressions of discrete emotions is not primitive. Thus, primitivity is not a property that can be attributed to discrete emotions as part of the property cluster. The next property that is claimed to be part of the natural property cluster kind of discrete emotions is the semantic structure of discrete emotion terms.

### 3.2.1.2- Semantic Structure (P<sub>2</sub>)

Barrett argues that if emotions are best studied as discrete categories, then self-reports of emotional experience should exhibit a simple structure (2006b). That is, when subjects are asked about their emotional experiences in self-reports, the terms they use to describe these experiences should be grouped around discrete categories (*fear, anger, happiness*, and so on). In other words, “measures of emotion experience should produce evidence of discrete, discriminable categories” (Barrett 2006b, 35). According to Barrett, evidence of this sort would indicate that reports of any given emotion (e.g. *sadness*) constitute a unified content area (e.g. feelings of *sadness*) and not another content area (e.g. feelings of *anger*). In addition, if semantic studies showed a consistent grouping of discrete categories, this would indicate that the given emotion could not be broken down into more basic parts, like valence and arousal.

Barrett argues that there is little evidence in self-reports that people classify their emotional experience in terms of discrete categories. That is, subjects do not consistently group their experience of emotions into discrete units. For example, a number of studies have reported that subjects do not consistently group pictures of negative facial expressions, such as *disgust, fear, and sadness* (e.g. Ekman 1971). Reports of negative emotions like *fear* and *sadness* are so highly correlated that there is often no unique variance to distinguish these discrete emotions. Barrett claims that “scales that are explicitly built to measure discrete emotions tend to show high correlations between like-valenced states” (2006b, 35). For example, the fact that facial expressions of discrete emotions like *fear* and *sadness* are not distinguished consistently from one another may be explained by the fact that they are both negatively valenced. Barrett concludes that there is not a unified, simple semantic structure to discrete emotions terms and

therefore semantic structure is not a property that can be part of the natural property cluster of discrete emotions.

### 3.2.1.3- Physiological Distinctness (P<sub>3</sub>)

Since James (1884) proposed that each discrete emotion has a unique bodily pattern, researchers have been attempting to identify these signature bodily responses. Researchers are primarily concerned with the autonomic nervous system responses because these are generally not under conscious control. The assumption is that responses not under conscious control demonstrate patterns that have been adaptive and indicate an underlying causal mechanism. Several researchers argue that there are distinct bodily responses associated with discrete emotions (e.g. Ekman, Levenson, and Friesen, 1983). Barrett believes that studies supporting this claim are misleading. She says that “meta-analytic reviews of this literature generally find that categories like *anger*, *fear*, *sadness*, *disgust*, and *happiness* cannot be fully differentiated by autonomic activity alone (Cacioppo et al., 1997; 2000)” (2006b, 41). Discrete emotion theorists have similarly reported that subjects have trouble *distinguishing between* negative emotions, like *disgust*, *anger*, and *fear* (e.g. Ekman 1971; Izard 1977).

At the same time, Barrett acknowledges that *some* distinctive differences have been found for discrete emotions, but not for *most* of the discrete emotions. Meta-studies have shown that there is a high amount of heterogeneity between the bodily responses recorded for given discrete emotions. For example, *anger* has been found to increase the heart rate in comparison with *happiness* and decrease heart rate in comparison to *fear* (for a review see Cacioppo et al. 2000). Thus, there is no characteristic level of heart rate or change in heart rate associated with *anger*. From this and other empirical evidence, Barrett concludes that “such heterogeneity likely indicates that unspecified variables are moderating the relations between emotion categories and

these peripheral nervous system responses<sup>32</sup> (2006b, 41). Barrett believes that these unspecified variables are most likely valence and arousal.

Barrett also offers alternative explanations for the data supporting distinct physical responses associated with discrete emotions. First, vascular patterns that appear to distinguish between *fear* and *anger* also distinguish between *threat* and *challenge* appraisals according to several studies (e.g. Blascovich and Mendes 2000). Barrett argues that researchers may inadvertently manipulate appraisals of challenge and threat while trying to generate expressions of *anger* and *fear*. Second, skin conductance responses that appear to configure around discrete emotions may be explained by attention allocation, which some studies have shown (e.g. Frith and Allen, 1983).

A third alternative explanation Barrett provides is that autonomic nervous system responses may be in the service of behavior or expected behavior. The appearance of distinct physiological response for the discrete emotions may be explained by metabolic demands associated with the actual or expected behavior of the given discrete emotion. For example, the increased heart rate associated with *fear* may be due to the expected or actual metabolic demands associated with a fight or flight response. Additionally, Barrett claims that a diverse set of behaviors are evident for discrete emotions so that there are not behaviors unique to discrete emotions like *fear*. She concludes that “if autonomic activity is in the service of behavior (or expected behavior), and if a heterogeneous range of behaviors is associated with a given emotion category, then emotion-specific autonomic patterns are unlikely on *a priori* grounds” (2006b, 41-42). Instead, the apparent configuration of autonomic nervous system reactions may be due to the subjects’ preparation for specific behaviors, rather than the discrete emotion.

---

<sup>32</sup> The peripheral nervous system is the system in which sensory, motor, and autonomic nerves are linked to the central nervous system.

Barrett concludes that the evidence for distinct autonomic nervous system responses for the discrete emotions is problematic, since there are inconsistent findings across studies and alternative explanations have not been ruled out. There does not appear to be a property of unique bodily responses associated with each discrete emotion. Therefore, physiological distinctiveness is not a property that can be attributed to the discrete emotions as part of the property cluster.

#### **3.2.1.4- Behaviors (P<sub>4</sub>)**

The next property that is signature behaviors for each discrete emotion. For example, it has been claimed that the behavior associated with *anger* is attacking or the urge to attack and *fear* is associated with freezing or fleeing or the urge to freeze or flee (Frijda, 1986). Barrett claims that signature behaviors are not and cannot be associated with specific discrete emotions. Emotions are “specific, context-bound attempts to deal with a situation” and “functional demands vary with situations, making it likely that instances of the same emotion can be associated with a range of behaviors [instead of a distinct behavior]” (Barrett 2006b, 41).

As evidence for the claim that signature behaviors are not observed for each discrete emotion, Barrett cites the animal learning literature. Several studies have shown that animal behavioral responses correspond to situational demands (e.g. Bouton, 2005). For example, rats that are put into a *fear* situation (that is defined by the presence of a predator) show a range of behaviors from orientating their attention to the predator, freezing, jumping, or attacking depending on the proximity of the predator (for a review see Bouton, 2005). Barrett concludes that behaviors do not group consistently around specific discrete emotions and that many different behaviors are associated with many different discrete emotions. Therefore, behavior

does not constitute a property that is indicative of each discrete emotion and that can be claimed to be part of the natural property cluster kind.

In sum, Barrett has argued that the discrete emotions do not fulfill requirement 1 for natural kind status. There is no distinct facial expression that is automatic, pan-cultural, and primitive for the discrete emotions. Additionally, she argues that there is little evidence for the discrete emotions in semantic structure, no distinct physical response for them, and no signature behaviors associated with the discrete emotions. Therefore, there is not a cluster of often co-occurring properties.

### **3.2.2- Do the Discrete Emotions Share Causal Mechanisms?**

The causal mechanism that underlies the homeostatic property cluster (requirement 2) of each discrete emotion is assumed to be a distinct neural mechanism in the brain.<sup>33</sup> Many discrete emotion theorists have claimed that discrete emotions have specific neural causes (e.g. Damasio, 1999; Ekman, 1992; Izard, 1993; LeDoux, 1996; Panksepp, 1998). For example, it has been claimed that the basal ganglia is the causal mechanism for *disgust*. Barrett cites two meta-analyses of neuroimaging studies of discrete emotions over the past 10 years that do not show distinct neural activations corresponding to the discrete emotions (Murphy et al., 2003; Phan et al., 2002).<sup>34</sup> According to Barrett, these studies found that “unique activation patterns for each category of emotion were difficult to discern, and those that materialized were less consistent than expected” (2006b, 43). In particular, the meta-reviews showed that there was a wide

---

<sup>33</sup> A causal mechanism can be cashed out in several different ways: as distinct neural circuits, psychological mechanisms (e.g. affect programs- Ekman, Izard, Tomkins), cognitive modules (Cosmides and Tooby 2000), neurotransmitters, or motivations and it is assumed that all of these options will have some sort of correspondence with the brain. A term that is often used when discussing causal mechanisms is “brain markers.” One way of finding such a “brain marker,” popular in the literature is to apply neuroimaging techniques (MRI, fMRI, PET) to find increased activation.

<sup>34</sup> These meta-reviews included a variety of experimental paradigms, including emotion perception (showing subjects pictures of facial expressions of emotions) and emotion induction (where subjects are presented a stimulus to induce an emotional state). See Barrett 2006b, 43-45 for a detailed explanation of the meta-reviews.

divergence between reported neural localizations for *anger*, *happiness*, and *sadness*. I proceed to give a particular example that Barrett emphasizes.

The meta-reviews reported the most agreement between studies regarding the association of *fear* and the amygdala, but according to Barrett the “correspondences were lower than what might be expected if the amygdala represented a core *fear* system in the brain” (2006b, 43). For example, Phan et al. (2002) reported that only 60% of the studies reviewed found increased activation in the amygdala in studies involving *fear* and Murphy et al. (2003) reported that less than 40% of the studies reviewed found increased activation in the amygdala in studies involving *fear*. In addition, subjects with amygdala damage can correctly classify facial expressions of *fear* when their attention is directed towards the stimuli (Adolphs et al., 2005) and the amygdala has been found to be active in situations that do not involve *fear*, such as eye gaze (Adams et al., 2003). It seems there is not a strong association between *fear* and the amygdala. Thus, Barrett claims that the amygdala is likely not the causal mechanism associated with *fear*.

Barrett also argues that there are alternative explanations for the apparent association between the amygdala and *fear*. For example, stimulus features such as uncertainty or novelty may have produced this correspondence. Similar problems have been observed for other discrete emotions and proposed causal mechanisms. Given the empirical evidence and the possibility of alternative explanations, Barrett concludes that there is little support for the claim that discrete emotions have distinct neural causes and therefore the discrete emotions fail to fulfill requirement 2 for natural kind status.

Let us take stock of the argument so far. Barrett has argued that there is neither a cluster of properties associated with discrete emotions, thus failing to fulfill requirement 1, nor a causal mechanism associated with them, thus failing to fulfill requirement 2. In particular, there are no

facial expressions that are automatic, pan-cultural, and primitive for each discrete emotion, there is no evidence of distinct groupings of the semantic terms for discrete emotions, there are not unique physical reactions associated with the discrete emotions, and a diverse set of behaviors are associated with the discrete emotions. Additionally, there is little to no evidence of a causal mechanism for each of the discrete emotions.

According to core affect theorists, such as Barrett and Russell, one of the main reasons that the discrete emotion approach has not established natural kinds is because it uses a folk category rather than a scientific one. Now that we have seen Barrett's argument against the natural kind status of discrete emotions, I will turn to the alleged theoretical advantages of core affect.

### **3.3- Alleged Theoretical Advantages of Core Affect**

The goal of the scientific study of emotion is to find natural kinds about which inductive generalizations can be formed. As we just saw, Barrett argues that the empirical evidence does not support the thesis that discrete emotions are natural kinds. Core affect theorists believe that their theory avoids the problems associated with the discrete emotion approach by proposing a more elegant scientific theory. Further, they suggest that the case for core affect as a natural kind is more promising than the case for discrete emotions as natural kinds. I will consider these alleged theoretical advantages in turn.

