THE NATURE OF THE EMOTIONS

by

ROBERT VIDIR GUNNARSSON

B.A., University of Iceland, 1993

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

Department

of

Philosophy

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

April 1996

©Robert Vidir Gunnarsson, 1996
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Philosophy

The University of British Columbia
Vancouver, Canada

Date April 26, 1996
ABSTRACT

The question I ask in this thesis is “What is the nature of emotions?” I then argue against essentialist answers to the question and suggest a non-essentialist approach to the study of emotions.

The first chapter is a general introduction, where I try to specify what I mean by the term “emotion”, without assuming anything about their nature. I do that by looking briefly at the general characteristics of some paradigm emotions.

In the second chapter I argue against affect theories about emotions. My arguments turn mostly on the point that mere affects do not have the cognitive depth to account for some of the most conspicuous aspects and characteristics of emotions.

The third chapter is directed against cognitive theories of emotions. I argue that the characteristics of emotions (e.g. intentionality and rationality), which have been taken to necessitate cognitivism, can either be accounted for without beliefs or cognitions or they do not apply to all emotions.

In the fourth chapter I suggest a non-essentialist approach to emotion theory. I make use of Wittgenstein’s notion of family resemblance. I then try to outline the place of cognitive emotions within the family resemblance model and come to the conclusion that a cognitive essentialist model does in fact apply to certain emotion types.
In the fifth and final chapter I ask myself whether anything important can be learnt about emotions from looking at their functions, i.e. at what they do instead of what they are. After looking at two accounts of the functions of emotions, and rejecting them as, at best, incomplete, I come to the conclusion that at least some emotions have the function to serve as commitment devices to carry out rational strategies.
**TABLE OF CONTENTS.**

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td></td>
<td>TABLE OF CONTENTS</td>
<td>iv</td>
</tr>
<tr>
<td></td>
<td>ACKNOWLEDGMENTS</td>
<td>v</td>
</tr>
<tr>
<td>Chapter I</td>
<td>INTRODUCTION.</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>I. 1. What do I Mean by “Emotion”?</td>
<td>2</td>
</tr>
<tr>
<td>Chapter II</td>
<td>EMOTIONS AS AFFECTS.</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>II. 1. Introduction.</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>II. 2. Two Types of Affect Theories.</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>II. 3. Emotions are Independent of Feelings.</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>II. 4. Feelings and the Intentionality of Emotions.</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>II. 5. Feelings and the Rationality of Emotions.</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>II. 6. Feelings and the Individuation of Emotions.</td>
<td>28</td>
</tr>
<tr>
<td>Chapter III</td>
<td>COGNITIVE THEORIES OF EMOTIONS.</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>III. 1. Introduction.</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>III. 2. Beliefs and the Intentionality of Emotions.</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>III. 3. Beliefs and the Rationality of Emotions.</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>III. 4. Beliefs and the Individuation of Emotions.</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>III. 5. Beliefs and Emotions in Conflict.</td>
<td>55</td>
</tr>
<tr>
<td>Chapter IV</td>
<td>NON-ESSENTIALISM ABOUT EMOTIONS.</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>IV. 1. Introduction.</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>IV. 2. Emotions and Family Resemblance.</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>IV. 3. Family Resemblance and the Cognitive Emotions.</td>
<td>66</td>
</tr>
<tr>
<td>Chapter V</td>
<td>FUNCTIONS OF EMOTIONS.</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>V. 1. Introduction.</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>V. 2. Emotions and Instrumental Rationality.</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>V. 2. 1. Ronald de Sousa: Emotions as Patterns of Salience.</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>V. 2. 2. Frijda on Vengeance.</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>V. 2. 3. Emotions as Commitment Devices.</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>A SUMMARY OF CONCLUSIONS.</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>BIBLIOGRAPHY.</td>
<td>107</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

I would like to thank all my roommates and follow philosophers in the House of Philosophy, on 10th Avenue, for countless enjoyable and helpful discussions about issues directly and indirectly relevant to my thesis. I am also grateful to Wilhelm Emilsson who read a completed draft of the thesis and gave countless helpful comments on grammar, style, typos and, most importantly, content. I thank professor Earl Winkler who read the final version. But most of all I thank my supervisor, professor Paul Russell for his constructive criticisms, comments and guidance through the process of writing this thesis.

By the way, all the faults are mine, much of the good stuff theirs.
CHAPTER I.

INTRODUCTION.
I. 1. What do I Mean by “Emotion”?

The main question of this thesis is an old one, posed by a number of philosophers and other thinkers. The question is: What are emotions? The question seems straightforward, but it is in fact not at all clear what states or conditions are being referred to by the term “emotion” and what states are excluded. Before plunging into emotion theory, I am therefore obliged to attempt to answer this terminological question, and I must try to do so without presupposing anything substantial about the nature of the emotions. The phenomena we generally call “emotions” do not form a neatly organized natural kind and it may seem hard to point out criteria which clearly distinguishes emotions from other mental states such as non-emotional beliefs (e.g., that snow is white) or non-emotional sensations (e.g. physical pain) without actually postulating something about the nature of emotions.

One way to set a starting point for an inquiry into the nature of emotions is to list examples of commonly accepted emotions and thus give a rough idea of what class of phenomena is meant by the term “emotion”. This can then be followed by some observations about typical characteristics of, or tentative generalizations about the listed emotions and possibly of what prima facie characteristics distinguish them from other conditions and mental states.

It is perhaps impossible to compile an extensive list of conditions that are undisputedly emotions. But on my list I believe I can fairly safely put conditions like anger, fear, guilt,
hope, pride, shame, indignation and resentment. This is indeed a varied list and somewhat randomly chosen, but few would oppose calling each of these conditions an emotion. Some other conditions, such as love and hate, might raise some question, but there are at least occasional manifestations of both love and hate which must count as emotions, whether or not these are instances of love or hate per se.

The first prima facie characteristic common to these emotions was pointed out by Annette Baier who wrote that emotions,

are different from moods in that they typically have objects, are about something, not everything, while moods, if they are about anything, seem to be about nearly everything.¹

Other philosophers, such as Anthony Kenny, go a step further and claim that emotions do not only typically, but essentially have objects. One gets angry that one’s car was stolen; one hopes that one gets a job; one is afraid of large animals, etc. Resentment and indignation are for something someone did or is believed to have done. It is rather uncommon to be just angry or afraid or indignant without directing one’s anger, fear or indignation at something. This does seem to distinguish emotions from mere objectless feelings, such as the pain from a strained muscle, and from moods, but the question of which states can count as having objects is left unanswered. One of the questions concerning the intentionality of emotions is whether we must follow in the footsteps of Spinoza who wrote:

Modes of thinking such as love, desire, or whatever emotions are designated by name, do not occur unless there is in the same individual the idea of the thing loved, desired, etc.²

¹ Baier, 1990, p. 3.

²
The question here is whether the intentionality of emotions requires that we embrace cognitivism about emotions.

A second characteristic is that all these emotions typically imply or involve evaluations of their objects or some kind of a pro or con attitudes towards them. If one loves or hates, is angry or relieved or hopeful etc. one seems to care one way or another about the object of one’s emotion. When one is angry that one’s car was stolen one wishes, implicitly or explicitly, that it had not been stolen, or thinks in some sense unfavorably of the fact that it was stolen. Negative emotions involve negative evaluations, positive emotions positive ones, and it is hard to imagine an emotion neutral towards its object. It has even been argued that different emotions may differ only in the evaluations involved.\(^3\)

The third characteristic is that emotions are typically associated with affects, which are normally feelings or sensations and sometimes physiological changes. One often feels anger, guilt and shame very strongly, hope and pride often less so. Loving or hating a person or fearing an animal affects one in a way that cold (although perhaps evaluative) beliefs do not. In everyday speech people tend to talk about their “feelings and emotions” and take it mean the same thing, which shows that in folk psychology the tie between emotions and feelings is strong. It has been claimed that all beliefs affect people to some extent, but emotions seem prone to do so much more intensely. This helps to distinguish emotions from ordinary, non-emotional beliefs.

\(^2\) Spinoza, E. II. axiom 3.
\(^3\) See Lyons, 1980, ch. 3.
Finally, emotions often motivate or incite one to particular types of action. An angry man seeks revenge and may have a hard time holding himself back, and a frightened man flees, hides or tries to protect himself. A guilty man makes amends, or feels that he should, and a shamed man avoids the gaze of other people. Other emotions may seem less inciting. Hope may for example not incite one to do anything, except wait and see what happens.

These observations shall be my starting point. I have now given a rough idea of what group of conditions I refer to by “emotion”. Emotions are conditions that normally involve feelings or physiological affects, but also have evaluated external objects and often incite one to action. My task now is to move from these initial observations to an understanding of what emotions are and how they function, and in doing so it is essential not to lose sight of the starting point. A theory about the nature of emotions should account for, or at least allow that emotions are normally intentional, evaluative, affective and motivational, or else explain how common sense can be so drastically mistaken about them.

Some writers on emotions have been tempted to ignore things that in common language are termed emotions. They have for example claimed that some alleged emotions, such as love, are in fact durable dispositional traits whereas proper emotions are occurrent and have a limited duration. Their theories, some of which purport to be complete theories of all there is to emotions, sever a large portion of what commonly are considered to be emotions, from the rest, i.e. from what they consider to be central cases of emotions. Which emotions one considers central, and which of their features most significant will depend to a large degree on one’s background and interests and the context in which one is thinking about emotions. When Aristotle discussed emotions he was concerned with how they could be manipulated
by rhetoric, hence his emphasis on the role of beliefs in emotions; William James' viewpoint was that of a physician-turned-psychologist, hence the emphasis on the physiology of emotions; and Martha Craven Nussbaum views them with the mind of someone who grew up viewing philosophical and broadly moral questions through literature, hence her emphasis on a narrative structure of emotions.

Each of these theories may provide important insights into emotions and be good theories about particular emotion types or aspects of emotions. But they ought not to purport to be complete. Emotions are indeed heterogeneous and often seem to have nothing interesting or significant in common, nothing which unifies them under one banner. But although common language may often be confused and fit badly the class of actual phenomena it is meant to be about, I shall remain as faithful to it as I deem reasonable and try to understand emotions with as little Procrusteanism as possible. Lyons claims that to study emotions via emotion-terms is liable to be a very confusing business.\(^4\) But ignoring the lines drawn in common language may blind us to important relations and interconnections. It might prove to be a more fruitful strategy to make distinctions among emotions and perhaps place them on various scales, such as between cognitive and non-cognitive, cultural and natural, passive and active and so on, while also trying to understand why the different categories share the name "emotion" in common language.

\(^4\) Lyons, 1980, p. 56.
CHAPTER II.

EMOTIONS AS AFFECTS.

since feeling is first
who pays any attention
to the syntax of things
will never wholly kiss you

(E.E. Cummings)
II. 1. Introduction.

The philosophical study of emotions is now dominated by cognitivism. Although the roots of cognitivism can be traced back to the times of Aristotle and the Stoics, its rise to domination is fairly recent. The wave of cognitivism arose only a few decades ago and then as a response to the felt inadequacies of affect theories which had dominated the scene in the first decades of this century. A look at affect theories and their problems will therefore enhance our understanding of contemporary cognitivism, and give a vantage point from which to assess the cognitive theories.

Affect theories have been held by a number of philosophers and psychologists and come in a variety of different versions. For most of them the affect in question is an experienced feeling or a sensation and their theories share the view that essential to each type of emotion is a particular subjectively felt sensation. According to such theories I am angry when I feel the special sensation of anger, which I know by experience to be different from the sensations of other emotions, such as the special sensation of envy, hope, shame or guilt. Theories of this kind agree with a popular view of the emotions, one that is reflected in the fact that people often talk of their "feelings and emotions" as if the two were the same. Emotions are, to most people, something that they feel inside, sometimes turbulently, sometimes softly, but always privately and subjectively.
Affect theorists may differ on whether the affect is always a sufficient condition for emotions, but for all of them an affect is a necessary condition; without the affect there is no emotion. The most important distinctions among affect theories is between those that claim that an emotion is a psychic feeling and those that hold it to be a bodily feeling.\(^5\) The difference between them can be stated briefly as follows: One holds that emotions are feelings caused by thoughts, beliefs or knowledge of facts, whereas the other holds that emotions are feelings of bodily changes which are caused by direct perceptions of facts.

II. 2. Two Types of Affect Theories.

The former of the two types was held by John Locke and later by David Hume. In *An Essay Concerning Human Understanding* Locke observes that the simple ideas of pain and pleasure, can, although they are strictly speaking always in the mind, be either of the body or of the mind. The pains and pleasures of the mind are produced by thoughts of the pains and pleasures we get from other things. Locke wrote that, “Passions are moved by things, only as they appear to be the causes of Pleasure and Pain.” The love of grapes, which is by itself a pleasurable passion, is produced by the thought of the pleasure we get from actually eating grapes. Locke then goes on to describe various types of passions. Sorrow is, “the uneasiness of Mind, upon the thought of a Good lost, which might have been enjoyed

---

\(^5\) The distinction is made in these terms by Justin Oakley, (Oakley, 1992, ch. 1) and, in different terms, by Robert Solomon and Cheshire Calhoun. Solomon and Calhoun talk of sensation theories and physiological theories. (Calhoun and Solomon, 1984, p. 8)
longer; or the sense of a present evil,” whereas fear is the “uneasiness of mind upon the thought of future evil likely to befall us.”

Hume later adopted the core of Locke’s theory. He wrote:

As all the perceptions of the mind may be divided into impressions and ideas, so the impressions admit of another division into original and secondary. This division of the impressions is the same with that which I formerly made use of when I distinguish’d them into impressions of sensation and reflection. Original impressions or impressions of sensations are such as without any antecedent perception arise in the soul, from the constitution of the body, from the animal spirit or from the application of objects to the external organs. Secondary, or reflective impressions are such as proceed from some of these original ones, either immediately or by the interposition of its idea. Of the first kind are all the impressions of the senses, and all bodily pains and pleasures: Of the second are the passions, and other emotions resembling them.

Emotions are thus sensations which arise from either other sensations or from ideas. Hume distinguishes between direct and indirect emotions. The direct ones arise “immediately from good or evil, from pain and pleasure,” whereas the indirect result from the conjunction of pain or pleasure with other qualities. Among the direct emotions Hume counts desire, aversion, grief, hope, fear, despair and security, but among the indirect one’s are pride, humility, ambition, love, hatred, vanity, envy, pity, malice and generosity. But in spite of the complexity of Hume’s theory of the emotions, he agrees with Locke that the constitutive essence of each type of emotion is a particular simple and unanalyzable

---

6 Locke, 1975, pp. 230-31
sensation, known by introspection and as such indescribable by words or in theory. As Locke put it, the passions, "like other simple Ideas cannot be described, nor their Names defined; the way of knowing them is, as of the simple Ideas of the Senses, only by Experience." A theory of the various types of emotions, is therefore, for Hume and for Locke, never more than a theory of the causes, effects and conditions associated with emotions, while the emotions themselves have no theoretical or cognitive depth.

The latter of the two types of feeling theories is often named the James-Lange theory after its proponents William James and Carl Lange. In "What is an Emotion?", James acknowledges that there are feelings of pleasure or pain, interest and excitement bound up with mental operations, without physiological effects. We may take pleasure or pain in certain sounds and colors without it being accompanied by any detectable physiological changes or causing movement in the body or the face, and similarly ideas may charm us or tire us without a bodily manifestation. He takes this to prove that there are pleasures and pains inherent in certain forms of neural processes, wherever they occur. But, without giving a reason, James chooses to leave these aside and confine his attention to, "the more complicated cases in which a wave of bodily disturbance of some kind accompanies the perception of the interesting sights or sounds or the passage of the exciting train of ideas." James wrote:

Our natural way of thinking about these standard emotions [i.e. surprise, curiosity, rapture, fear, anger, lust, greed, and the like] is that the mental perception of some fact excites the mental affection called the emotion . . . My thesis on the

---

9 Locke, 1975, p. 229.  
10 James, 1984, p. 128.
contrary is that the bodily changes follow directly the PERCEPTION of the exciting fact, and that our feeling of the same changes as they occur IS the emotion.¹¹

James here explicitly attacks the common sense idea that when we resent, fear, are ashamed etc., we do so because of something we know or believe, i.e. he rejects the psychic feeling theories. A summary of James’ theory might go as follows: Objects affect one or more of the sense organs, and impulses pass to the cortex, where the object is perceived. The perception triggers a neural mechanism which causes currents to run down to the muscles and viscera and changes them in a variety of ways. Muscles get tense, the heart throbs, blood flushes to the cheeks, breathing gets difficult. We feel these changes and that feeling is the emotion. James then claims that “...without the bodily states following on the perception, the latter would be purely cognitive in form, pale, colorless, destitute of emotional warmth”.¹²

James makes three fairly simple points in support of his theory. The first is his claim that the nervous system is nothing but a bundle of predispositions to react in predetermined ways to environmental stimuli. The neural system is a mere hyphen between perceptions and responses. James likens us to locks which presuppose certain forms of keys. The keys are not attached to the locks, but we are sure to come across them in the course of our lives, and when we do the locks open. Similarly, triggers in our neural system presuppose the existence of certain kinds of stimuli. When we come across the stimuli they trigger the appropriate response.¹³

¹¹ James, 1984, p. 128.
¹² James, 1984, p. 128.
¹³ James, 1984, p. 129.
The susceptibility to have emotions is among our natural neurological predispositions. We are hardwired to react angrily to certain stimuli, and to become afraid at others. Without ever having seen an elephant a child is predisposed to be afraid if one were to come charging toward it, and a man in the wilderness, says James, is bound to feel excited at the sight of the shape of another human in the distance.\textsuperscript{14}

The second step in James' argument is his claim that every bodily change is felt, either acutely or obscurely, the moment it occurs. “Our whole cupic capacity is sensibly alive; and each morsel of it contributes its pulsations of feeling, dim or sharp, pleasant, painful, or dubious, to that sense of personality that every one of us unfailingly carries with him.”\textsuperscript{15}

This is followed by the crucial point that, “if we fancy some strong emotion, and then try to abstract from our consciousness of it all the feelings of its characteristic bodily symptoms, we find that we have nothing left behind, no “mind-stuff” out of which the emotion can be constituted, and that a cold and neutral state of intellectual perception is all that remains.”\textsuperscript{16}

From this James apparently wishes to draw his conclusion, that emotions are nothing but the subjectively felt sensations, caused by physiological changes.\textsuperscript{17}

Given the premises it is clear that any feelings involved in emotions are feelings of bodily changes. But there are several things to notice about James' argument. The first is that the method of introspection is highly suspect and has been debunked as a reliable source

\textsuperscript{14} James, 1984, p. 130.
\textsuperscript{15} James, 1984, p. 131.
\textsuperscript{16} James, 1984, p. 131.
\textsuperscript{17} It is interesting to note that cognitivists have turned this argument on its head and argue that if we abstract from an emotion all the beliefs or cognitions involved, what we have left is mere feeling. (See Calhoun, 1984, p. 333)
of knowledge of the mind. The thought experiment, which is the core of James' argument, is therefore not likely yield the knowledge he needs for his theory. I will come back to problems concerning introspection later.

The second point is that the argument supports James' theory only if we presume that an emotion is a feeling. That emotions are feelings is a hidden and a highly contentious premise in the argument. The argument is thus a defense only against other feeling theories of emotions, but does in no way refute the possibility that the remaining "cold and neutral state of intellectual perception" is sufficient to constitute an emotion.

The last point I want to make is that James' conclusion is too strong. All that could possibly follow from the premises of the argument, whether hidden or stated, is that feelings of bodily changes are necessary for emotions, but they are not shown to be sufficient. Therefore, even if we accept James' claims, based on introspection, the argument fails to support his central claim, that an emotion is nothing but feeling of bodily changes. It does not show, for example, beliefs to be superfluous or unnecessary for the occurrence of emotions.

The Humean theory and the James-Lange theory may on the surface look very different, but on a closer look, they are essentially the same; they both identify emotions with subjectively felt affects. That the psychic-feeling theory and the bodily-feeling theory are essentially the same can perhaps best be shown by looking briefly at Descartes' theory of emotions. According to Descartes an emotion is the subjective awareness of the activities of the animal spirits in the body. But he acknowledges two ways in which an emotion may
occur, one of which is closely akin to Locke's and Hume's psychic-feeling theory, and the
other equally close to the James-Lange theory.