Core affect theorists insist that their theory is more elegant than discrete emotion theories for two reasons: it deals with a more basic scientific concept and it is ontologically more parsimonious. The basicness of core affect is related to the fact that its components - valence and arousal - are simple and fundamental. Core affect theorists claim that valence and arousal cannot be captured by a "lower" concept and that they are more fundamental constituents of affective

phenomena than discrete emotion categories (Russell 2003). Additionally, many theorists have described something along the lines of core affect as a basic psychological process across species (for example, Wundt, 1897; Cannon, 1927; Davidson, 2000; Cacioppo, Gardner, and Berntson 1999).

Second, core affect is thought to be more ontologically parsimonious because it posits fewer theoretical entities than the discrete emotion approach. The discrete emotion approach would have us believe that there is a natural kind for each of the discrete emotions. Core affect theorists are convinced instead that the affective realm can be further simplified:

A frugal ontology may be all that is needed: Emotional life consists of the continuous fluctuations in core affect, in pervasive perception of affective qualities, and in the frequent attribution of core affect to a single Object, all interacting with perceptual, cognitive, and behavior processes. Occasionally, these components form one of the prototypical patterns [i.e. a prototypical or discrete emotion].<sup>35</sup>

Simplicity and ontological parsimony are epistemic virtues that any scientific theory tries to fulfill. All things being equal, the simpler and more parsimonious theory should be preferred. This being said, general epistemic virtues are not what is most important to assess the quality of a scientific theory. What matters most is that the theory matches the causal structure of the world. Boyd's account of natural kinds is formulated so that natural kinds are defined as the kinds which best conform to the causal structure of the world. If core affect theory were simpler and invoked fewer theoretical entities, but performed poorly in its search for natural kinds, it would not be preferable. So what is really at issue is the evidence core affect theorists provide in support of the thesis that core affect is a natural kind. I will accept for the sake of argument that core affect theory is simpler and ontologically more frugal than the discrete emotions, but leave this advantage in the background in what follows.

---

<sup>35</sup> Russell 2003, 152.

In support of the claim that core affect may be a natural kind, Barrett provides empirical evidence that members of the core affective category share a cluster of properties non-accidentally (thus fulfilling requirement 1) in virtue of a common causal (homeostatic) mechanism (thus fulfilling requirement 2). As a result, scientifically interesting generalizations can be formed about all instances of core affect. I will now reconstruct Barrett's argument that core affect fulfills each of these requirements. In chapter four I will argue that core affect is not a natural kind. I will present some disconfirming data or objections to her arguments here, but my argument will be further explained in the subsequent chapter

### **3.3.1- Do Instances of Core Affect Share Properties?**

Barrett alleges that the four properties commonly mentioned by discrete emotion theorists (those discussed above in chapter 3.2) better-support core affect as a natural kind. These properties shared by core affect are: (P<sub>1</sub>) automatic, pan-cultural, and primitive signature facial expressions, (P<sub>2</sub>) the semantic structure of affective terms, (P<sub>3</sub>) physiological distinctness, and (P<sub>4</sub>) behavioral distinctiveness. I will now consider the evidence for each and then assess whether it supports the case for the natural kind status of core affect.

#### **3.3.1.1- Facial Expressions (P<sub>1</sub>)**

As I pointed out in chapter 3.2, there are two general methodologies employed in the scientific study of facial expressions of emotion: perception-based studies (examining whether subjects can accurately identify emotions in pictures) and production-based studies (examining whether subjects produce facial expressions that are signatures of different emotional states). The goals of these studies are to find facial expressions of emotions that are automatic, pan-cultural, and primitive. Now, do the facial expressions associated with core affect have these characteristics?

### 3.3.1.1.1- Automaticity

Barrett argues that there is evidence for the automaticity of the perception of facial expressions associated with core affect. The perception is said to be automatic if it is fast and occurs without rational deliberation. Several studies have found that the initial perception of valence and arousal in facial expressions takes place within 25 ms of encountering the stimulus (Bargh 1997; Bargh, Chaiken, Gollwitzer, and Trötschel 1992; Bargh, Chaiken, Raymond, and Hymes 1996; Fazio, Sanbonmatsu, Powell, and Kardes 1986). Studies are conducted by presenting subjects with one stimulus (the primer) and a second stimulus (the target) in short time intervals. It is assumed that the recognition of the target as having a certain property, such as valence, in short time intervals indicates that a minimal amount of rational deliberation is involved, thus, providing evidence for automaticity. Other studies have shown that faces of positive versus negative emotional states that are presented subliminally influence subsequent judgments of the valenced value of ambiguous stimuli, whereas supraliminal presentations did not (Murphy and Zajonc 1993). In other words, the subjects' perception of the target is influenced by the primer only when the target and primer are presented in short time periods and not in longer time periods. This seems to indicate that the perception of facial expressions of core affect are automatic.

However, such findings depend on specific primers and on the strength of the association of the primer with the target. As Fazio et al. note "the present findings regarding attitude activation following repeated expression can be attributed to the resulting strength of the object-evaluation association and not to the extremity of the associated stimuli" (1986, 235). In fact, there was only an automatic association when primers were followed by the same type of target, so that the automatic association only appeared when, for example a positively valenced primer

was followed by a positively valenced target. In trials where they presented subjects with, for example a negative primer followed by a positive primer, the subjects did not automatically choose the correct valenced dimension. Additionally, these studies have been critiqued because the subjects were asked to memorize the primer and therefore may have been consciously assessing the target, rather than automatically associating the primer and the target. However, in later studies (such as Bargh et al. 1996) subjects did not memorize the primer, and there remained a high correspondence between presentation of the valenced primer and subjects' ability to categorize the valenced target,<sup>36</sup> but only when the valence was the same. Therefore, facial expressions of core affect appear to be automatic.

### **3.3.1.1.2- Panculturality**

Second, there have been studies that Barrett claims show that facial expressions of core affect are pancultural. Some studies on facial expressions of core affect have asked people to choose dimensions rather than discrete emotions that correspond to the expression; “two multidimensional-scaling studies of the feelings conveyed by facial expressions both yielded the same pleasure and arousal dimension in different languages” (Russell 1991., 439). For example, studies in Japanese and English (Osaka, 1986), in Greek, Chinese and English (Russell et al., 1989), in Greek (Triandis and Lambert, 1958.), and in Dutch (Frijda, 1953) all show that subjects group the expressions consistently around the same level of pleasure and arousal. In other words, subjects reliably discern the location of the facial expressions of core affect on the circumplex. As an example, subjects rate facial expressions corresponding to “lethargic” and “depressed” as negatively valenced and not highly arousing. Further, “no one has yet shown any influence of

---

<sup>36</sup> The consistency was statistically more significant (ranging from 100% to 97%) when there was a strong association between the primer and the target. When there was a weak correspondence the consistency was still statistically significant, but was much lower (ranging from 87% to 73%) (Bargh et al. 1996, 115).

language or culture on the nature of the dimensions found in these studies” (Russell 1991, 440). This seems to indicate that perception of facial expressions of core affect is pan-cultural.

These studies are compelling, but there are fewer studies that focus on dimensions rather than discrete emotions and in addition it seems that no one has done a study on a preliterate, isolated culture. The lack of data from a preliterate, isolated culture is necessary because it could be objected that the consistency in ratings of facial expressions and core affective levels is due to cultural influence and not due to some innate mechanism. This was an important critique of the basic emotions theory and a study conducted by Ekman (1971) showed that even people in isolated, preliterate societies could correctly identify the facial expressions with the corresponding emotion. Thus this is an objection that core affect theorists will need to address.

#### **3.3.1.1.3- Primitivity**

Third, there is evidence that facial expressions of core affect are primitive. Core affect theorists claim that we are born with the ability to feel pleasure and displeasure. There is much evidence that babies can feel pleasure and displeasure in the psychological literature (e.g., Emde, 1976, Osgood, 1975; Spitz, 1965, Sroufe, 1979) and that they communicate these feelings through facial expressions. Infants express negative affect through distinct facial expressions (Camras, 1992; Camras, Oster, Campos, & Bakemant, 2003) or in terms of arousal (Messenger, 2002).

In addition to the production-based studies on infants, there is also evidence from studies on children. Children discriminate and label faces largely on the basis of valence and arousal (Bullock & Russell, 1984, 1985, 1986; Hosie, Gray, Russell, Scott, & Hunter, 1998; Russell & Bullock, 1986). So, there appears to be evidence from both production and perception-based studies of facial expressions that children produce and recognize facial expressions of core

affect. Therefore, facial expressions of core affect appear to be primitive because there is evidence that infants and children produce and perceive facial expressions of emotions in terms of valence and arousal.

What is missing from this story is a connection with animal studies, as Panksepp (2007) rightly points out. If emotions emerge from valence and arousal, then there should be evidence in studies conducted on animals. Panksepp states that it is not clear if core affect is only supposed to underwrite emotional expression found in humans and not in non-humans.<sup>37</sup> If facial expressions of core affect are only applicable to humans, then I think the evidence given does support the claim that facial expressions of core affect are primitive. But Russell claims that core affect is a biological product of evolution, in part to prepare for action and influence behavioral choices, suggesting that there may be homologues in non-humans (2003, 156). If facial expressions of core affect do have homologues, then the claim that core affect is a natural kind may be difficult to prove because it is not clear how scientists would go about studying the facial expressions of core affect in non-human animals. Core affect theorists recognize the need for evidence of this sort (Barrett 2007c).

### **3.3.1.2- Semantic Structure**

According to Russell and Barrett, the semantic structure of core affective terms, as judged by subjects in self-reports, group consistently around levels of valence and arousal. Russell says “part of the meaning of all mood and emotion-related words in any language can be summarized by these two underlying dimensions [valence and arousal]” (2003, 153). Anna Wierzbicka (1999) has found that all human languages have phrases that roughly translate as “I feel good” and “I feel bad.” This suggests that the concepts of feel, good, and bad are recognized cross-

---

<sup>37</sup> This is a problem for core affect theorists because as I argued in chapter 2.3 valence and arousal have not been clearly defined. The phenomenological notion of valence and arousal seem to exclude non-humans or at least it is not clear how researchers would go about studying valence and arousal in this sense in non-humans.

culturally, which supports the claim that valence, at least, may have such a semantic role.

Additionally, Russell argues that arousal is also “likely [a] universal semantic dimension of emotion” (2003, 153). In support of the claim that valence and arousal are pancultural, Barrett (1995, 53) states:

The two dimensions have been identified in the semantic structure of affective terms and together typically account for a substantial amount of variance in self-reports of affective experience. Previous research has suggested that the dimensions represent the semantic components that individuals use to interpret and communicate their conscious, affective experience.

Below is a semantic based circumplex from studies that asked subjects to judge the levels of arousal and valence for affective terms like “nervous” and “happy.” Subjects were asked questions like whether “nervous” and “afraid” were similarly valenced and arousing, and subjects responded that they were. Subjects were also asked if “disappointed” and “satisfied” were similarly valenced and arousing, and subjects responded that they were not. This semantic based circumplex supports the core affect circumplex as a reliable projection of the affective space.

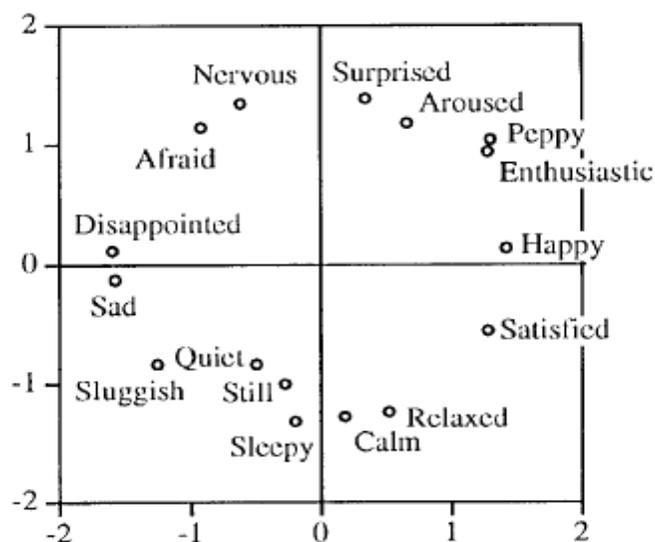


FIG. 1. The semantic-based circumplex structure of affect (Feldman, 1995a). Valence is the horizontal axis and arousal is the vertical axis.

The circumplex above provides good reasons for thinking that subjects categorize emotions in terms of core affect, but the subjects in these studies were all from Western cultures. A good theory of the emotions must be able to account for emotions across cultures and people. Barrett does cite one scientific study that she claims supports the cross-cultural grouping of core affect (Mesquita, 2003). Upon further inspection, though, Mesquita is concerned with providing a cultural model for the emotions that is consistent with the basic emotions theory.<sup>38</sup> She says that “some of these responses [facial expressions, peripheral nervous system responses, etc.] are demonstrably hardwired, but others may be socially learned” (2003, 887). If core affect is socially learned, then the theory will not apply to all cultures and humans. Therefore, reliable generalizations regarding the core affective category will not be forthcoming.

### **3.3.1.3- Physiological Distinctness**

In addition to evidence from facial expressions and the semantic structure of affective terms, Barrett claims that there are distinct autonomic nervous system responses for core affective states (Cacioppo et al., 2000; Lang et al., 1993). For example, negative affect is associated with higher diastolic blood pressure, increased heart rate, and decreases in electrodermal response duration. These measures are relative to positive affect and *contrastive properties* are seen for negative affect. So, whereas negative affect is associated with higher blood pressure and decreased electrodermal responses, positive affect is associated with decreases in blood pressure and heart rate and increases in electrodermal response duration. There are additional studies that she claims show that core affect involves distinct peripheral

---

<sup>38</sup> “The very course of emotions- the ways in which the potential of emotions is realized- appears to differ as a function of cultural models. Culture is, therefore, not spoiling or distracting from the basic theory of emotions, as has long been assumed. Rather, it should be considered a necessary aspect of emotion theory. Basically, emotions are cultural, even though that is not the only thing they are” (Mesquita 2003, 887).

nervous system responses (Cacioppo et al., 1997, 2000; Bradley & Lang, 2000). Therefore, Barrett argues that there are distinct physical reactions associated with core affect.

One objection to the claim that core affective states have distinctive physiological signatures is that all of these studies were done on humans. Again, a good theory of the emotions should be able to explain emotional phenomena wherever it is present and many people would argue that animals experience emotions. Thus, the distinctive physiological signatures found in humans, must also be found in animals, or an argument must be provided for why the physiological measurements of animals do not matter. Additionally, Panksepp (2007) notes that if core affect better explains affective states than discrete emotion approaches, then core affective theorists need evidence to explain away the evidence from animals given by the discrete emotion theorists

#### **3.3.1.4- Behaviors**

According to Barrett, voluntary actions, behaviors, or action tendencies (readiness to achieve or maintain a particular sort of relationship with the environment) are distinct for core affect. There is evidence from scientific studies that some behaviors can be considered expressive, in the sense that organisms' behaviors are indicative of its core affective states. For example, Cacioppo and Gardner (1999) argue that approach behavior indicates positive valence and withdrawal behavior indicate negative valence. Additionally, valence and arousal have been used to explain diverse sets of behaviors (see Russell 2003 for a review). The experience of negatively valenced states over prolonged periods of times has been used to explain psychological problems involving sex (Abramson & Ponkerton 1995), drug abuse (R.L. Solomon 1977), aggression (Berkowitz 1993), and eating (Pinel, Assanand & Lehman 2000). While an understanding of core affect may contribute to explanations for behaviors and help an individual

to identify the emotional state of another, there is no evidence that there are unique behaviors that correspond to core affective states.

### **3.3.2- Do Instances of Core Affect Share a Causal Mechanism?**

Lastly, Barrett provides some reason to think that there is an underlying causal mechanism to core affective states and believes that this is the strongest evidence for core affect as a natural kind. There may also be a neurotransmitter associated with core affect. Barrett says that “the dopamine system in the nucleus accumbens is associated with negative affective motivational states more broadly (Berridge & Robinson, 1998; Reynolds & Berridge, 2002, 2003)” (Barrett 2006b, 45). Core affect theorists claim that areas of activation in the brain during emotional episodes coincide with areas of activation for valence and arousal, which would indicate a neurobiological substrate or brain marker for core affect (Panksepp 2000; Bradley & Lang 2000; Cacioppo et al. 1997, 2000; Wagner et al. 2003). Barrett states that “neural activations provide a strong empirical basis for hypothesizing that a general affect system constitutes the most basic building block of emotional life” (2006b, 49). Finding such a brain marker for core affect would explain the causal mechanism that underlies the category and provide evidence for the claim that core affect constitutes a natural kind.<sup>39</sup>

Barrett relies heavily on two meta-analyses of emotion research (Phan, et al 2001 & Wager, et al 2003). The meta-analysis by Wager reviewed 65 PET and fMRI studies of emotional tasks. The meta-analysis by Phan reviewed 55 PET and fMRI studies of emotional tasks. Barrett claims that these studies give us a starting point in analyzing the data collected to see if core affect does constitute a natural kind and that their findings support the claim that there may be a brain marker for core affect. She claims that “both analyses observed greater left-sided

---

<sup>39</sup> I should emphasize that the neurobiological information for this only provides a preliminary sketch of very broad brain areas. No one has yet been able to causally reduce any emotional experiences to neurobiological processes. (Barrett 2006b)

activations (left-lateralization) for approach- (vs. withdrawal) related affect” (2006c, 81). This would indicate that there may be a brain marker for at least the valence dimension.<sup>40</sup> When examining the empirical evidence for the arousal dimension, she admits that the evidence is less clear. In fact she believes that the causal mechanism underlying core affect has yet to be fully analyzed, but that from these meta-analyses we can at least draw a tentative conclusion that core affect activates particular areas of the brain, indicating that there is an underlying causal mechanism. In addition, the core affect theorists claim that there is evidence that psychological disorders arise from or are associated with brain areas relevant to core affect (Barlow 2002; Hariri & Holmes 2006; Hariri et al. 2005). I will now assess the evidence in more detail.

#### **CHAPTER 4: ASSESSING CORE AFFECT**

Now that we have considered Barrett’s arguments against the natural kind status of discrete emotions and for the natural kind status of core affect we are finally in a position to address the central question of this thesis: Is core affect a natural kind? In this chapter I will argue that, contra Barrett (2006), the evidence does not suggest that core affect is a natural kind. Before doing so, I want to remind the reader of what it takes to be a natural kind *sensu* Boyd (see section 3.1 for a more detailed analysis). Here is a compact summary of Boyd’s core idea:

A kind is natural if and only if its members tend to share a family of co-occurring properties and a causal (homeostatic) mechanism that brings about their co-occurrence non-accidentally.

In light of this account, there are two requirements a kind K must satisfy in order to qualify as natural:

*Requirement 1:* There is a family of often co-occurring properties  $P_1 \dots P_n$  associated with instances of K.

---

<sup>40</sup> I am glossing over some issues regarding the different dimensional approaches. I am here assuming that the approach dimension is related to the valence dimension (see Ch. 2), but only because Wager states that the “valence results were very similar to those for approach/withdraw.” (520) I assume this is also why Barrett includes it in her work. (2007b)

*Requirement 2:* There is a causal (homeostatic) mechanism that brings about the family of often co-occurring properties seen in instances of K.

In section 3.2 I considered Barrett's argument that the discrete emotions *do not* satisfy requirements 1 and 2. In section 3.3 I introduced her reasons to think that core affect *does* fulfill requirements 1 and 2. In the next section I will argue that despite Barrett's arguments to the contrary, core affect does not appear to fulfill requirements 1 and 2.

#### **4.1- The Status of Core Affect as a Natural Kind**

Barrett's key hypothesis on core affect is the following:

The empirical case supporting the hypothesis that core affect is a natural kind is suggestive (Barrett 2006b, 48).

Since core affect is a combination of valence and arousal, this hypothesis appears to suggest three things. Firstly, that valence is a natural kind. Secondly, that arousal is a natural kind. Thirdly, that the combination of valence and arousal – core affect - is a natural kind. In the subsequent sections of this chapter, I will argue that valence is not likely to be a natural kind (4.1.2), that arousal is not likely to be a natural kind (4.1.3), and that core affect is not likely to be a natural kind (4.1.4). In 4.1.5 I will argue that even though core affect may not be a natural kind, four potential candidates for natural kind status can be located in subsets of core affect.

##### **4.1.2- Is Valence a Natural Kind?**

Valence, as represented by the horizontal axis of the circumplex, ranges on a continuum from positive to negative. In order to claim that valence is a natural kind it would need to be the case that all instances ranging from positively to negatively valenced would tend to share a family of co-occurring properties and a causal (homeostatic) mechanism that brings about their co-occurrence non-accidentally. As I will demonstrate below, the empirical evidence does not support the case that the class of all valenced states constitutes a natural kind. The evidence

supports splitting valence into two groups corresponding to the extremes of the valence continuum: positive valence and negative valence. *Contra* Barrett, the empirical evidence suggests that positive valence and negative valence may constitute *distinct* natural kinds (I will return to the consequences of this point in 4.1.4 and 4.1.5).

The empirical evidence provided below will suggest that there are diverse causal mechanisms underlying positive valence and negative valence. If this is the case, then valence does not fulfill requirement 2. There also appear to be distinct properties clustered around each type of valence, showing that valence does not fulfill requirement 1 either. Thus, valence does not seem to meet the requirements for natural kindhood because it is too heterogeneous and does not conform to the causal structure of the world. It should be noted that the scientific studies I am about to discuss contrast these two extremes and the properties that are measured are thus relational.<sup>41</sup> I will begin by looking at the evidence for a cluster of properties for valence and then examine the empirical data for a causal mechanism for valence.

#### **4.1.2.1- Shared Properties?**

There are four *contrastive* pairs of properties that are associated with positive valence and negative valence and thus not with valence as an entire class. These properties involve facial expressions, physiological distinctness, behaviors, and effects on the health of subjects. The shared properties fulfill requirement 1, but only for the extremes of valence and not for valence

---

<sup>41</sup> I would like to briefly remind the reader of the methodology employed in these studies that was discussed in 3.2 and emphasize that these extremes (positive and negative valence/high and low arousal (discussed in the preceding section) are contrasted. Most of the studies use facial expressions or other pictures that are emotionally salient (either production-based studies or perception-based studies), measure certain data, and correlate these findings with self-reports in order to ensure that the subjects agree that the pictures represent the specific emotion state that they are supposed to represent. For example, subjects may be asked to view pictures depicting negative emotional expressions or states and positive emotional expressions or states. The measured data collected is correlated with self-reports and contrasted between positive and negative emotionally relevant states. For example, while viewing a negative emotional expression blood pressure increases in subjects in relation to the blood pressure of subjects viewing positive emotional expressions. Thus the reader should keep in mind that these properties are all relational properties. The same applies to high and low arousal states.

as a class. So there appears to be evidence that positive valence may be a natural kind and that negative valence may be another natural kind. I will now consider evidence which suggests this split of the valence category into two distinct natural kinds.