Descartes distinguishes between the functions of the body and of the soul. The functions
of the body are movement and heat, whereas the chief function of the soul is thought.
Thought is again of two sorts, on one hand there are actions and desires, and on the other are
passions. The passions are "all those kinds of perceptions or forms of knowledge which are
found in us."\(^{18}\) Emotions are among these kinds of perceptions. They start with the
perception of some object, in the case of fear it may be of a strange and frightful animal
approaching. This perception is then transmitted via the pineal gland to the soul. In the soul
the perception is compared with the memory of previous similar perceptions, and if they are
found to have been harmful to us it "excites the passion of apprehension in the soul and then
that of courage, or else that of fear and consternation according to the particular
temperament of the body or the strength of the soul."\(^{19}\) This is a psychic-feeling theory; the
perception of an animal, previously hurtful, gives rise to the belief that the animal is
dangerous, which in turn gives rise to the feelings of either fear or courage.

But Descartes has another account of emotions, which does not involve beliefs or
comparison to previous perceptions. Perceptions can dispose the brain,

\begin{quote}
in such a way that the spirits reflected from the image thus formed on the gland,
proceed thence to take their places partly in the nerves which serve to turn the back
and dispose the legs for flight and partly in those which so increase or diminish the
\end{quote}

\(^{18}\) Descartes, as quoted in Lyons, p. 3.
\(^{19}\) Descartes, as quoted in Lyons, p. 2.
orifices of the heart ... [that it] sends to the brain the spirits which are adapted for the
maintenance and strengthening of the passion of fear.\footnote{Descartes, as quoted in Lyons, p. 3.}

Fear becomes, on this account, the awareness of the unusual activities in the body, or, to
use James’ words, feelings of bodily changes. This is the James-Lange theory in a nutshell.
Descartes thus represents both psychic feeling theories and bodily feeling theories, but they
are only theories of different causes of emotions. The emotions themselves are, in any case,
simple perceptions in the soul. The perception of the object, the belief that it is dangerous
and the bodily reaction are not a part of the emotion of fear.

This serves to emphasize that, at the core, the feeling theories are the same, and they are
therefore perhaps vulnerable to the same objections. There are several standard criticisms to
theories which identify emotions with sensations, some of which have been taken to support
cognitive theories of emotions. I will look at some of these criticisms and try to assess to
what extent they are successful at severing the tie between feelings and emotions. Later I
will discuss to what extent they can be used as arguments in favor of theories which make
thought and beliefs essential to emotions.

\section*{II. 3. Emotions are Independent of Feelings.}

A common criticism is based on the observation that emotions often seem to be
independent of feelings. Robert C. Solomon wrote that, “however predictable the
association of feeling and emotion, the feelings no more constitute or define the emotion

\footnote{Descartes, as quoted in Lyons, p. 3.}
than an army of fleas constitutes a homeless dog.\footnote{21} And just like a dog may get rid of its fleas it seems that feelings may fade away while the emotion remains. Long lasting emotions are often mentioned to support claims of this kind. To say that Jones has loved his wife since the day they met or that he is still angry with Smith after all these years, is not to say that he has, over the years, constantly felt the warmth of love or the agitation of anger. He remains angry without feeling agitated and he loves his wife without the feeling of warmth. Therefore these emotions can not be identified with the feelings often associated with them.

The typical reaction to this sort of objections involves a bit of Procrustean butchery, something I intend to avoid as far as possible. Feeling theorists make a distinction between occurrent emotions and dispositional emotions (or a disposition to have an emotion, which by itself may not amount to an emotion). They then claim that their theory is only meant to be about occurrent emotions, or even that only occurrent emotions are proper emotions, and that they happen to be essentially feelings.

The distinction between occurrent emotions and dispositional emotions is often made as follows: We can call a person angry and mean it either occurrently or dispositionally. The person’s anger is occurrent anger if he actually fulminates, raises his voice or shows other angry or threatening expressions, or has (possibly suppressed) angry thoughts and feelings. But we mean it dispositionally if we mean only that the person is prone to have fits of occurrent anger on particular occasions and in certain circumstances, but are not referring to a particular episode or instance of anger. To this it is often added that a dispositional

\footnote{21} Solomon, 1978, p. 159.
emotion would better be thought of as a character trait than as an emotion and therefore needs not to be dealt with in emotion theory.\textsuperscript{22}

This distinction may make it look acceptable to sever dispositional emotions from the study of emotions. But there is a complication here. We have to make a clear distinction between two types of emotional dispositions; general dispositions and focused or specific dispositions. A general disposition to anger, i.e. a disposition to become angry frequently and for trivial reasons, is probably best thought of as a character trait. One can have such a character trait and yet, at some given moments, not be angry at anyone or anything. But a focused or specific dispositional anger, such as John's yearlong anger at Smith, has, at any given moment all these years, been real anger directed at a specific object. It had a beginning and it may come to an end, and all the time it was real, actual anger. This can less plausibly be written off as a character trait; one can be subject to specific dispositional anger without having an angry character, i.e. without the general disposition to anger. If we leave specific dispositional emotion out of emotion theory we will likely be closing our eyes to important aspects of emotions, namely their endurance and ability to recede to the back of our minds, while still playing an explanatory role in the way we perceive, think of and interact with the world. To be affected by an emotion is not necessarily to have an emotional episode. Some emotions seem actually to be characteristically enduring rather than merely episodic and they may be manifested solely in the way they inconspicuously affect our thoughts and actions. John may unconsciously shun and ignore Smith, and only when that is pointed out to him, realize that deep down he is still angry. The distinction, as

it is frequently made, does thus not cover all the ground. There are indeed character traits which involve dispositions to have emotional episodes, but are themselves not emotions. What is termed occurrent emotions, I would prefer to call emotional episodes. This leaves out the long term emotions which may never find expression in emotional episodes, and would in some cases be incorrectly thought to involve dispositions to anything like what Lyons and others describe as occurrent emotions.

If feeling theorists were to give us good reasons to exclude this class of conditions from a general theory of emotions, they would still not be entirely off the hook. Short term emotions or emotional episodes may also be independent of feelings. To use an example from Justin Oakley, when we see a snake on our path we may undergo physical changes, caused by our perception of danger, a throbbing heart, high pulse rate, shallow breathing and we flee. This is clearly a story of an episode of fear, and it can all take place without us feeling anything at the time. We may not feel a thing until we are in a safe distance from the snake. Leaving this out of a purportedly general theory of emotions requires a special justification.

Feelings seem therefore not to be essential to emotions. But are other affects necessary? Justin Oakley believes they are (along with cognitions and desires). He recognizes that we can have emotions without experiencing any feeling at all. But he argues that, “the point here suggests rather that emotional affectivity need not be understood solely in terms of feelings.” Oakley reserves the word “feeling” for affects that one is conscious of or notices, whereas some affects involved in emotions can go unnoticed. He claims that an emotion

---

requires only physiological changes, or even psychic modifications, by which he means not merely a disposition to have emotions, but something that actively guides our attention and colors our perceptions of the world. An affect, he says, is,

a bodily or psychic condition which we are in, but which we need not feel, in having an emotion, and this condition is linked dynamically with the elements of cognition and desire in that emotion.

But it seems to me that Oakley has introduced a considerably thinner notion of affectivity than James or Hume would subscribe to, and one perfectly compatible with some cognitive theories, such as de Sousa's account of emotions as patterns of salience among objects of attention. His notion of a psychic modification is partly an attempt to avoid the charge against cognitive theories, that a mere cognition is cold and neutral and does not guide our attention, "color" our perception of the world or move us to action the way emotions do. But his answer to the question, why do emotions guide our attention and color our perception is vacuous. He merely says that they are not emotions, unless they do. Opium puts us to sleep because of its "virtus dormativa".

II. 4. Feelings and the Intentionality of Emotions.

The second criticism against feeling theories is the claim that mere feelings are not intentional states and are unable to be about anything, as emotions normally are. Feeling theories can not adequately account for the fact that, at least some emotions are directed

---

towards an object, and they therefore offer an inadequate account of the emotions. Feeling theorists typically account for the apparent object directedness of emotion in terms of a focus of attention or directedness of behavior. Hume said that in pride and humility our view is fixed on the self and, “when self enters not into consideration, there is no room either for pride or humility.” But for Hume and other feeling theorists the tie between the emotion and its object is a matter of empirical contingency. A person with the particular sensation of pride could, in principle, focus on someone other than himself and a person feeling the particular sensation of shame could in principle focus on what she regards her admirable qualities or deeds.

The first thing to notice about this is that strictly speaking it is not the emotion (sensation) that is focused on an object, the focusing is something that typically follows the emotion. This is not a serious objection to feeling theories, it shows only a need for a minor modification of the way we speak about emotions and their intentionality.

This does not let the feeling theorists off the hook though. A more forceful argument from intentionality can be built on the contention that the tie between emotions and particular types of objects is conceptual or logical, but not contingent and empirical. It is not a lucky coincidence that we are typically angry at what we perceive to be an insult or a harm inflicted on us. It is given in the concept of anger and its logical structure, and that is something that feeling theories can not account for. It is incompatible with identifying emotions with simple sensations, because simple sensations do not have the required cognitive depth.

---

This argument needs a further twist to hold, because there is actually a way to think of mere feelings as being intentional. The brutest of all feelings, physical pain, doubtlessly has the evolutionary function to represent physical injury. If one did not feel pain one might not avoid the fire and possibly burn to death. The ability to feel pain has been chosen for in evolution because pain represents physical harm and is conducive to one’s survival. Pain can in this way be thought of as being about something external to itself. But note that even if we can justify thinking of pain as having an intentional object we would certainly be taking it too far if we take the object to be anything but the body undergoing some physical harm. If I am blindfolded and my hand is pressed against a hot stove, the pain I feel will not represent, or be about a hot stove, it will only be about the palm of my hand burning.

If we apply this notion of representation to emotions as feelings we will see that this is not sufficient to give feelings the kind of intentionality emotions purportedly have. On a James-Lange account of emotions as bodily feelings, emotions would be about bodily changes, whereas on a Lockean or a Humean account of emotions as psychic feelings, they would be about one’s beliefs. That is certainly not the kind of intentionality an emotion theory should explain. We do not fear our pounding hearts nor our belief that there is a tiger jumping at us. What we fear the is the tiger itself, and that is what an adequate emotion theory should account for. In the absence of a proper account of how feelings can be intentional in the way emotions purportedly are, feeling theorists will have to either reject the alleged conceptual connection between an emotion and its objects, or content that emotions do not really have objects.
II. 5. Feelings and the Rationality of Emotions.