First, production-based facial studies have consistently shown that positively valenced emotions and negatively valenced emotions activate different facial muscles. Positively valenced emotions elicit more activation in the cheek region, specifically increased zygomatic (smile muscle) activity, as measured by facial electromyogram (EMGs).<sup>42</sup> In contrast, the *same* production-based facial studies have consistently shown that negatively valenced emotions show increased activation in the brow area (Cacioppo et al. 2000, 178-179) and the corrugator (frown) muscle. These heterogeneous and *contrasting* properties for positive valence and negative valence are observed while subjects are thinking about past experiences (i.e., in production-based studies) and while viewing pictures of positive emotional expressions (i.e., in perception-based studies) (Fridlund et al. 1984; Lang et al. 1993; Schwartz et al. 1980. For a review of the empirical studies see Cacioppo et al. 2000, 177-179). Production-based studies on infants show the same diverse facial muscle activation for positively valenced and negatively valenced emotions during emotionally salient stimuli. One study looked at infant cry-faces and infant smile-faces (Messinger 2002). Messinger found that the same facial muscles that are activated in adults are activated in infants. That is, cry-faces elicit more activation in the corrugator muscle and smile-faces elicit more activation in the zygomatic muscle. These diverse activations for positively valenced facial expressions and negatively valenced facial expressions, seen both in adults and infants, suggest that valence as a class is too heterogeneous to constitute a natural kind. Each type of valence appears to have its own distinct observable properties, in terms of

---

<sup>42</sup> EMGs record electrical currents associated with muscle contractions. Scientists have detailed maps of the facial muscles. See Ekman 1976.

facial expressions. So, just as there is no face of *anger*, there is no face of a valenced state simpliciter. At most the evidence suggests that there may be a face of a negatively valenced state and another face of a positively valenced state.

Second, the empirical evidence suggests that physical distinctness, defined as reactions in the autonomic nervous system, is only found at the extremes of positive valence and negative valence (not the entire class of valence). For example, during viewings of pictures that express positively valenced emotions, there is a general decrease in autonomic nervous system activity (Cacioppo et al. 2000, 183), although greater peak heart rate acceleration has been observed (Bradley and Lang 2000).<sup>43</sup> In contrast, the *same* studies found that there is greater autonomic activity when subjects view pictures representative of negatively valenced emotions, combined with significant heart rate deceleration. Notice that, as we saw above in the case of facial expressions, autonomic nervous system reactions and measurements in cases of negative versus positive valence form a *contrastive* pair. Once again, it seems that positively-valenced and negatively-valenced states are too heterogeneous to form a single natural kind.

Other studies on the distinctness of valence show similarly diverse reactions. When scientists measure blink responses in subjects viewing pictures of positively valenced emotional states, the number of blink responses decreases and quicker eye blinks are recorded (Lang et al. 1993; Bradley and Lang 2000). Again, the *same* studies show a contrastive reaction in the eye movements of subjects viewing pictures depicting negatively valenced emotional states. When subjects view pictures of negative expressions, the number of blink responses increases and slower eye blinks are recorded. Again, the evidence suggests that the properties are contrastive

---

<sup>43</sup> Bradley and Lang (2000) note that “as an index of emotional state, cardiac rate is less straightforward than other psychological measures.” See page 258 of their article for a detailed discussion of this point.

between negative and positive valence. There are no characteristic properties of valence as a class.

Third, subjects exhibit distinct behavioral reactions to positively valenced emotions and different behavioral reactions to negatively valenced emotions. For example, one study found that subjects rate pictures depicting positive emotional states as more interesting than pictures depicting negative emotional states (Lang et al. 1993). The *same* study observed that, despite this, subjects chose to view the pictures depicting positive states in shorter time periods and the pictures depicting negative states for longer time periods. Other studies have found that subjects shown pictures depicting negative emotional states show greater attention to them than to pictures depicting positive emotional states (Pratto and John 1991; Hansen and Hansen 1988). This means that subjects exhibit more orienting behaviors to pictures of negative emotional states than to pictures of positive emotional states. Once again, it appears that the properties do not cluster for valence as a class, but for the extremes of the continuum.

Fourth, there are increased health benefits associated with positively valenced emotions and decreased health benefits associated with negatively valenced emotions. Many studies have found that individuals who describe experiencing mostly positively valenced emotions in their daily lives live longer than those who report experiencing mostly negatively valenced emotions (e.g. Swenson et al., 1973; Herbert & Cohen, 1993; Kiecolt-Glaser et al, 1996; Lyubomirsky et al., 2005; Maruta et al., 2000). A study that examined the diaries of nuns, looking for positive and negative emotional terms, found that those who used more positive terms lived longer than those who used more negative terms (Danner et al., 2001). Another study found that the way men described bad emotional experiences (either in positive or negative terms), predicted their future health states. They found that those who described their experiences in positive terms

were healthier later in life than those who described their experiences in negative terms (Peterson et al., 1998). In a meta-review of the relation between positively valenced emotions and health benefits, Lyubomirsky et al. conclude that “our review of the cross-sectional empirical literature suggests that happiness [defined as positive valence] is positively correlated with indicators of superior mental and physical health” (2005, 825). One more time, the evidence suggests that there are contrasting health benefits for the extremes of valence.

To summarize, the empirical evidence from facial expression studies, physical distinctness, behaviors, and effects on health implies that valence is too heterogeneous to be a natural kind *sensu* Boyd. Even though ‘valenced states’ is not a homogenous kind, the classes of ‘negatively valenced’ and ‘positively valenced’ states may be. If X is positively valenced, X will tend to promote health effects, activate certain facial muscles, result in distinct behaviors, and have signature responses in the body. On the other hand, if X is negatively valenced, X will tend to promote negative health effects, activate certain (distinct) facial muscles, result in distinct behaviors, and have signature responses in the body. In sum, the evidence suggests there is one cluster of properties associated with positive valence, and another, distinct, even *contrastive* cluster of properties associated with negative valence. But a cluster of properties is only half of the story for being a natural kind *a la* Boyd. The key to a homeostatic property cluster is a causal mechanism. I now turn to examine the empirical evidence for a causal mechanism underlying valence.

#### **4.1.2.2- Shared Causal Mechanism?**

Empirical data suggests that there are different causal (homeostatic) mechanisms for positive valence and negative valence. As I will show, these causal (homeostatic) mechanisms appear to be *contrastive*, just as the properties above were. So there do appear to be reasons to

think that positive and negative valence fulfill requirement 2, but not the entire class of valenced states. Once more, valence appears to be too heterogeneous to form a natural kind, but the extremes may constitute natural kinds.

In emotion theory the assumption is that a causal mechanism for emotions will be found in the brain. We lack robust empirical evidence concerning causal mechanisms for positive valence and negative valence, but it should be noted that discrete emotion theorists have likewise failed to find strong correlations between individual types of discrete emotions and specific processes in the brain. Even though there is not a distinct brain area that we can call the causal mechanism for positive valence and negative valence, there is some general evidence that we can point to. Many studies have correlated positive valence with increased left-hemisphere activation and negative valence with increased activation in the right hemisphere. Cacioppo et al. states that most studies have found that “the left anterior region of the brain appears to be involved in the expression and experience of approach-related emotion [positive valence], and the right anterior region appears to be involved in the expression and experience of avoidance-related emotions [negative valence]” (2000, 185). The evidence for diverse causal mechanisms for positive and negative valence suggests that all valenced states are too heterogeneous to constitute a natural kind, but these extremes may be homogenous enough.

For example, a number of studies using electroencephalographic (EEG)<sup>44</sup> recordings of subjects during viewings of emotionally salient film clips have found that positive reactions (as indicated by self-reports) to positively valenced films are correlated with activity in the left-hemisphere. The *same* studies found that negative reactions to negatively valenced films were correlated with right-hemisphere activation (Cacioppo et al. 2000). Additionally, clinical observations (Robinson et al. 1984; Robinson & Downhill 1995) and experimental research

---

<sup>44</sup> EEGs record electrical activity in the brain by using sensors, called electrodes, attached to the subject’s head.

suggest that individual differences in the activation of the left-hemisphere result in a predisposition to experience positive emotions and individual differences in the activation of the right-hemisphere result in a predisposition to experience negative emotions (Cacioppo et al. 2000, 185). Clinical observations also support the right lateralization for negatively valenced emotions. One study found that a patient who had undergone a right temporal lobectomy responded just as a normal patient would to positive stimuli, but not to negative stimuli (Morris et al. 1991).<sup>45</sup> While these are extremely general, they at least point to diverse causal mechanisms.

Given the current empirical data the most that we can say is that increased activation in the left-hemisphere appears to be correlated with positive valence, and increased activation in the right-hemisphere appears to be correlated with negative valence. Even though the evidence does not show a robust causal mechanism for either positive valence or negative valence, it does suggest that there may be diverse causal mechanisms for the class of valence. Thus, it appears that all valenced states do not fulfill requirement 2 for natural kindhood, but the extremes of positive and negative may fulfill this requirement. I will now examine the evidence for the natural kind status of the other dimension of core affect: arousal.

#### **4.1.3- Is Arousal a Natural Kind?**

Arousal, as represented along the vertical axis of the circumplex, is a continuum from high levels to low levels. I now consider the hypothesis that all aroused states (ranging from high to low) tend to share a family of often co-occurring properties and a causal (homeostatic) mechanism that brings about their co-occurrence non-accidentally. I will suggest that the

---

<sup>45</sup> This was found not only in measurements while the subject was viewing negative pictures, but also was reported by the patient's family and friends.

evidence points to a split within the arousal dimension, analogously to the case of valence. The evidence suggests that high arousal and low arousal may be *distinct* natural kinds.

It should be noted that it is difficult to evaluate the natural kind status of arousal. There are several reasons for this. First, it is difficult to precisely define “arousal” as I argued in section 2.3. Arousal is most often thought of and measured by reactions in the autonomic nervous system, but the effects of the autonomic nervous system are far-ranging. Due to the far-ranging effects, it seems that there may be alternative explanations for these visceral properties that need to be eliminated. Barrett admits to this: “Patterns of autonomic and somatic responses relate in some way to feeling active and alert, slowed down and sleepy, or angry or sad, but there is no simple, one-to-one correspondence” (2004b, 685). It is important for scientists to keep in mind that if they want to find natural kinds they need to precisify their terms. According to Boyd, “natural kinds are solutions...to problems about how to sort things so as to facilitate reliable induction and explanation” (1999, 72). In order to accomplish this, the categories must be revised to fit the causal structure of the world. Second, there are few studies solely on arousal, whereas there are a plethora of studies on valence. Lastly, there are problems with identifying a causal mechanism for arousal.

The empirical evidence which is available suggests that the entire class of arousal is not a natural kind. Similarly to valence, arousal appears to split into two possible natural kinds (the extremes): high arousal and low arousal. The evidence I elucidate below suggests that there are clusters of distinct properties for high and low arousal, but not for the entire class of arousal. Therefore, arousal does not fulfill requirement 1. The evidence for a causal mechanism underlying arousal is sketchy at best, but I will suggest that there are reasons to think that there may be two underlying causal mechanisms in the category. Therefore, arousal does not appear to

fulfill requirement 2. Again, I will return to what this might imply about the natural kind status of core affect in 4.1.4 and 4.1.5.

#### **4.1.3.1- Shared Properties?**

There appear to be at least three *contrastive* pairs of properties for high arousal and low arousal, but no properties shared by the entire class. These properties are facial expressions, physical distinctness, and behaviors. These properties support requirement 1 for natural kindhood, but only for the extremes: high arousal and low arousal. I will present the empirical evidence for each of these properties below.

First, there is evidence that when a positively or negatively valenced emotion is viewed the level of tension in the relevant facial muscle group is related to the level of arousal. In other words, higher levels of arousal correspond to greater tension in the relevant muscles, depending on whether the emotion is positively or negatively valenced (Cacioppo et al. 1986). In particular, zygomatic (smile muscle) activity increases as arousal increases in viewings of emotionally positive pictures (Lang et al. 1993). Cacioppo et al. conclude that “the present research indicates that facial EMG activity varies as a function of...the intensity of affective reactions” (1986, 266). Messinger (2002) believes that his data on facial expression in infants also indicate that high arousal corresponds to stronger contraction of the associated facial muscles. These empirical data seem to indicate that high levels and low levels of arousal have characteristic effects that are contrastive and therefore arousal as a whole is too heterogeneous to be a natural kind.

Second, there is evidence that physiological responses are distinct for levels of arousal. For example, skin conductance responses determined by activity in the sweat glands increase as arousal increases, regardless of whether the emotionally salient pictures are rated as positive or negative (Lang et al. 1993; Bernstein 1969; Maltzman et al. 1971; Bradley et al. 1990; Cook et

al. 1991; Greenwald et al. 1989; Manning & Melchiori 1974; Winton et al. 1984). In other words, subjects produce more sweat when highly aroused and less sweat when they are in states of low arousal. Lang et al. conclude that “skin conductance response increased monotonically with ranked arousal” (1003, 265). Again, there appear to be contrastive properties associated with levels of arousal.

Third, there are distinct behavioral responses correlated with high arousal and contrasting responses to low arousal. For example, measures of attention, leading to orienting behaviors, are related primarily to arousal levels (Lang et al. 1993). Subjects tend to exhibit orienting behaviors to highly arousing emotional stimuli. Additionally, Bradley et al (1992) showed that memory is better for pictures that are rated as highly arousing.<sup>46</sup> Gold and McGaugh’s (1975) study also supports the claim that memory increases for highly arousing stimuli. Gold and McGaugh manipulated arousal levels in animal subjects and showed that they also demonstrated better memory of training stimuli when the arousal levels were high. These properties are once again clustered around the extremes of arousal and not arousal as a whole. Even though arousal states are not homogeneous, high arousal and low arousal may be. In sum, the limited empirical data suggest that the extremes of arousal may each be associated with a distinct cluster of properties and may therefore be natural kinds. Again, the causal (homeostatic) mechanism is key and so I turn to that requirement below.

#### **4.1.3.2- Shared Causal Mechanism?**

Causal mechanisms related to arousal are much more difficult to distinguish, but given what was said in the previous section, it seems likely that there are (at least) two distinct causal mechanisms at work in the category of arousal. One study did find that a patient who had

---

<sup>46</sup> It should be noted that this is not the case for pathologically fearful subjects. When these subjects are highly aroused they have poor memories.

portions of the right amygdala removed consistently reported lower arousal levels (Bradley and Lang 2000, 259). So it might be the case that the right hemisphere is correlated with arousal, but this is a tentative conclusion.

The problem with finding a causal mechanism might be attributed to the fact that scientists are not looking for the right kind of causal mechanism. Causal mechanisms are assumed to be found in regions or areas in the brain, but there may be good reasons to think that this will not be the case for arousal. All of the measurable effects associated with arousal involve visceral properties and the effects are diverse and widespread. Given such diverse effects it seems unlikely that one brain area or region would be correlated with arousal. It may be the case that scientists need to search for a causal mechanism for arousal at a different level of analysis. Perhaps the causal mechanism for arousal is better thought of as a change in hormone levels or neurotransmitters. Further empirical research needs to be done, but if clusters of properties are indicative of an underlying causal mechanism (this is usually assumed to be the case), then in all likelihood there are different causal mechanism for high and low arousal.

To summarize, the limited empirical data suggest that arousal is too heterogeneous to be a natural kind. Instead, homogeneity appears at the extremes of high arousal and low arousal. If X is a high arousal state, then X will tend to exhibit increased tension in facial muscles, increased production in the sweat glands, and orientating behavior. On the other hand, if X is a low arousal state, then X will tend to exhibit decreased tension in facial muscles, decreased production in the sweat glands, and less orientating behavior. Thus, high arousal and low arousal may both constitute natural kinds, but the evidence suggests that arousal as an entire class does not.

#### 4.1.4- Is Core Affect a Natural Kind?

I have now evaluated whether valence and arousal are natural kinds. Could it be the case that core affect (the combination of valence and arousal) may be a natural kind? In order for core affect to be a natural kind all instances ranging from positive valence-low arousal to positive valence-high arousal to negative valence-low arousal to negative valence-high arousal would have to tend to share a family of co-occurring properties and a causal (homeostatic) mechanism that brings about their co-occurrence non-accidentally. In other words, it would have to be the case that core affect fulfilled requirements 1 and 2 for natural kind status. Under this view, the circumplex as a whole would constitute a visual representation of a natural kind.

As was shown in the previous two sections of this chapter, both valence and arousal are too heterogeneous to be natural kinds. There are distinct *contrastive* properties corresponding to the extremes of valence and arousal. It seems to follow that core affect is most likely not a natural kind because combining the heterogeneity of valence and arousal leads to even more heterogeneity. For example, instances of negatively valenced-highly aroused states are associated with an increase in autonomic activity and orientating behavior, whereas instances of positively valenced- low aroused states are associated with a decreases in autonomic activity and a lack of orientating behavior. These are heterogeneous properties that cannot be attributed to core affect as part of the natural property cluster. Decreased health benefits associated with negative valence and increased health benefits associated with positive valence is another heterogeneous property of core affect. Several other heterogeneous properties associated with the core affect category were discussed in sections 4.1.2 and 4.1.3. Thus, there is not a family of often co-occurring properties for core affect and core affect fails to fulfill requirement 1 for a natural kind *sensu* Boyd. Likewise there are multiple causal mechanisms associated with core affect. For example,

right-hemisphere activation is associated with negative valence and left-hemisphere activation is associated with positive valence. Thus core affect fails to fulfill requirement 2.

To answer the question of this thesis (and this chapter): *No*, core affect is *not* a natural kind. The properties exhibited by core affective states are too heterogeneous for the entire class of core affect to form a natural kind. Under the view that the set of all core affective states is a natural kind the entire circumplex is a visual representation of a single natural kind. This is not the case. If any natural kinds are visually represented on the circumplex (i.e. if core affect remains a fruitful methodology for studying the emotions), then they must be sub-regions of the circumplex. In fact, empirical data appear to support the hypothesis that subsets of core affect are natural kinds.

#### **4.1.5- Are Subsets of Core Affect Natural Kinds?**

Even though core affect may not be a natural kind, there is a sense in which Barrett's key hypothesis may be correct- *certain sets* of core affective states are likely natural kinds. Core affective states are varied combinations of valence and arousal levels. The empirical evidence supports the hypothesis that the set of states corresponding to the four quadrants of the circumplex are natural kinds. That is, four subsets of core affective states appear to be natural kinds: the set of negatively valenced- high arousal states, the set of positively valenced- high arousal states, the set of negatively valenced- low arousal states, and the set of positively valenced- low arousal states. These four seem to fulfill requirements 1 and 2 for natural kind status. This claim is a consequence of the evidence presented in sections 4.1.2 and 4.1.3. Table 1 below compiles all of the data provided in these previous sections. The table maps directly onto the four quadrants of the circumplex. I strongly encourage the reader to compare table 1 to the core affective circumplex on page 10.

Table 1

<b>Negative Valence/High Arousal</b>	<b>Positive Valence/ High Arousal</b>
<p data-bbox="272 289 667 317"><u>Cluster of Properties (Requirement 1):</u></p> <ul data-bbox="191 321 748 940" style="list-style-type: none"> <li>• Increased activation in brow area and corrugator (frown) facial muscle (in adults and infants) (-V)</li> <li>• Increased tension in corrugator (frown) facial muscle (↑ A)</li> <li>• General increase in autonomic nervous system activity, but significant heart rate deceleration (-V)</li> <li>• Increased production in sweat glands (↑ A)</li> <li>• Increase in number of blink responses and shorter eye blinks (-V)</li> <li>• Subjects rate emotionally salient pictures less interesting, but choose to view them in longer time periods (-V)</li> <li>• Subjects attend more to emotionally salient pictures (-V)</li> <li>• Allocation of more attention to emotionally salient picture- Orientating behavior towards more- Increased memory allocation (↑ A)</li> <li>• Decreased health benefits, such as shorter life spans and worse overall mental and physical health (-V)</li> </ul> <p data-bbox="233 945 708 972"><u>Possible Causal Mechanism (Requirement 2):</u></p> <ul data-bbox="191 976 727 1024" style="list-style-type: none"> <li>• In general, more right-hemisphere activation (-V)</li> </ul>	<p data-bbox="865 289 1260 317"><u>Cluster of Properties (Requirement 1):</u></p> <ul data-bbox="784 321 1341 915" style="list-style-type: none"> <li>• Increased activation in zygomatic (smile) facial muscle (in adults and infants) (+V)</li> <li>• Increased tension in zygomatic (smile) facial muscle (↑ A)</li> <li>• General decrease in autonomic nervous system activity, but greater peak heart rate acceleration (+V)</li> <li>• Increased production in sweat glands (↑ A)</li> <li>• Decrease in number of blink responses and quicker eye blinks (+V)</li> <li>• Subjects rate emotionally salient pictures more interesting, but choose to view then in shorter time periods (+V)</li> <li>• Subjects attend less to emotionally salient pictures (+V)</li> <li>• Allocation of more attention to emotionally salient pictures- Orientating behavior- Increased memory allocation (↑ A)</li> <li>• Increased health benefits, such as longevity and better overall mental and physical health (+V)</li> </ul> <p data-bbox="826 976 1300 1003"><u>Possible Causal Mechanism (Requirement 2):</u></p> <ul data-bbox="784 1008 1320 1035" style="list-style-type: none"> <li>• In general, more left-hemisphere activation (+V)</li> </ul>
<b>Negative Valence/Low Arousal</b>	<b>Positive Valence/Low Arousal</b>
<p data-bbox="272 1071 667 1098"><u>Cluster of Properties (Requirement 1):</u></p> <ul data-bbox="191 1102 748 1724" style="list-style-type: none"> <li>• Increased activation in brow area and corrugator (frown) facial muscle (in adults and infants) (-V)</li> <li>• Less tension in brow area and corrugator (frown) facial muscle (↓ A)</li> <li>• General increase in autonomic nervous system activity, but significant heart rate deceleration (-V)</li> <li>• Decreased production in sweat glands (↓ A)</li> <li>• Increase in number of blink responses and shorter eye blinks (-V)</li> <li>• Subjects rate emotionally salient pictures less interesting, but choose to view them in longer time periods (-V)</li> <li>• Subjects attend more to emotionally salient pictures (-V)</li> <li>• Allocation of less attention to emotionally salient picture- Lack of orientating behavior ----- Decreased memory allocation (↓ A)</li> <li>• Decreased health benefits, such as shorter life spans and worse overall mental and physical health (-V)</li> </ul> <p data-bbox="233 1728 708 1755"><u>Possible Causal Mechanism (Requirement 2):</u></p> <ul data-bbox="191 1759 691 1808" style="list-style-type: none"> <li>• In general, more right-hemisphere activation (-V)</li> </ul>	<p data-bbox="865 1071 1260 1098"><u>Cluster of Properties (Requirement 1):</u></p> <ul data-bbox="784 1102 1341 1696" style="list-style-type: none"> <li>• Increased activation in zygomatic (smile) facial muscle (in adults and infants) (+V)</li> <li>• Less tension in zygomatic (smile) facial muscle (↓ A)</li> <li>• General decrease in autonomic nervous system activity, but greater peak heart rate acceleration (+V)</li> <li>• Decreased production in sweat glands (↓ A)</li> <li>• Decrease in number of blink responses and quicker eye blinks (+V)</li> <li>• Subjects rate emotionally salient pictures more interesting, but choose to view then in shorter time periods (+V)</li> <li>• Subjects attend less to emotionally salient pictures (+V)</li> <li>• Allocation of less attention to emotionally salient picture- Lack of orientating behavior (↓ A)- Decreased memory allocation (↓ A)</li> <li>• Increased health benefits, such as longevity and better overall mental and physical health (+V)</li> </ul> <p data-bbox="826 1755 1300 1782"><u>Possible Causal Mechanism (Requirement 2):</u></p> <ul data-bbox="784 1787 1320 1814" style="list-style-type: none"> <li>• In general, more left-hemisphere activation (+V)</li> </ul>