The third criticism is based on the claim that mere feelings cannot be assessed as being either rational or irrational, appropriate or inappropriate, proportionate or disproportionate, as emotions supposedly can. In arguing against feeling theories Robert Solomon noted, "by way of a clue rather than a full-blown argument," that emotions can often be assessed as being rational or irrational, appropriate or inappropriate, justifiable or unjustifiable.

Yet no such evaluations are appropriate in the realm of headaches and bellyaches, warm flushes and nausea. Headaches are neither reasonable nor unreasonable. . ..Yet we say without hesitation that "you were wrong to be angry at him,” that “he was unjustified in being jealous,” that “loving her was the wrong thing to do.” . .. [T]hese sorts of evaluations should give us a further reason to reject the seemingly indisputable thesis that emotions are mere feelings, occurrences that happen to us beyond our own control. Feelings and occurrences are not "reasonable” or “unreasonable.” They simply are. Only what we do can be so assessed.28

An argument along these lines was one of the reasons feeling theories were largely rejected and cognitivism embraced. But this passage from Solomon does in fact leave us without a clue as to what the argument really is. He seems to have in mind only a distinction between passivity (things that happens to us) and activity (things we do). Our actions can be assessed as being either rational or irrational, whereas what happens to us and is not in our control cannot be so assessed. This leaves me at loss about assessment of ordinary beliefs. If I see a glass on the table in front of me, it is far beyond my control whether or not I believe there is a glass on the table. I believe it whether I choose to or not. Does that mean

that beliefs of this kind cannot be assessed as either rational or irrational? To make anything of Solomon’s clue it needs to be shown that mere feelings can never be in our control, and for that purpose the analogy to headaches and bellyaches is not sufficient. Headaches and bellyaches tend to have physical causes beyond our control, but that does not exclude the possibility of mere feelings being induced by something more voluntary, such as thoughts, or by internalized norms for the rationality of emotions, and to the extend that thoughts are in our control, feelings could be said to be in our control. The clue, as presented by Solomon, cannot have swayed anyone either from feeling centered theories nor towards cognitivism. We need a more convincing account of how the clue can be developed into a refutation of feeling theories.

But first, what kind of rationality is it that we attribute to emotions? We sometimes assess the instrumental or pragmatic value of emotions. “There is no use crying over spilled milk” we say, meaning that grieving a loss is not going to bring back the lost thing. Grieving serves no material purpose and does not bring about any good. But in another sense it does make sense to cry over spilled milk. Grief is warranted by a loss and is rational as a response to a loss, but not as a response to winning the jackpot; anger is warranted by an offense but not by display of goodwill; gratitude is warranted by generosity but not by stinginess. This, but not instrumental rationality, is the sense of rationality we are concerned with. This sense of rationality is closely related to epistemic warrant. Emotions are warranted by circumstances, events or actions, regardless of their instrumental value. It can be rational, in this sense of rationality, to get angry (because of an offense) although getting angry would be harmful and thus instrumentally irrational.
Errol Bedford gives the argument from the rationality of emotions a different slant from that of Solomon's:

I am inclined to think that if an emotion were a feeling no sense could be made of them [i.e. the distinctions between appropriate and inappropriate or justified and unjustified emotions] at all. It may be said that an emotion is unjustified when a feeling is inappropriate or unfitting to a situation. But I find this unintelligible. Feelings do not have the character that makes this relationship possible . . . . In general, I do not think it can be maintained that logical predicates apply either to feelings or to sensations. What reasons could be given for or against a feeling, or for or against its "inappropriateness" to a situation? If someone were to say "I felt a pang this afternoon" it would be meaningless to ask whether it was a reasonable or unreasonable pang. The matter is different if he says "pang of regret," but the phrase "of regret" does not name the feeling, . . . and the pang is justified, if it is, not as a feeling, but because his regret is justified . . . .

Whereas Solomon based his clue on the alleged passivity of feelings (they are not something we do, they simply are), Bedford picks on the simplicity of feelings. They do not have the character (presumably the cognitive depth) that allows us to apply logical predicates to them.

To rescue his theory from arguments like this a feeling theorist might suggest that in order to assess emotions as being either appropriate or inappropriate, it is sufficient that they are responsive to norms for appropriateness. He would concede that the assessment is, strictly speaking, not an assessment of the emotion/feeling itself, but rather of its antecedent conditions. This, like the similar response to the argument from intentionality, would

---

29 Bedford, 1984, p. 274.
require a minor modification of the way we speak of the emotions, but appears to rescue feeling centered theories from the argument from rationality. If Bedford’s argument is to count as an argument against feeling theories it must therefore be supplemented by an argument showing that feelings are not responsive to norms, which might prove hard to do. He would have to show for instance why it is that people from different cultures are nauseated by different things

In response to this a cognitivist might make a distinction between, on the one hand, a claim that an emotion is appropriate to a situation and, on the other hand, a claim that it makes sense for the subject to have that emotion. For an emotion to be rational in the second sense it must stand in the right kind of epistemic relation to the situation that warrants it. We can imagine a situation where it makes sense to be afraid, such as when a tiger is sneaking up behind one’s back, but where the fear one actually feels is utterly irrational. The fear may have nothing to do with the tiger, but be caused by the suspicion that the sky is about to collapse on one’s head. For the fear to be rational it seems that it also needs to be epistemically related to the danger. A belief or some equivalent of a belief about the tiger seems therefore to be necessary for an emotion to be assessed as rational.

Most cognitivists, including Bedford, would not resort to this response though. This does not make a belief or an equivalent of a belief, a part of the emotion and it allows the norms for the appropriateness of emotions to be contingent or conventional, whereas what cognitivists would presumably claim is that the norms for the rationality and appropriateness of emotions are given in each emotion concept, or at least that the range within which the norms for an emotion can vary is limited by the concept.
The argument here is closely related to the previous one, about the objects of emotions. If emotions were mere feelings they could in principle have any object whatsoever (insofar as they have objects at all). But they cannot have just any object, the kinds of objects they can have is given in the concept of each emotion. Similarly, if emotions were mere feelings, they could in principle be considered rational or irrational by any kind of norm whatsoever. The norms could be some kind of social or conventional moral norms (as opposed to logical or conceptual norms). So the argument here is that the norms can not be conventional or contingent. It is inconceivable that it can be rational to feel angry at one's benefactor or to feel guilty for one's acts of kindness. The kinds of situations that warrant each type of emotion is given in each emotion concept, and identifying emotions with mere feelings does not allow the theoretical or conceptual depth needed to make sense of that.

This move by cognitivists certainly needs a further justification, or else it merely begs the question. Feeling theorists would simply reject the contention that the concept of an emotion does entail norms for the rationality of that emotion. Ronald de Sousa argues (along the lines of Anthony Kenny) that the vocabulary of emotions is learned from association with paradigm scenarios, first in daily life, later through stories and fairy tales and finally through literature and art. Paradigm scenarios involve two aspects: a paradigm situation providing the characteristic object of the emotion; and a set of characteristic or normal responses to the situation.30

De Sousa then argues that the norms for the rationality of the emotions are, at least partly, derived from the paradigm situations. “Since emotions are learned in terms of these

paradigms, they cannot, at least within a given social context, be criticized for inappropriateness if they occur in response to the paradigm situation." The upshot of the picture drawn by de Sousa is that the concept of an emotion entails norms for its rationality. The norms for the rationality of emotions are therefore not entirely contingent as they would be if emotion terms referred to nothing but a simple, experienced sensation.

II. 6. Feelings and the Individuation of Emotions.

The fourth criticism, and what I take to be decisive against feeling theories, is that introspection is incapable of individuating feelings in any way parallel to our taxonomy of emotions. Solomon asks himself,

What is the difference between a common pair of emotions—for example, embarrassment and shame? People are rarely, if ever, confused about which emotion they “have,” but, when asked to differentiate between them, they find themselves speechless. [...] In fact, the feelings and the sensations associated with the one emotion may be and usually are no different from those associated with the other.32

William James might not be disturbed by this objection. He believed that emotions, insofar as they are separate and distinguishable types, could be distinguished by their physiological differences, mostly visceral changes. He wrote that,

the various permutations and combinations of which these organic activities are susceptible, make it abstractly possible that no shade of emotion, however slight,

---

should be without a bodily reverberation as unique, when taken in its totality, as is the mental mood itself.\textsuperscript{33}

The most straightforward response to this claim is that, even when we do not pay attention to the physiology of jealousy and envy or shame and embarrassment or any other emotion, we generally know when we are jealous but not envious, ashamed but not embarrassed, indignant but not annoyed. So even if James were right about the physiological uniqueness of separate emotions we seem in reality to distinguish emotions by some criteria other than differences in feelings or physiology. There is therefore a distinction among the emotions, of a different sort.

Although James’ theory was a scientific hypothesis, subject to experimental verification he argued only by way of a thought experiment. “Imagine away the feelings of physiological changes, and all you have left is a cold and emotionally neutral cognition,” his argument went. But later experimental evidence has not turned out to be in James’ favor. The physiologist, Walter B. Cannon, tested James’ hypothesis and found that there is no correspondence between types of physiological events and types of emotions. He found, for instance, that the same visceral changes can occur in very different emotional states as well as in non-emotional states and that artificial induction of physiological changes, typical of intense emotions, does not necessarily produce any emotions. Athletes injected with adrenaline felt “on the edge”, or “keyed up” as just before a race while other subjects felt “as if afraid”, “as if awaiting great joy” or “as if moved”.\textsuperscript{34} In a later psychological study

\textsuperscript{33} James, 1984, pp. 130-131.
\textsuperscript{34} Cannon, 1984, pp. 143-151.
physiological reaction ordinarily involved in occurrences of fear, anger and euphoria were mimicked with drug injections. The subjects were then given different psychological cues. Most of those cued to react angrily claimed to experience anger, those cued for fear claimed to experience fear and those cued for euphoria were euphoric. These differences could not be accounted for in terms of physiological differences.

We seem to have to fall back on introspection and verbal reports of the sensations themselves, and that is highly problematic. Feelings are often not sufficiently distinct to be individuated into types of emotions. We may, for example, not be able to tell envy from jealousy, regret from guilt etc., because they often feel (almost) the same. Feeling theorists might resort to arguing that this does not matter, because envy and jealousy, if they feel the same, are indeed the same emotion being thought of as two different emotions by reference to contingencies external to the emotion itself. “Might envy and jealousy not be the same emotion being caused by different events?” they could ask.

But this response faces the further problem that separate instances of the same type of emotion may feel completely different. We sometimes feel fear as paralyzing, sometimes as agitating, and fear may even at times be manifested only in feelings associated with an effort made to block typical symptoms of fear. If the feeling theorists take the apparently logical step here and claim that in such cases we do not have the same emotion they do so at the cost of a huge loss in plausibility. Their theory would simply be too greatly at odds with common language taxonomy of emotions and would probably have to be taken as a theory.

36 See also Oakley, 1992, p. 19.
37 See Greenspan, 1988, pp. 21-22.
of something else than emotions. (It would be like to theorize about oranges and say “Look at these apples. Let’s call them oranges” and then go on to theorize about apples by their new name.) A theory which does not allow us to distinguish envy from jealousy, shame from guilt and regret, grief from pity, and annoyance from indignation, is simply not a theory of emotions. The conclusion to be drawn from this is that there is no such thing as the feeling of fear or the feeling of anger or of any other emotion. But we must note that this does neither show that there is no such thing as emotions, nor that emotions have nothing to do with feelings.
CHAPTER III.

COGNITIVE THEORIES OF EMOTIONS.
III. 1. Introduction.

Affect theories have largely been abandoned for the reasons I have just described. Affects lack the depth and complexity that allows them to account for some important aspects of emotions. They gave way to what may be called intellectualist cognitivism, i.e. theories that make propositional beliefs and thoughts or something equivalent to propositional beliefs or thoughts, central to emotions. Intellectualist cognitivism promises to solve or avoid the problems by identifying emotions with propositional thoughts or beliefs rather than with feelings or affects. According to intellectualist cognitivism the concept of each type of emotions is characterized by a particular type of beliefs, about a particular type of objects. The study of emotions thus becomes a matter of conceptual analysis and emotions come to be individuated by the different beliefs and thoughts they involve. The problems with introspective or physiological individuation therefore vanish. Intellectualist cognitivism purports to explain the object directedness of emotions in terms of the objects of the beliefs involved and the rationality of emotions comes to be tied to the rationality of the beliefs. And finally, the apparent independence of emotions from feelings becomes a truism, a change in belief is a change in emotion, regardless of whether the feelings change accordingly.

I have talked here of contemporary cognitivism as intellectualist to distinguish it from a different brand of cognitivism. In the first half of this century some English psychologists
and philosophers held cognitive theories which did not tie the emotions to the intellect or require them to have propositional content. The concept of thought they associated with emotions was broad enough to apply to all states and actions of the mind with objective content, including sensation, imagination, and memory as well as the intellectual operations of understanding and judgment. C. D. Broad defines cognition as any experience with an “epistemological object”, which distinguishes it from mere feeling. Emotions, accordingly, are cognitions which have a felt quality or tone. But what enabled Broad this feat was the fact that he accepted introspection as a source of knowledge about the mind. Emotions as non-intellectual, non-propositional but yet cognitive states were to be individuated by introspection. Broad’s theory, and other theories of the kind, thus share a fatal flaw with theories which equate emotions with feelings, and they do not avoid the problems with affect theories as intellectualist cognitivism purports to do by shifting the study of emotions from introspection to conceptual analysis. In what follows I am concerned with cognitivism of the intellectualist brand and by “cognitivism” I shall refer only to intellectualist cognitivism. On the one occasion I wish to bring up the non-intellectualist brand I shall, following John Deigh, talk of “traditional cognitivism.”

Cognitive theories differ in various ways. They differ for instance on the question whether a cognition is only a necessary part of an emotion or whether it is also sufficient to constitute an emotion. They also differ on whether emotions must involve actual beliefs or whether it is enough, for example, that a thought is entertained without necessarily being

---

endorsed. But these differences are largely irrelevant to my discussion. I am concerned mostly with the question whether cognition is necessary to emotions.

Among the most notable contemporary proponents of cognitivism are Robert C. Solomon and Martha Craven Nussbaum. Nussbaum professes her allegiance to the Stoic belief that, "the emotion is itself identical with the full acceptance of, or recognition of, a belief." Solomon identifies emotions with evaluative judgments.

An emotion is a judgement (or a set of judgments), something we do. An emotion is a (set of) judgment(s) which constitute our world, our surreality, and its "intentional objects."

In what follows I will assess whether the failures of affect theories necessitate cognitivism. My conclusion is that they do not. I will then discuss a problem cognitivists have to deal with, namely the apparent possibility of having emotions which go against ones explicitly held beliefs. I argue that suggested solutions to that problem either beg the question in favor of cognitivism or slide into something very much like traditional cognitivism, which in turn fails to solve some of the problems of affect theories.

III. 2. Beliefs and the Intentionality of Emotions.

Emotions typically have objects. That was one of the reasons feeling theories were rejected and replaced with cognitive theories. Anthony Kenny goes as far as to claim that it is essential to emotions that they are directed at an object. He wrote:

---

The most important difference between a sensation and an emotion is that emotions, unlike sensations, are essentially directed to objects. It is possible to be hungry without being hungry for anything in particular, as it is not possible to be ashamed without being ashamed of anything in particular.\textsuperscript{41}

Kenny uses this as a starting point in arguing for the necessity of beliefs for emotions. But is the transition from object-directedness to belief justified?

The first thing to notice is that some emotions appear to be objectless. We sometimes seem to be afraid without being afraid of anything specific or without knowing what we are afraid of, and we can be angry for no reason and without directing our anger at anything or anybody. Kenny concedes that there may be apparently objectless states which we are tempted to call anger or fear, but he claims that these states do not show the tie between emotions and objects to be contingent. For one thing, these emotions often have objects in spite of the apparent objectlessness. An apparently objectless depression, which makes all things seem black, is precisely about all the things that seem black, and phobia is not fear without an object but fear without an adequate object. But there are still cases where an (alleged) emotion is truly objectless. About such cases, Kenny argues that, insofar as we are right to call these conditions “emotions”, it is only by reference to emotions which do have objects. An agitated person may claim to be afraid, and later invent or pick an object for his fear. But if he repeatedly claims to be afraid without latching his “fear” onto an object, his utterances gradually loose their meaning. The use of the word “fear”, about an objectless emotion, is therefore dependent upon its use in cases where fear has an object.\textsuperscript{42}

\textsuperscript{41} Kenny, 1963, p. 60.
\textsuperscript{42} Kenny, 1963, pp. 60-62.
For the sake of the argument I will temporarily grant this to Kenny. But the transition from object directedness to the necessity of belief still needs to be justified. Kenny’s basic premise is that not only do emotions have objects, but the concept of each emotion limits what kinds of objects it can have. He makes this point in terms of his notion of a “formal object.”

The formal object of φing is the object under that description which must apply to it if it is to be possible to φ it. If only what is P can be φd, then “thing which is P” gives the formal object of φing.43

Some descriptions of formal objects are trivial. Only edible things can be eaten and only inflammable things can be burnt. “Things which are edible” thus gives the formal object of eating, and “things which are inflammable” that of burning. But these are descriptions given merely by modalizing a verb and do not specify any characteristics the objects must have. Other descriptions are not trivial in the same way. Only dirty things can really be cleaned and only wet things can really be dried. Cleaning a clean thing is not really cleaning, and drying a dry thing is not really drying. “Things which are dirty” therefore gives the formal object of cleaning, and “things which are wet” that of drying. These descriptions limit the range of possible objects by specifying particular characteristics the objects must have.

Kenny then contends that emotions have non-trivial formal objects. The formal objects of emotions can not be given merely by modalizing the emotion terms. The formal object of envy is not only “Things which are enviable”, it must be described as something like “Persons who possess something oneself does not possess but desires to possess.” What

43 Kenny, 1963, p. 189
kinds of objects an emotion can have is thus limited by characteristics given in the
description of the formal object of the emotion. According to Kenny’s contention, it is not
merely a matter of contingent fact that people are not angered by being benefited and do not
become envious of the poor and wretched, it is given in the concepts of anger and envy.

But here is an important twist, namely that it is in fact possible to be angry toward one’s
benefactors and it is possible to envy the poor and wretched. But, claims Kenny, being
angry at one’s benefactors is possible, only if one believes the help actually harms or that it
is offered with questionable intentions, and one can envy the poor only if one believes
poverty brings happiness. What is impossible is to be angry for being helped when one
believes one actually benefits from being helped and that the help is offered with the best of
intentions, or to envy those one believes to live a miserable and wretched life due to their
poverty. “Only what is wet can in fact be dried; but something which is merely believed to
be an insult may provoke anger.”

It follows from this, argues Kenny, that the description of a formal object of emotions
must contain reference to a belief. A modified description of the formal object of envy
would be “Persons one believes to possess something that oneself does not possess but
desires to possess.” Hence the idea that to any emotion concept there is logic such that to
say one has emotion x toward z implies that one believes such and such about z.

Kenny’s surprisingly influential argument can be dismissed rather lightly. One need
only point out that, from the fact that something may be an object of an emotion in virtue of

---

the subjects beliefs about it, it does not follow that whenever the thing is the object of the emotion the subject has the required type of beliefs. A belief about a thing might be only one of potentially many ways of making some thing an object of an emotion. All that follows from the possibility of envying the miserable and being angry at one’s benefactor, is that the formulation of a formal object of anger and envy and other emotions, must allow that a thing can, appropriately or inappropriately, be the object of an emotion in virtue of either true or false beliefs held by the subject. The description of a formal object of emotions needs not exclude other ways of directing an emotion to an object.

I have not yet questioned Kenny’s claim that the concept of an emotion sets limits on the types of objects it can have. To give accounts of the object directedness of emotions, without referring to beliefs, we may not need to sever the tie Kenny believes there to be between emotions and types of objects. If we were to sever the tie completely there would be nothing against imagining a society where the emotions have gone topsy turvy. Let us assume that there is a set number of emotion types and that in our society we are confident about what kinds of situations or objects call for each type of emotions. If the tie between the types of emotions and the types of objects is purely contingent we can imagine a society where the pairing of emotions and objects has been shuffled. In that society gratefulness might have been replaced by anger, such that anger would typically and appropriately be directed towards one’s benefactors (to what we are now typically thankful for), and hope might have been replaced by jealousy, such that we would feel jealous towards possible future pleasures (to what we now typically hope for). I am here, of course, talking about a conceptual possibility. This is very likely a practical impossibility, because the shape of our
emotional life is to a large degree determined by a natural basis which might prove impossible to overcome. A dog is by nature prone to anger towards other dogs trespassing on his territory and a child is by nature prone to fear at the sight of an attacking elephant. But the interesting question here is whether this shift is a conceptual possibility? And if not, why?