In light of the evidence there seem to be four natural kinds. That is, negative valence/high arousal, positive valence/high arousal, negative valence/low arousal, and positive valence/low arousal all appear to share their own a family of co-occurring properties and a causal (homeostatic) mechanism that brings about their co-occurrence non-accidentally. For example, the set of positively valenced and low arousal states is associated with increased activation in the zygomatic (smile) facial muscle, a general decrease in autonomic activity, lack of orientating behavior, and more left-hemisphere activation. From this information, scientists may be able to reliably predict that these properties will be observed when a subject is either viewing a picture rated as positively valenced and low arousal or experiencing this core affective state. Thus, the set of positively valenced and low arousal states would seem to support some preliminary generalizations.

It should also be noted that in so far as discrete emotions can be plotted on the circumplex they are likewise amenable to generalizations phrased in terms of valence and arousal. Since *fear*, *anger*, and *disgust* map onto the negative valence/high arousal quadrant of the circumplex, we can expect them to be associated with increased activation and tension in the corrugator (frown) facial muscle, orientating behavior, a general increase in autonomic activity, and more right-hemisphere activation. Thus, preliminary generalizations and predictions retain the basicness of valence and arousal, and apply to some discrete emotions.

The claim that there may be four subsets of core affective states that are natural kinds is an empirical hypothesis and so is revisable in light of further empirical data. In fact, we should expect that these proposed four natural kinds will be revised. At present we have only a rudimentary understanding of the causal mechanisms. We can localize a causal mechanism for positive valence somewhere in the left-hemisphere and for negative valence somewhere in the

right-hemisphere. The causal mechanism is even less clear for arousal. Generalizations and predictions will become more powerful as we gain a further understanding of the causal mechanisms underlying these natural kinds.

## CHAPTER 5: CONCLUSION

To summarize, I have explained a new and interesting theory of the emotions: core affect (Ch.2). I then presented Barrett's arguments against the natural kind status of discrete emotions and in favor of the natural kind status of core affect (Ch.3). I then argued that while core affect is likely *not* a natural kind, four subsets of core affective states (negative valence/high arousal, positive valence/high arousal, negative valence/low arousal, and positive valence/low arousal) may be natural kinds (Ch.4). The empirical evidence presented in table 1 supports my argument: there is a family of often co-occurring properties and a causal (homeostatic) mechanism that brings about the co-occurrence non-accidentally for each of these subsets of core affect. Thus, there appear to be four natural kinds within the core affective category. I would like to end with a brief discussion of what I think core affect can add to emotion theory and a general lesson for emotion theorists looking for natural kinds: splitting proposed kinds is the best avenue for discovering natural kinds.

As I previously mentioned, the dominant paradigm in emotion theory is the discrete emotion approach and the central concern has been locating natural kinds of discrete emotions. I think that there is something that discrete emotion theorists can take from this notion of splitting core affect and the tentative correlation between arousal and valence. One of the main reasons Barrett claims that the discrete emotions are not natural kinds is because they assume that folk categories of emotions, such as "anger," *must* corresponds to a scientific category, *anger*. Typically folk terms must be refined so as to fit the causal structure of the world. The discrete emotion theorists

seem to have forgotten that in order to find a natural kind we must fit our terms to the causal structure of the world (for a similar argument see Griffiths 2004a). Perhaps it is time for discrete emotion theorists to take seriously Russell's suggestion that arousal levels may distinguish a number discrete emotions that are otherwise difficult to separate scientifically (such as *sadness*, *disgust*, *contempt*, *anger*).

A similar moral ought to be drawn by core affect theorists. Even when we start with "scientific" terms or concepts, we still may need to revise them in light of empirical data. Our goal is to make our terms fit the causal structure of the world if we want to find natural kinds *sensu* Boyd. Core affect theorists appear to be right that valence and arousal are important in understanding the emotions, but their own research suggests that core affect is not a natural kind. The evidence suggests that core affect needs to be further refined if it is to be scientifically fruitful as a natural kind.

Several philosophical points follow from my tentative conclusion that there are four potential natural kinds within the core affective category. First, it may often be the case that splitting a proposed kind is more likely to yield natural kinds. Second, philosophy has more to offer scientists than just conceptual analysis. Of course this is an important part of philosophy of science. I have pointed to several problems with the core affect theory (see chapter 2). But what my tentative conclusion shows is that philosophy can also contribute and direct future scientific research by a careful analysis not only of the theory, but also of the empirical data. Given my conclusion further research on the core affective category should focus on the extremes of the dimensions and further correlating these extremes. Thus, philosophy can assist in directing scientific research not only by arm chair analysis or by using anecdotal evidence in arguments, but also by getting our hands dirty and really analyzing the available evidence.

In conclusion, the best shot at locating natural kinds is empirically reformed revisions of the category. In the case of core affect, the category appears to need to be split. Four subsets of core affect are likely natural kinds. That is, the set of negative valence-high arousal states may be a natural kind, as well as each of the other three core affective states corresponding to the quadrants of the circumplex. It may also be the case that some of the discrete emotions need to be split, perhaps on the basis of differing arousal levels. If we are after natural kinds, then we must fit our categories to what Boyd calls the sometimes messy and complex causal structure of the world (1991, 143). This may inevitably mean splitting what we thought before was a natural kind. This is surely a small price to pay for the potential consequence of finding new categories suitable for scientific study and analysis.

## BIBLIOGRAPHY

- Adams, R.B., Gordon, H.L., Baird, A.A., Ambady, N., and Kleck, R.E. "Effects of gaze on amygdala sensitivity to anger and fear faces." *Science* 300 (2003): 1536-1537.
- Adolphs, R., Gosselin, F., Buchanan, T.W., Tranel, D., Schyns, P. and Damasio, A.R. "A mechanism for impaired fear recognition after amygdala damage." *Nature* 433 (2005): 68-72.
- Aristotle. "Metaphysics." *The Basic Works of Aristotle*. Ed. R. McKeon et al. New York: Random House, 1941. 682-926.
- Arnold, M.B. *Emotion and Personality*. New York: Columbia University Press, 1960.
- Balaban, M.T. "Affective influences on startle in five-month-old infants: reactions to facial expressions of emotion." *Child Development* 66 (1995): 28-36.
- Bargh, J.A., Chaiken, S., Govender, R., and Pratto, F. "The generality of the automatic attitude activation effect." *Journal of Personality and Social Psychology* 62.6 (1992): 893-912.
- Bargh, J.A., Chaiken, S., Raymond, P., and Hymes, C. "The automatic evaluation effect: unconditioned automatic attitude activation with a pronunciation task." *Journal of Experimental Social Psychology* 32 (1996): 104-126.
- Bargh, J.A. "The automaticity of everyday life." *The automaticity of everyday life: Advances in social cognition*. 10 Ed. R.S. Wyer Jr. Mahwah, NJ: Erlbaum, 1997, 1-61.
- Barlow, D.H. *Anxiety and its Disorders: The Nature and Treatment of Anxiety and Panic*. New York: Guilford, 2002.
- Barrett, Lisa Feldman. "Hedonic tone, perceived arousal, and item desirability: three components of self-reported mood." *Cognition & Emotion* 10.1 (1996): 47-68.
- Barrett, Lisa Feldman. "The relationship among momentary emotion experiences, personality descriptions, and retrospective ratings of emotion." *Personality and Social Psychology Bulletin* 23.10 (1997): 1100-14.
- Barrett, Lisa Feldman and Russell, James A. "Independence and bipolarity in the structure of current affect." *Journal of Personality and Social Psychology* 74.4 (1998): 967-984.
- Barrett, Lisa Feldman and Russell, James A. "Core Affect, Prototypical Emotional Episodes, and Other Things Called Emotion: Dissecting the Elephant." *Journal of Personality and Social Psychology* 76.5 (1999a): 805-819.
- Barrett, Lisa Feldman and Russell, James A. "The Structure of Current Affect: Controversies and Emerging Consensus." *Current Directions in Psychological Science* 8.1 (February 1999b): 10-14.
- Barrett, Lisa Feldman and Fossium, Thyra. "Mental representations of affect knowledge." *Cognition and Emotion* 15.3 (2001): 333-363.
- Barrett, Lisa Feldman. "Feelings or words? Understanding the content in self-report ratings of emotional experience." *Journal of Personality and Social Psychology* 87 (2004a): 266-281.
- Barrett, L. F., Quigley, K., Bliss-Moreau, E., and Aronson, K. R. "Arousal focus and interoceptive sensitivity." *Journal of Personality and Social Psychology* 87 (2004b), 684-697.
- Barrett, Lisa Feldman. "Solving the emotion paradox: Categorization and the experience