A hint towards the answer might be found in looking at the way we learn emotion concepts. Kenny has an account of the learning of emotion words, based on Wittgenstein’s description of the way we learn to replace natural expressions with verbal expressions. Wittgenstein wrote:

How do words refer to sensations? [...] Here is one possibility: words are connected with the primitive, the natural, expressions of the sensation and used in their place. A child has hurt himself and he cries; and then adults talk to him and teach him exclamations and, later, sentences. They teach the child new pain-behaviour.⁴⁶

Kenny applies this to the learning of emotion words. When a child comes running to his mother, crying and trembling, she will tell him not to be afraid. But, says Kenny, here we meet a difficulty. Because there is no pattern of behavior common to every manifestation of a single emotion, no matter what its object, we cannot say that emotion-words are taught as a replacement of emotional behavior. For no matter how we describe a piece of behavior, claims Kenny, “it will only be emotional behaviour if it occurs in the appropriate circumstances.”⁴⁷ If a child cries we can not know whether the cry is a pain behavior or an

⁴⁶ Wittgenstein, 1958, #244.
⁴⁷ Kenny, 1963, p. 66.
emotion behavior unless we know whether it is crying because it bumped its head or because it has been left alone.

The language of emotions must therefore be taught in connection not only with emotional behavior, but above all in connection to objects of emotion. It is in connection with fearful objects, pleasant tastes, and annoying circumstances that the child learns the verbal expressions of fear, pleasure, and anger.48

What are we to conclude from this? Kenny is probably right in his contention that no emotion is always manifested by the same behavior. But the same goes for mere physical pain, and yet we are not tempted to draw the same conclusions about pain as Kenny does about emotions. Pain can be expressed in various different ways, but that does not mean that the concept of pain is essentially tied with causes or objects of pain.

Another thing to notice is that although no emotion is always manifested in the same behavior, some behavior or expressions are always or nearly always manifestations of the same emotion. Bodily gestures, such as an angry threatening pose, or a defensive crouch are signs of anger and fear. Facial expressions also give away emotions. An angry face is as easily recognizable as a joyful one. Bodily gestures and facial expressions can indeed be faked, and may occur without the emotion they normally manifest. But the correlation between these expressions and the emotions is strong enough for us to attribute emotions on the basis of them alone, given that we have no special reasons not to do so. It seems therefore to be possible to recognize some instances of emotions without them having objects. Claims that one is angry or afraid, without being angry about, or afraid of anything

particular can make sense without the use of these emotion terms relying on cases where the emotions have objects. I revoke the concessions I have made to Kenny. At least some emotions, i.e. emotions of types to which an expression is sufficiently closely tied, can be taken to occur without having an object. And from that it follows as well that the tie between these instances of emotions and types of objects is contingent. It could plausibly be argued that emotion types are somehow inherently tied to types of objects, but that does not support the stronger essentialist claim that every instance or token of an emotion type must have an object of the requisite type.

Now it remains to be shown how an emotion can be directed at an object without a belief about the object, or at least without its directedness having to be explained in terms of beliefs. The cognitivists are right in their criticism of feeling theories: object directedness of emotions can not be properly accounted for within the framework of feeling theories. But the problem is not that feeling theories exclude beliefs from emotions proper, but rather that they do not allow us to think of expressions or tendencies to behavior as a part of the occurrence of emotions, i.e. that they treat all concomitants of feelings as mere contingencies. There is nothing in a simple sensation that is directed at anything, but emotional actions and expressions can be, and typically are directed towards objects.

Allan Gibbard suggests a way to break out of the deadlock of feeling theories without having to resort to cognitivism. He suggests that we think of emotions as biological adaptive syndromes. Their symptoms are typical expressions and tendencies to action, and each of them typically responds to a special class of circumstances. Gibbard trivializes the
question of what exactly constitutes an emotion. He looks at a typical example of animal emotions, the anger of a dog towards another dog.

Think of a dog whose territory is approached by another dog. The dog stands up, it takes a special kind of stance or it runs back and forth, it barks, and it is primed to attack if the other dog keeps approaching. This story combines features of various kinds: a cause (territorial intrusion), expressive behavior (barking and taking threatening stances), and other behavioral tendencies (the dog is primed to attack). 49

We can also assume that the dog perceives some salient features of the situation and feels some kind of heated agitation. Where in all this is the emotion? This is clearly a story about anger and the anger is directed at an object. But yet we are not forced to, and might actually be reluctant, to attribute a belief to the dog. The directedness of its anger is sufficiently accounted for in terms of the directedness of its behavior and expressions. When a human reacts angrily to an insult or an intrusion we will indeed be strongly tempted to attribute beliefs to her, but we do not need to resort to these beliefs to account for the object directedness of her anger. It may even be questioned whether the beliefs and evaluations we can justifiably attribute to her, would, in the absence of angry expressions or (possibly suppressed) behavioral tendencies, suffice to constitute anger.


One argument against feeling theories went as follows: Emotions can be assessed as being either rational or irrational, appropriate or inappropriate, justified or unjustified. Mere

49 Gibbard, 1990, p. 132.
feelings cannot be so assessed. Therefore emotions are not mere feelings. This is, strictly speaking, a valid argument, but the premises need to be scrutinized before we can accept them. What kind of rationality is it that we assign to emotions and why can it not be attributed to feelings?

The kind of rationality we are attributing to emotions is closely analogous to epistemic warrant of belief. It is the sense of rationality in which it is irrational (in spite of a possible gain of some kind) to be angry at the messenger who brings the bad news he had no part in making, and the sense in which it does make sense to cry over spilled milk (in spite of its alleged uselessness). We do normally not have in mind pragmatic or instrumental rationality. For an emotion to be rational in the sense we are concerned with, is not for it to be desirable or advantageous. Anger can be rational in the one sense while it is not rational in the other. It can make sense to want not to be angry (because it is disadvantageous) while anger is fully appropriate and rational (because one was unjustly offended).

The most plausible and convincing reading of the argument I could think of was the following: The norms for the rationality of emotions, i.e. norms for what kinds of circumstances warrant them, are given in the concept of each emotion. Thinking of emotion terms as referring to simple sensations does not allow the conceptual depth needed to make sense of this. If emotions are simple sensations they can at best be contingently or conventionally tied to norms. It would be conceivable that it would be considered rational to be angry at ones benefactor, when gratitude would in fact be in order, and to envy the poor and wretched, which one ought in fact to pity. Therefore emotions can not be identified with simple sensations.
This argument can only purport to refute feeling theories. An additional argument is needed to justify the claim that, since emotions are subject to assessment of rationality, they *necessarily* involve beliefs. Errol Bedford offers such an argument. He contends that when emotions are inappropriate to circumstances they can only be accounted for in terms of beliefs involved.

Suppose that B does something that is to A’s advantage, although A thinks that it is to his disadvantage [e.g. B, a solicitor administering A’s affairs, sells some shares that A believes (wrongly) will appreciate]. Now it would be misleading to say simply, except to a fully informed audience, “A resents what B did”—this surely carries the incorrect implication that B has injured A. To guard against this it is necessary to add “but his resentment is quite unjustified,” or some equivalent expression. A’s belief that B has done something that affects him adversely, is, however, a necessary condition if the word “resentment” is to be used at all.\(^{50}\)

One could raise the immediate objection to this that there are cases of irrational emotions whose irrationality cannot be explained in terms of false beliefs held, but, on the contrary, by their going against, or not responding to actual, rational beliefs. Imagine that you become scared when you are standing on the edge of a cliff, although you are in no danger of falling. Suppose you do not know that you are perfectly safe (there could be an invisible, unbreakable wall of glass on the edge). In that case the fear is fully rational, given your information. But if you do know that you are perfectly safe, and you feel afraid anyway, the fear is irrational. It goes against a true, rational belief and its irrationality can therefore not be accounted for in terms of a false or irrational beliefs.

\(^{50}\) Bedford, 1984, p. 273.
John Deigh offers another argument against the transition to cognitivism. He points out that not all experiences of emotions are subject to assessment of rationality. The emotions of young babies and of animals that lack reason are obvious examples. It makes no sense to ask whether fear in an infant or a rabbit is appropriate or inappropriate. If a creature lacks reason, Deigh argues, "it lacks the faculty whose operations are presupposed in descriptions of states of mind as rational or irrational, reasonable or unreasonable." These emotions are instinctive reactions to a perceived threat, and they neither involve operation of reason nor are they responsive to norms. The conclusion to be drawn from this is that our practice of assessing emotions as rational or irrational does not provide sufficient ground for a general conception of emotions.

Cognitivists do not seem to be bothered by this seemingly obvious objection. They have either ignored it or offered some reasons for excluding these (alleged) emotions from the range of conditions they are dealing with. Ronald de Sousa wrote animal emotions off as mere responses that fall short of being full-fledged intentional states. Robert M. Gordon describes animal emotions as syndromes consisting of transfixed attention, overt behavior, and autonomic changes in the subject's physiological condition. Fear in an animal is thus a mere "flight arousal syndrome", a "state of fear", which falls short of being actual fear. State of fear is not essential to proper fear and only occasionally accompanies the human, essentially cognitive attitude of fear. Gordon limits his emotion theory to human emotions, which are essentially cognitive or propositional.

51 Deigh, 1994, p. 847.
Generally, the main reason the cognitivists offer for excluding animal "emotions" from the study of emotions is that they lack what cognitivists take to be the requisite type of intentionality, i.e. the intentionality of propositional beliefs, or in de Sousa's case, they are incapable of "singular reference" which requires "certain essentially logical or languagelike resources."\(^54\) This smacks of question begging. As we saw earlier the fact that emotions are typically directed at an object is not a sufficient ground for inferring propositional belief or thought. Object directedness does not imply cognitive intentionality. Deigh claims that, "the argument is merely a variation on the ... fallacy of inferring propositional thought from intentionality. In this variation it is a fallacy of equivocation."\(^55\) Intentionality can be a property of emotions whether or not their subjects possess reason, and if, as de Sousa claims, emotions require singular reference, that should pose no problem; a dog's anger toward another dog is clearly directed toward that dog and none other. If a singular reference is required it seems to follow only that a dog is incapable of anger toward an absent dog, to whom he can not express his anger. Similarly the dog could only fear immediate dangers, but not future ones.\(^56\) The cognitivists therefore need a further justification for excluding from their study, emotions which do not involve reason or propositional thought. They need to show that propositional intentionality is a defining feature of emotions, and that therefore the threatening behavior of a dog toward a trespasser is not an expression of an emotion

\(^{54}\) De Sousa, 1987, 98.