- of emotion.” *Personality and Social Psychology Review* 10 (2006a): 20-46.
- Barrett, Lisa Feldman. “Are Emotions Natural Kinds?” *Perspectives on Psychological Science* 1.1 (2006b): 28-58.
- Barrett, Lisa Feldman and Wager, Tor D. “The Structure of Emotion: Evidence from Neuroimaging Studies.” *Current Directions in Psychological Science* 15.2 (2006c): 79-83.
- Barrett, Lisa Feldman. “Valence as a basic building block of emotional life.” *Journal of Research in Personality* 40 (2006d): 35-55.
- Barrett, Lisa Feldman, Mesquita, Batja, et al. “The Experience of Emotion.” *Annual Review of Psychology* 58 (2007a): 373-403.
- Barrett, Lisa Feldman, Bliss-Moreau, Eliza, et al. “The Amygdala and the Experience of Affect.” *Social, Cognitive, and Affective Neuroscience* 2 (2007b): 73-83.
- Barrett, L. F., Lindquist, K., Bliss-Moreau, E., Duncan, S., Gendron, M., Mize, J., & Brennan, L. “Of mice and men: Natural kinds of emotion in the mammalian brain?” *Perspectives on Psychological Science* 2 (2007c): 297-312.
- Ben-Ze’ev, Aaron. “Emotion as a Subtle Mental Mode.” *Thinking about Feeling: Contemporary Philosophical on Emotions*. Ed. Robert Solomon. New York: Oxford University Press, 2004. 250-268.
- Bernstein, A.S. “The orienting response and direction of stimulus change.” *Psychonomic Science* 12 (1969): 127-128.
- Blascovich, J. and Mendes, W.B. “Challenge and threat appraisals: the role of affective cues.” *Feeling and Thinking: The role of affect in social cognition*. Ed. J.P. Forgas. New York: Cambridge University Press, 2000. 59-82.
- Block, Ned. “On Confusion about a Function of Consciousness.” Reprinted in *The Nature of Consciousness: Philosophical Debates* (1997): 375- 415.
- Bouton, M.E. “Behavior systems and the contextual control of anxiety, fear, and panic.” *Emotions: Conscious and Unconscious*. Eds L.F. Barrett et al. New York: Guilford, 2005. 205-227.
- Boyd, Richard. “Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds.” *Philosophical Studies* 61 (1991): 127-148.
- Boyd, Richard. Kinds. “Complexity and Multiple Realization.” *Philosophical Studies* 95 (1999): 67-98.
- Bradley, M.M., Cuthbert, B.N., and Lang P.J. “Startle reflex modification: Emotion or attention?” *Psychophysiology* 27 (1990): 513-522.
- Bradley, M.M., Greenwald, M.K., Petry, M., and Lang, P.J. “Remembering pictures: pleasure and arousal in memory.” *Journal of Experimental Psychology: Learning, Memory and Cognition* 18 (1992): 379- 390.
- Bradley, M.M. and Lang, P.J. “Measuring emotion: behavior, feeling, and physiology.” *Cognitive Neuroscience of Emotion*. Eds. Richard D. Lane and Lynn Nadel. Oxford: Oxford University Press, 2000. 242-276.
- Cacioppo, J.T., Perry, R.E., Losch, M.E., and Kim, H.S. “Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions.” *Journal of Personality and Social Psychology* 50 (1986): 260-268.
- Cacioppo, J.T., Berntson, G.G., Klein, D.J., and Poehlmann, K.M. “The psychophysiology of emotion across the lifespan.” *Annual Review of Gerontology and Geriatrics* 17 (1997): 27-74.

- Cacioppo, J.T. and Gardner, W.L. "Emotion." *Annual Review of Psychology* 50 (1999): 191-214.
- Cacioppo, JT, Berntson, GG, Larsen, JT, Poehlmann, KM, and Ito, TA. "The psychophysiology of emotion." *Handbook of Emotion*. 2<sup>nd</sup> Edition. Eds. M. Lewis and JM Haviland-Jones. New York: Guilford Press, 2000, 173-191.
- Cannon, WB. "The James-Lange theory of emotion: A critical examination and an alternative theory." *American Journal of Psychology* 39 (1927): 106-124.
- Caron, R.F., Caron, A.J., and Myers, R.S. "Do infants see emotional expressions in static faces?" *Child Development* 56 (1985): 1552-1560.
- Chalmers, D.J. "What is a neural correlate of consciousness?" *Neural Correlates of Consciousness: Empirical and Conceptual Questions*. Ed. T. Metzinger. Cambridge: MIT Press, 2000. 17-40.
- Chalmers, D.J. "First-person methods in the sciences of consciousness." *Consciousness Bulletin: Center for Consciousness Studies* University of Arizona (1999).
- Chalmers, D.J. "On the search for the neural correlates of consciousness." *Towards a science of consciousness II: The Second Tuscan Discussions and Debates*. Eds. S. Hameroff et al. Cambridge: MIT Press, 1998. 219-230.
- Charland, Louis C. "The Natural Kind Status of Emotion." *British Society for the Philosophy of Science* 53 (2002): 511-537.
- Charland, Louis C. "The Heat of Emotion: Valence and the Demarcation Problem." *Journal of Consciousness Studies* 12.8 (2005): 82-102.
- Cook, E.W., Hawk, L.H., David, T.L., and Stevenson, V.E. "Affective individual differences and startle reflex modulation." *Journal of Abnormal Psychology* 100 (1991): 5-13.
- Corr, Philip J., Kumari, Venna, et al. "Harm avoidance and affective modulation of the startle reflex: a replication." *Person. Individ. Diff.* 22.4 (1997): 591-593.
- Craver, Carl F. "Dissociable Realization and Kind Splitting." *Philosophy of Science* 71 (December 2004): 960-971.
- Damasio, A.R. *The feeling of what happens: Body and emotion in the making of consciousness*. New York: Harcourt Brace, 1999.
- Damasio, A.R. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: G.P. Putnam's Sons, 1994.
- Danner, D.D., Snowdon, D.A., and Friesen, W.V. "Positive emotions in early life and longevity: Findings from the nun study." *Journal of Personality and Social Psychology* 80 (2001): 804-813.
- Darwin, C. *The Expression of the Emotions in Man and Animals*. 1859. Chicago: University of Chicago Press, 1965.
- Davidson, RJ. "The functional neuroanatomy of affective style." *Cognitive neuroscience of emotion*. Eds. RD Lane and L Nadel New York: Oxford University Press, 2000. 371- 388.
- Duncan, Seth and Barrett, Lisa Feldman. "Affect is a form of cognition: a neurobiological analysis." *Cognition and Emotion* 21.6 (2007): 1184-1211.
- Duffy, E. "An explanation of 'emotional' phenomena without the use of the concept 'emotion'." *Journal of General Psychology* 25 (1941): 283- 293.
- Ekman, Paul. "Constants across cultures in the face and emotion." *Journal of Personality and Social Psychology* 17.2 (1971): 124-129.

- Ekman, P. and Friesen, W. "Measuring facial movement." *Environmental Psychology and Nonverbal Behavior* 1 (1976): 56-75.
- Ekman, P., Levenson, R.W., and Friesen, W.V. "Autonomic nervous system activity distinguishes between emotions." *Science* 221 (1983): 1208-1210.
- Ekman, P. "An argument for basic emotions." *Cognition and Emotion* 6 (1992): 169-200.
- Ekman, Paul. "Basic Emotions." *Handbook of Cognition and Emotion*. Eds. T. Dalgleish and M.J. Power. West Sussex: John Wiley & Sons Ltd, 1999b. 45-59.
- Ekman, Paul. "Facial Expressions." *Handbook of Cognition and Emotion*. Eds. T. Dalgleish and M.J. Power. West Sussex: John Wiley & Sons Ltd, 1999a. 301-320.
- Elfenbein, H.A. and Ambady, N. "Universal and cultural differences in recognizing emotions." *A meta-analysis: Psychological Bulletin* 128 (2002): 203-235.
- Elfenbein, H.A. and Ambady, N. "Universals and cultural differences in recognizing emotions." *Current Directions in Psychological Science* 12 (2003): 159-164.
- Ellsworth, P.C. and Scherer, K.R. "Appraisal processes in emotion." *Handbook of Affective Sciences*. Ed. R.J. Davidson et al. New York: Oxford University Press, 2003. 572- 595.
- Emde, RN, et al. "Emotional expressions in infancy: a biobehavioral study." *Psychological Issues* 10 (1976):1-200.
- Fazio, R.H., Sanbonmatsu, D.M., Powell, M.C., and Kardes, F.R. "On the automatic activation of attitudes." *Journal of Personality and Social Psychology* 50.2 (1986): 229-238.
- Fehr, B. and J. A. Russell. "Concept of emotion viewed from a prototype perspective." *Journal of Experimental Psychology: General* 113 (1984): 464- 86.
- Fodor, J.A. *The Modularity of Mind*. Cambridge: The MIT Press, 1983.
- Frith, C.D. and Allen, H.A. "The skin conductance orientating response as an index of attention." *Biological Psychiatry* 12 (1983): 27-39.
- Fridlund, A.J., Schwartz, G.E., and Fowler, S.C. "Pattern recognition of self-reported emotional state from multiple-site facial EMG activity during affective imagery." *Psychophysiology* 21 (1984): 622-637.
- Frijda, N.H.. *The Emotions*. Cambridge: Cambridge UP, 1986.
- Frijda, N.H. and Tcherassof, A. "Facial expressions as modes of action readiness." *The Psychology of Facial Expression*. Eds. JA Russell and JM Fernandez-Dols. New York: Cambridge UP, 1997, 78-102.
- Frijda, N.H. "The psychologists' point of view." *Handbook of Emotions*. 2<sup>nd</sup> ed. Eds. M Lewis and J Haviland. New York: Guilford, 2000. 59-74.
- Gold, P.E. and McGaugh, J.L. "A single-trace, dual-process view of memory storage processes." *Short-term memory*. Eds. D. Deutsch and J.A. Deutsch New York: Academic Press, 1975. 51-64.
- Greenwald, M.K., Cook, E.W., and Lang, P.J. "Affective judgment and psychophysiological response: dimensional covariation in the evaluation of pictorial stimuli." *Journal of Psychophysiology* 3 (1989): 51-64.
- Griffiths, P. E. *What Emotions Really Are: The Problem of Psychological Categories*. Chicago: University of Chicago Press, 1997.
- Griffiths, P. E. "Is Emotion a Natural Kind?" *Thinking about Feeling: Contemporary Philosopher on Emotions*. Ed. Robert Solomon. New York: Oxford University Press, 2004a. 233-249.

- Griffiths, P.E. "Emotions as natural and normative kinds." *Philosophy of Science* 71.5 (2004b): 901-911.
- Hansen, C.H. and Hansen, R.D. "Finding the face in the crowd: an anger superiority effect." *Journal of Personality and Social Psychology* 54 (1988): 917- 924.
- Hariri Ar, Drabant EM, Munoz KE, Kolachana BS, Mattay VS, et al. "A susceptibility gene for affective disorders and the response of the human amygdala." *Archives of General Psychiatry* 62 (2005): 146-52.
- Hacking, Ian. "A Tradition of Natural Kinds." *Philosophical Studies* 61 (1991): 109-126.
- Hebb, DO. "Drives and the CNS (conceptual nervous system)." *Psychological Review* 62 (1955): 243- 254.
- Izard, CE. *Human Emotions*. New York: Plenum, 1977.
- Izard, CE. "Four systems for emotion activation: cognitive and noncognitive processes." *Psychological Review* 100 (1993): 68-90.
- Izard, C.E. and Ackerman, B.P. "Motivational, organizational, and regulatory functions of discrete emotions." *Handbook of Emotions*. 2<sup>nd</sup> ed. Eds. M Lewis and J Haviland. New York: Guilford, 2000. 253-264.
- Izard, CE. "Basic emotions, natural kinds, emotion schemas, and a new paradigm." *Perspectives on Psychological Science* 2.3 (2007): 260-275.
- James, William. "What is an Emotion?" *Mind* 9 (1884): 188-205.
- Kripke, Saul. *Naming and Necessity*. Harvard: Harvard University Press, 1980.
- Lang, P.J., Greenwald, M.K., Bradley, M.M., and Hamm, A.O. "Looking at pictures: Affective, facial, visceral, and behavioral reactions." *Psychophysiology* 30 (1993): 261-273.
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). *International affective picture system (IAPS): Technical manual and affective ratings*. Gainesville: University of Florida, Center for Research in Psychophysiology.
- LaPorte, J. *Natural Kinds and Conceptual Change*. Cambridge: Cambridge University Press, 2003.
- Lazarus, R.S. "Progress on a cognitive-motivational-relational theory of emotion." *American Psychologist* 46.8 (1991): 819-834.
- Leary, T. *Interpersonal diagnosis of personality*. New York: Ronald, 1957.
- LeDoux, J.E., Sakaguchi, A. and Reis, D.J. "Subcortical efferent projections of the medial geniculate nucleus mediate emotional responses conditioned by acoustic stimuli." *Journal of Neuroscience* 4 (1984): 683-698.
- LeDoux, J.E. *The Emotion Brain: The Mysterious Underpinnings of Emotional Life*. New York: Simon and Schuster, 1996.
- LeDoux, J.E. and Phelps, E.A.. "Emotional Networks in the Brain." *Handbook of Emotions*. 2<sup>nd</sup> ed. Eds. M Lewis and J Haviland. New York: Guilford, 2000. 157-172.
- Locke, J. *An Essay Concerning Human Understanding*. 1690. Reprinted. New York: Prometheus Books, 1995.
- Lyubomirsky, S., King, L., and Diener, E. "The benefits of frequent positive affect: Does happiness lead to success." *Psychological Bulletin* 131 (2005): 803-855.
- Machery, Edouard. "How to Split Concepts: Reply to Piccinini and Scott." Forthcoming in *Philosophy of Science*.
- Machery, Edouard. "Concepts are not a Natural Kind." *Philosophy of Science* 72 (July

- 2005): 444-467.
- Maltzman, I., Harris, L., Ingram, E., and Wolff, C. "A primacy effect in the orienting reflex to stimulus change." *Journal of Experimental Psychology* 87 (1971): 202-206.
- Manning, S.K., and Melchiori. "Words that upset urban college students: measured with GSRs and rating scales." *Journal of Social Psychology* 94 (1974): 305-306.
- Maurta, T., Colligan, R.C., Malinchoc, M., and Offord, K.P. "Optimists vs. pessimists: Survival rate among medical patients over a 30-year period." *Mayo Clinic Proceedings* 75 (2000): 140-143.
- Mead, G.H. *Mind, self and society*. Chicago: University of Chicago Press, 1934.
- Messinger, D.S. "Positive and negative: infant facial expressions and emotions." *Current Directions in Psychological Science* (2002): 1-6.
- Mesquita, Batja. "Emotions as Dynamic Cultural Phenomena." *Handbook of Affective Sciences*. Ed. R. Davidson et al. New York: Oxford University Press, 2003. 871-890.
- Mill, J.S. *A System of Logic*. 1843. London: Routledge University Press, 1905.
- Millikan, Ruth Garrett. "Historical Kinds and The 'Special Sciences'." *Philosophical Studies* 95 (1999): 45-65.
- Morris, M., Bradley, M., Bowers, D., Lang, P., and Heilman, K. "Valence-specific hypoarousal following right temporal lobectomy." *Journal of Clinical and Experimental Neuropsychology* 13 (1991):42.
- Moruzzi, G. and Magoun, H.W. "Brain stem reticular formation and activation of the EEG." *Electroencephalography and Child Neurophysiology* 1 (1949): 455-473.
- Murphy, S.T., and Zajonc, R.B. "Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures." *Journal of Personality and Social Psychology* 64 (1993): 723- 739.
- Murphy, F.C., Nimmo-Smith, I., and Lawrence, A.D. "Functional neuroanatomy of emotion: A meta-analysis." *Cognitive, Affective, and Behavioral Neuroscience* 3 (2003): 207-233.
- Nussbaum, M. *Upheavals of Thought: The Intelligence of the Emotions*. Cambridge: Cambridge University Press, 2001.
- Nussbaum, M. "Emotions as Judgments of Value and Importance." *Thinking about Feeling: Contemporary Philosophical on Emotions*. Ed. Robert Solomon. New York: Oxford University Press, 2004. 181-199.
- Osgood, CE, May, WH, & Miron, MS. *Cross-cultural universals of affective meaning*. Urbana: University of Illinois Press, 1975.
- Ortony, A. and Turner, T.J. "What is basic about basic emotions?" *Psychological Review* 97 (1990): 315- 331.
- Panksepp, J. *Affective Neuroscience: The Foundations of Human and Animal Emotions*. New York: Oxford University Press, 1998.
- Panksepp, J. "Emotions as Natural Kinds within the Mammalian Brain." *Handbook of Emotions*. 2<sup>nd</sup> ed. Eds. M Lewis and J Haviland. New York: Guilford, 2000. 137-156.
- Panksepp, J. "Affective consciousness: core emotional feelings in animals and humans." *Consciousness and Cognition* 14 (2005): 30-80.
- Panksepp, J. "Neurologizing the Psychology of Affects." *Perspectives on Psychological Science* 2.3 (2007): 281-296.

- Peterson, C., Seilgman, M., and Vaillant, G. "Pessimistic explanatory style is a risk factor for physical illness: A thirty-five year longitudinal study." *Journal of Personality and Social Psychology* 55 (1988): 23-27.
- Phan, K.L., Wager, T.D., Taylor, S.F., and Liberzon, I. "Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI." *NeuroImage* 16 (2002): 331-348.
- Piccinini, Gualtiero and Scott, Sam. "Splitting Concepts." Forthcoming in *Philosophy of Science*.
- Plato. *The Statesman*. Trans. Julia Annas. New York: Cambridge University Press, 1995.
- Plato. *The Republic*. Trans. Benjamin Jowett. New York: Dover Publications, 2000.
- Plutchik, R. *Emotion: A Psychoevolutionary Synthesis*. New York: Harper & Row, 1980.
- Posner, J., Russell, JA., and Peterson, B. "The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology." *Development and Psychopathology* 17 (2005): 715-734.
- Pratto, F. and John, O.P. "Automatic vigilance///; the attention-grabbing power of negative social information." *Journal of Personality and Social Psychology* 61 (1991): 380- 391.
- Prinz, Jesse. "Which Emotions are Basic?" *Emotion, Evolution, and Rationality*. Eds. D Evans and P Cruse. New York: Oxford University Press, 2004a. 69-88.
- Prinz, Jesse. *Gut Reactions: A Perceptual Theory of Emotion*. New York: Oxford University Press, 2004b.
- Prinz, Jesse. "Embodied Emotions." *Thinking about Feeling: Contemporary Philosophers on Emotions*. Ed. Robert Solomon. New York: Oxford University Press, 2004c. 44-60.
- Putnam, Hilary. *Reference and Truth*. Cambridge: Cambridge University Press, 1981.
- Quine, W. "Natural kinds." *Ontological Relativity and Other Essays*. New York: Columbia University Press, 1969. 114-138.
- Reisenzein, Rainer. "Pleasure-Arousal theory and the intensity of emotions." *Journal of Personality and Social Psychology* 67.3 (1994): 525-539.
- Rorty, Amelie Oksenberg. "Enough Already with 'Theories of the Emotions.'" *Thinking about Feeling: Contemporary Philosophers on Emotions*. Ed. Robert Solomon. New York: Oxford University Press, 2004d. 269-278.
- Russell, B. *Human Knowledge: Its Scope and Limits*. London: Allen and Unwin, 1948.
- Russell, JA. "Evidence of convergent validity on the dimensions of affect." *Journal of Personality and Social Psychology* 36 (1978): 1152-1168.
- Russell, JA. "A circumplex model of affect." *Journal of Personality and Social Psychology* 39 (1980): 1161-1178.
- Russell, JA. "Pancultural aspects of human conceptual organization of emotion." *Journal of Personality and Social Psychology* 45 (1983): 1281-88.
- Russell, JA. "Culture and the categorization of emotions." *Psychological Bulletin* 110.3 (1991): 426-450.
- Russell and Lemay. "Emotion concepts." *Handbook of Emotions*. 2<sup>nd</sup> ed. Eds. M Lewis and J Haviland. New York: Guilford, 2000. 491-503.
- Russell, JA, "Core Affect and the Psychological Construction of Emotion." *Psychological Review* 110.1 (2003): 145-172.
- Russell, JA. "Emotion in human consciousness is built on core affect." *Journal of*

- Consciousness Studies* 12.8 (2005): 26-42.
- Saucier, G., Ostendorf, F., and Peabody, D. "The non-evaluative circumplex of personality adjectives." *Journal of Personality* 69.4 (August 2001): 537-582.
- Scherer, K.R. "The role of culture in emotion-antecedent appraisal." *Journal of Personality and Social Psychology*. 73 (1997) 902-22.
- Scherer, K.R. "Appraisal Theory." *Handbook of Cognition and Emotion*. Eds. T. Dalgleish and M.J. Power. West Sussex: John Wiley & Sons Ltd, 1999. 637-663.
- Schachter, S. and Singer, JE. "Cognitive, social, and physiological determinants of emotional state." *Psychological Review* 69 (1962): 379-399.
- Schiffenbauer, A. "Effects of observer's emotional state in judgments of the emotional state of others." *Journal of Personality and Social Psychology* 30 (1974): 31-35.
- Seilgman, M. "Optimism, pessimism, and mortality." *Mayo Clinic Proceedings* 75 (2000): 133-134.
- Seo, M-G., Barrett, Lisa Feldman, and Jin, S. "The Structure of Affect: History, theory, and implications for emotion research in organizations." *Research companions to emotion in organizations*. Eds. C. Cooper and N. Ashkanasy. Northampton, MA: Edwars Elgar, In Press.
- Seyfarth, R.M. and Cheney, D.L. "Meaning and emotion in animal vocalization." *Emotion inside out: 130 years after Darwin's The Expression of the Emotions in Man and Animals (Annals of the New York Academy of Sciences Vol. 1000)*. Ed. P. Ekman et al. New York: New York Academy of Sciences, 2003. 32-55.
- Soloman, Robert. "Emotion, thoughts, and Feelings: Emotions as Engagements." *Thinking about Feeling: Contemporary Philosopher on Emotions*. Ed. Robert Solomon. New York: Oxford University Press, 2004. 76-90.
- Spitz, RA. *The First Year of Life*. New York: Int. University Press, 1965.
- Sroufe, LA. "Socioemotional development." *Handbook of Infant Development*. Ed JD Osofsky, New York: Wiley, 1979. 462-516.
- Strongman, KT. *The Psychology of Emotion*. 5<sup>th</sup> Ed. West Sussex: John Wiley & Sons Ltd, 2003.
- Sullivan, H.S. *The interpersonal theory of psychiatry*. New York: Norton, 1953.
- Swenson, W., Pearson, J., and Osborne, D. *An MMPI source book: Basic item, scale, and pattern data on 50,000 patients*. Minneapolis: University of Minnesota Press, 1973.
- Tooby, J. and Cosmides, L. "Evolutionary Psychology and the Emotions." *Handbook of Emotions*. 2<sup>nd</sup> ed. Eds. M Lewis and J Haviland. New York: Guilford, 2000. 91-115.
- Turati, C. "Why faces are not special to newborns: An alternative account of the face preference." *Current Directions in Psychological Science* 13 (2004): 5-8.
- Venn, J. *The Logic of Chance*. London: Macmillan, 1866.
- Widen, S.C. and Russell, J.A. "A closer look at preschoolers' freely produced labels for facial expressions." *Developmental Psychology* 39 (2003): 114-128.
- Wierzbicka, A. *Emotions across Languages and Cultures*. Cambridge: Cambridge University Press, 1999.
- Wiggins, J.S. "An informal history of the interpersonal circumplex tradition." *Journal of Personality Assessment* 66 (1996): 217-234.
- Winton, W.M., Putnam, L.E., and Krauss, R.M. "Facial and autonomic manifestations of

- the dimensional structure of emotion." *Journal of Experimental Social Psychology* 20 (1984): 195-216.
- Wundt, W.M. *Outlines of Psychology*. Trans. Charles Hubbard Judd. 1897. St Clair Shores, Michigan: Scholarly Press, 1969.
- Wundt, W.M. *Principles of Physiological Psychology*. Trans. Edward Bradford Titchener. 1873. New York: Kraus Reprint Co., 1969.
- Zajonc, RB. "Feeling and thinking: Preferences need no inferences." *American Psychologist* 35 (1980): 151-175.