\(^{55}\) Deigh, 1994, p. 848.

\(^{56}\) De Sousa's idea of the singularity of reference rests on a contestable causal-externalism about linguistic reference, which I have some sympathy for. But I would rather not have a theory of emotions rely heavily on it.
(assuming we do not want to attribute a propositional thought to the dog) or else they must acknowledge that their study is only a study of a particular type or aspect of emotions.

Some philosophers have chosen the latter direction. They intend their study only to apply to human psychology and thus only to concern the thoughts and feelings of rational beings. But that move requires a justification as well. What is it about human psychology that warrants this special attention? What most clearly separates humans from animals, and mature humans from young babies, is the special importance of language in human life. As we mature we learn to speak, we learn to encode our thoughts in language and we develop a system of beliefs. Gradually the system of beliefs becomes influential in shaping our feelings and experiences. It therefore makes sense to pay special attention to beliefs in the study of human thought and feelings, and to assume that the process of thought in humans often follows logical relations among beliefs. This provides a plausible distinction between human psychology and animal or infant psychology and possibly a justification for restricting the scope of emotion theory to human emotions.

But we can still question the inference of belief from the rationality of emotions. Deigh argues that,

the difficulty would remain, because the distinction, being based on the observation that thought is ubiquitous in human experience, implies only that belief is always present in the experiences of human emotion and not that it is, in every case, an essential element of such experiences.\(^{57}\)

\(^{57}\) Deigh, 1994, p. 850.
An emotion can be infused with beliefs, and yet be intelligible without reference to the beliefs that infuse it. Some emotions may not be intelligible without reference to beliefs, such as the fear of a plummeting stockmarket. We cannot conceive of such fear without attributing beliefs about stockmarkets and finance, to the subject of the fear. But some other emotions, such as the fear of falling when one looks over the edge of a cliff, are intelligible regardless of what the subject believes. The former of these two fears is subject to assessment of rationality (it is as rational as the subject's beliefs about the stockmarket), whereas the latter, if it does not involve beliefs, is not. It seems then that even if we allow the cognitivists to ignore the emotions of animals and infants, not all emotions are subject to assessment of rationality. The rationality of human emotions is therefore an insufficient ground for arguing that emotions necessarily involve beliefs or propositional thoughts.

III. 4. Beliefs and the Individuation of Emotions.

Both my objection against Errol Bedford's argument and John Deigh's more general objection against the transition from rationality of emotions to cognitivism are based on the assumption that we have criteria for attributing emotions, which does not rely on beliefs. A cognitivist might therefore be tempted to ignore these objections and claim that emotions can only be identified in terms of beliefs. But this is not in accordance with common practice. We commonly judge from people's behavior and physical expressions, not from beliefs they express, what emotions they are experiencing. The onus is therefore on the cognitivists to show this common practice to be misguided, or to show that whenever we are justified in attributing an emotion we are at the same time justified to attribute the
corresponding belief. The question here is, do we need beliefs to identify emotions and
discriminate between them, or can we attribute an emotion without attributing a belief?

It has been argued that we cannot distinguish between different emotions by the way
they feel. Jealousy may feel just like envy and embarrassment just like shame. Two
instances of the same emotion type may also feel different. Fear, for instance, may be felt as
agitating or as paralyzing. Feeling theories, identifying emotions with particular sensations,
do therefore not allow us to classify emotions into anything like the classification given in
the common language of emotions.

Some have argued that this may not be a sufficient reason to reject feeling theories.
Feeling theorists might claim, in defense of their theory, that the distinctions we commonly
make among the emotions are artificial and do not correspond to any real differences.
William James complained that philosophers and psychologists treat emotions to much as if
they were "absolutely individual things . . . eternal and sacred psychic entities, like the old
immutable species in natural history."\textsuperscript{58} Perhaps so, but a theory that allowed us to hold on
to our folk psychological or common language taxonomy of emotion and made sense of the
distinctions between shame and embarrassment, envy and jealousy, etc., would have a \textit{prima
facie} plausibility which a Jamesian or a Humean theory lacks. I am furthermore inclined to
ask what makes a Jamesian theory, if it must cut across all lines and distinctions we
commonly make among the emotions, a theory of emotions at all, but not a theory of
something else?

\textsuperscript{58} James, 1950, p. 449.
But do these considerations necessitate cognitivism or do we have other ways to distinguish emotions? For the feeling theorists, the common and constitutive property of every instance of some particular type of emotion, is its felt quality. For the cognitivists, on the other hand, it is the type of thought the emotion is said to entail. Any experience of guilt, for instance, includes the thought that one has done something wrong. This thought identifies the experience as one of guilt rather than, say, shame or embarrassment.

Jerome Neu, in arguing against Hume’s feeling theory, wrote:

One cannot move from the ‘very port and gait of a swan’ to the claim that it has a concept of self. And if one cannot claim that, one cannot, in logic, claim that it feels pride. Indeed, even for humans, port and gait or other behaviour is only evidence of a particular emotion in the context of certain beliefs ascribed to the subject. . . . [A]ny behaviour (almost) can express any emotion. Even in a creature capable of a particular emotion (say, anger), a bit of behaviour which might usually be typical of that emotion (say, hitting), will only be an expression of that emotion under certain conditions, which include the thought behind it. The particular emotion being expressed (if any) will depend on the particular thought; just as the character of an action depends on the intention.  

I have already criticized feeling theories on the ground that mere feelings do not suffice to determine which emotion one is experiencing. Here Neu makes the further claim that one’s behavior, gestures or expressions are also insufficient. Any behavior can express any emotion. Therefore we can take a piece of behavior to express a particular emotion only if we can attribute to the subject, a thought that makes the behavior an expression of that emotion rather than some different emotion. This is the argument Neu gives here, but it is

59 Neu, 1977, pp. 52-53.
clearly incomplete. It needs the further premise that expressions and behavior, on the one hand, and thoughts on the other, are the only alternatives by which to identify and distinguish between the emotions, after feelings have been discarded.

But even given the added premise we may still doubt whether the conclusion is warranted. Any behavior can be an expression of any emotion. I grant that to Neu. But are not some behaviors and expressions so strongly tied to particular emotions that we are justified to attribute those emotions on the basis of the behavior alone if there are no special reasons not to do so? We attribute anger or fear to dogs on the basis of their behavior alone. Why should we not do the same to humans? Maybe because it makes no sense to attribute either pretense or hidden, complex reasons to a dog, whereas we can never be absolutely sure what motivations lurk in a human mind. When a human A hits human B it makes sense to ask why he did it. Was he angry or was he perhaps being affectionate to his masochist friend? Maybe it was jealousy but not anger? Or could he even have been expressing his gratitude to C who hates B? But when a dog attacks an intruding dog we assume right away that he was angry, because we are not willing to attribute to him anything more complex or sophisticated than a reaction to a perceived intrusion.

But sometimes it makes no sense to ask these questions about a human reaction or expression of an emotion. Gibbard’s story of the dog involves descriptions of a situation and of a spontaneous response. The emotion has an object typical to anger (an intruder) and a typical angry expression (a threatening stance). We can tell similar stories about humans, stories of spontaneous reactions, typical of specific emotions, to circumstances which can be described as presenting objects which are appropriate for the same emotions. Fear felt on
the edge of a precipice and the fear of big spiders are examples of this. These fears can take hold of people faster than they can have the thought of danger, and even in the face of the knowledge that there is no danger.

A cognitivist reaction to this would be to claim that, insofar as we are justified in attributing emotions, we are equally justified to attribute thoughts. Neu writes:

If we are prepared to ascribe certain desires (as opposed to bare needs) to non-language-users (e.g. animals and infants), we need not hesitate to ascribe certain thoughts to them as well.⁶⁰

A few pages later he implies that we can attribute fear of immediate danger to a dog, but not fear of dangers in extended future, because the dog does not have the “conceptual equipment” to have thoughts about the future.⁶¹ Since Neu mentions infants (who have not had the time to learn concepts), the conceptual equipment at work here is presumably innate and non-linguistic; we, as well as some other animals, are born with certain concepts which play a role in instinctive desires and emotions.

Neu is probably among the cognitivists John Deigh accuses of confusing being sensible of something with having the concept of it. Deigh writes:

To be sensible of a property is to be able to detect its presence and to discriminate between those things that have it and those that do not. To have the concept of a property, by contrast, is to be able to predicate it of some object and, hence, to locate it in a system of propositional thought.⁶²

⁶⁰ Neu, 1977, p. 44.
⁶¹ Neu, 1977, p. 52.
⁶² Deigh, 1994, p. 840.
Predication presupposes some system of propositions. This system is realized in an organization of thought that having a concept of some property implies. Such organization of thought is not necessary for detecting a property and discriminating between things that have it and those that do not have it, i.e. the perceptual ability to detect and discriminate does not require conceptual understanding. That an animal or an infant can sense danger in certain circumstances does therefore not imply that it has the concept of danger. We are not forced to attribute a thought to an animal when we attribute fear to it.\(^{63}\)

To this a cognitivist might reply that when an animal is aware of danger, it is aware of an abstract property, but not a sensory one, and that the awareness of an abstract property requires having the concept of it. Deigh has a quick response to this. We can reasonably suppose that the sensibilia, to which an animal reacts when it becomes aware of danger, are fixed by natural selection and the range of experiences characteristic of its environment, and they do not generalize to other dangers. An antelope is predisposed by nature to fear a natural predator, but not a handgrenade. It is, in other words, not the abstract property of dangerousness of which an animal is aware, but rather fixed, concrete properties, which through natural selection have come to be perceived as being scary. A thing or a situation is scary to a creature if the creature is predisposed (presumably neurologically) to be afraid of it. To be aware of and react emotionally to features of the environment does therefore not require a conceptual understanding of these features. These features are the keys to the locks in William James’ metaphor, and a lock need not have the concept of the key by which it can be opened.

\(^{63}\) Deigh, 1994, p. 840.
III. 5. Beliefs and Emotions in Conflict.

It is a problem shared by all belief-cognitivist theories, as different as Nussbaum’s, Solomon’s, Snow’s and Kenny’s, that emotions often appear to conflict with explicitly avowed beliefs. Our beliefs do not always accord with our feelings. Cheshire Calhoun tells us the story of Tess, who knows that spiders are harmless, but still fears them, knows there is nothing wrong with homosexuality but still experiences shock and revulsion when she finds out that her friend is a lesbian. Her beliefs and her emotional reactions are in conflict. There may indeed be cases where the conflict is only apparent, when it turns out that, “the person’s avowed beliefs are mere pretenses or product of self-deception; the emotion actually fits the real, unavowed beliefs.” But this needs not be so in all cases. One can be perfectly confident and be able to argue and give convincing evidence for one’s avowed beliefs, and yet consistently experience conflicting emotions.

Several attempts have been made to solve this problem within a cognitive framework of emotions. One is to claim that in cases of emotion-belief conflicts we actually have two opposing beliefs, namely our explicitly held belief and the belief manifested by the emotion. Cognitivists here deny that people always live up to the ideal of rationality. People do not hold a unified, consistent system of beliefs, but may hold different and inconsistent beliefs, perhaps at different levels. Emotional beliefs are thus somehow independent of beliefs held from a critical reflective viewpoint. Surely this may often be the case, but are we entitled to attribute beliefs to people on the basis of their emotions alone, without any other evidence?

---

64 Calhoun, 1984, p. 331.
65 Calhoun, 1984, pp. 331-332.
Calhoun notes that we could reasonably charge Tess with inconsistent beliefs if she vacillated in her belief-avowals and her emotional responses. But if there is no other evidence for a belief conflicting with the avowed belief, then attributing a belief on the grounds of the emotion alone would surely be to beg the question in favor of cognitivism.

Calhoun tries to solve the problem of belief-emotion conflict by claiming that in cases of this kind, the explicitly avowed belief is defective in an important way. It is held only intellectually (based on inference or authority), but not evidentially, i.e. it is not, in Calhoun's words, "bourne out by our experience." This allows one to emotionally "see things as...", which Calhoun believes is constitutive of emotion. Tess thus "sees" spiders as being dangerous and gays as being revolting, in spite of her intellectually held beliefs that spiders are harmless and gays are okay. In this manner Calhoun holds on to cognitivism, but rejects jugdmentalism or belief-cognitivism. The cognitive element in emotion is, on this account, not a belief or a judgment, it is a "seeing as...". Calhoun's article is aptly named with a question, "Cognitive Emotions?" She answers the question affirmatively, but what I believe she has actually shown is not that emotions are cognitions, but rather that we do not have a sharp notion of cognitions to apply to all emotions and that to attribute cognitions to many instances of emotions we have to stretch the notion of cognition far beyond any ordinary usage of the word. We are left with the question whether we are any better off with "seeing as..." than with a "feeling as...", or even whether "seeing as...", which is not actually believing, is anything but "feeling as..."?

---

What we have here then is something that is not an actual belief but has the epistemic content of regular beliefs, and is manifested only in an emotional state or a reaction. We have slid back into traditional cognitivism, and it seems to me that at this point we encounter the problem of introspection again. Calhoun has to show what constitutes seeing something as being one thing rather than another. What makes the experience of “seeing as...”, in a particular instance, one emotion rather than another? If a “seeing as...” is to count as guilt its content has to be specific enough to be recognizable as guilt rather than shame or embarrassment, if it is jealousy it must be distinguishable from envy and so on. If it is not an actual belief and not even a thought that is entertained, we seem to have to fall back on the emotional experience itself, its physiological concomitants or its spontaneous expressions. If there is no spontaneous expression or if the expression is not clearly distinct from the expressions of other emotions we have nothing but the affects to refer to, the experience or the physiology. Here we come across some of the same problems as we found with feeling theories, problems with introspective individuation of emotions, that cognitivism was designed to avoid.
CHAPTER IV.

NON-ESSENTIALISM ABOUT EMOTIONS.
IV. 1. Introduction.

I have so far rejected the traditional emotion theories individually. But it is tempting to ask whether there is a deeper problem common to their approach to the emotions, independent of the various theories. Did they perhaps ask the wrong question?

The question that has driven much of emotion theory is the old Socratic question of what unifies the many and different under one banner (Socrates asked what unifies the different virtues as virtues, etc.). The theorists have searched for unity in the plurality of emotions, a description that applies to them all. Socrates'/Plato's answer was that the many and different were unified by being copies of the same form or idea. The emotion theorists have answered, not in the spirit of Socrates, but of Aristotle. Aristotle's essentialism is the theory that some types of objects, no matter how described, have essences; that is they have certain properties, without which they could not exist or be the things they are or belong to the type they do. The traditional emotion theories, as different as they are, are an attempt to answer the same question and they consequently share the same basic structure. They are essentialist theories, they ask what property it is that makes a condition an emotion, and without which the condition would not be an emotion.

Essentialism does not imply that the essence of emotions is a single, tangible thing, element or property. Some of the essentialist theories require a complex structure for an emotion, a combination of belief and desire a belief; desire and an affect (e.g. Justin
Oakley); or a belief supplemented by an evaluation, which is not strictly cognitive (e.g. William Lyons). They can even emphasize a process or change as essential to emotion. But they all share the view that there is a general account of emotions to be discovered or revealed, one, perhaps abstract, description that applies to all emotions and sets them apart from other conditions.

I do not believe there is an *a priori* argument against essentialism to be found. But the discussion I have given of the problems that arise in identifying, explaining and characterizing the endless variety of emotions, in terms of the various essentialist models, can be taken as an indication that perhaps we ought to try a different way of thinking about emotions, a way that allows us to appreciate the similarities, relations and continuity among the different emotions, rather than compartmentalizing them in ways that blind us to these.

Patricia Greenspan is aware of the difficulties created by the immense variety of emotions.\(^67\) In her analysis she tries to avoid the Procrusteanism employed by so many other philosophers, but does so at the cost of becoming overly abstract. She analyses emotions as being a compound of an evaluative proposition and comfort or discomfort about that proposition. She allows for a variety of comfort or discomfort for each emotion and concedes that the evaluative proposition need not involve an actually held belief. The features of the internally related affect and evaluation can differ dramatically from case to case. But she admits that description of the variety of cases in accordance with her analysis will often sound stilted and or simply uninformative.\(^68\) So perhaps she has not entirely

\(^67\) Greenspan, 1988, ch. 2.
\(^68\) Greenspan, 1988, p. 15.
avoided Procrusteanism after all. Procrustes did not only chop the feet of those who were too long for his bed, he also stretched, and thus tore apart, those who were too short. Similarly, Greenspan does perhaps not chop as eagerly as some of her colleagues, but she runs the risk of stretching some emotions to fit her analysis of emotions as affects felt about an evaluative proposition. That is an equally undesirable type of Procrustean butchery. Why should we bother with a stilted and uninformative analysis of emotions?

I believe there is a more fruitful approach to emotion theory, which avoids both chopping and stretching and allows us to appreciate the similarities and relations as well as the differences between emotions. That is to treat emotions as having a family resemblance rather than an essence, and accept the fact that nothing unifies them, unless we are willing to resort to a highly abstract account, in terms of which descriptions of some emotions will sound “stilted and uninformative.” Thinking about emotions in terms of family resemblances allows us to stay closer to the descriptive ground.

IV. 2. Emotions and Family Resemblance.

I borrow the idea of a family resemblance from Wittgenstein’s later philosophy. In *Philosophical Investigations* he pictures someone complaining that in all his discussion of language games he has not explained what the essence of language is.

And this is true.-Instead of producing something that is common to all we call language, I am saying that these phenomena have no one thing in common which makes us use the same word for all,-but that they are related to one another in many
different ways. And it is because of this relationship, or these relationships, that we
call them all “language”.\textsuperscript{69}

The novelty of this idea is that Wittgenstein discards the Socratic presumption that there
must be something common to all language in virtue of which it is language. He uses games
as an example to illustrate this. Wittgenstein does not proffer an \textit{a priori} argument to show
that there is nothing common to all games. He only suggests that we look and see whether
there is anything common to all.

Consider for example the proceedings that we call “games”. I mean board-
games, card-games, ball-games, Olympic games, and so on. What is common to
them all?-Don’t say: “There must be something common, or they would not be
called ‘games’ ”-but look and see whether there is anything common to all. For if
you look at them you will not see something that is common to all, but similarities,
relationships, and a whole series of them at that.\textsuperscript{70}

Robert J. Fogelin gives a crude representation of the idea of family resemblance using
the following diagram:

\begin{center}
\begin{tabular}{cccccc}
01 & 02 & 03 & 04 & 05 & 06 \\
A & B & C & D & E & F \\
B & C & D & E & F & A \\
C & D & E & F & A & B \\
D & E & F & A & B & C \\
\end{tabular}
\end{center}

0\textsubscript{1} through 0\textsubscript{6} represent a set of objects and the letters A to F represent properties they
possess. Each object shares three properties with two of the other objects in the set, but no

\textsuperscript{69} Wittgenstein, 1958, #65.  
\textsuperscript{70} Wittgenstein, 1958, #66.  
\textsuperscript{71} Fogelin, 1987, p. 133.
single property is shared by all the objects. This is roughly what Wittgenstein has in mind when he says that our examination of games will show "a complicated network of similarities overlapping and crisscrossing: sometimes overall similarities, sometimes similarities in detail." These similarities are what Wittgenstein calls "family resemblance."

Now, what does the idea of a family resemblance imply, or, perhaps more importantly, what does it not imply? It should be made clear that it is imaginable that all games might have something in common and still only have a family resemblance in what makes each of them a game. We could imagine that they all share some trivial feature, for example that they all have a sixty minutes time limit. But that is not a feature that makes them games, so if we were to make this discovery we would not thereby have found the essence of all games. Similarly, if we were to find a feature common to all emotions, it would not follow that that feature was the essence of emotions. The emotions might very well still be emotions without it. We could, and probably would, be wrong to think that feature essential to emotions. We have hitherto confidently used the term "emotions" without any knowledge of such a feature, so there are clearly other properties that make a condition an emotion.

What do we see if we treat emotions the same way as Wittgenstein treats games, i.e. if we, as Wittgenstein suggests, just look and see, and do not think? When we resist the temptation to presume that something must unify emotions and set them apart from other conditions, we will most likely (like Wittgenstein in his make-believe search for the essence of games), find that there is no answer to the Socratic question. We will find no one thing

---

72 Wittgenstein, 1958, #66.
which unifies them and makes each of them an emotion, but only “similarities, relationships, and a whole series of them at that.” We will also find that nothing sets them clearly apart from other conditions. Between emotions and things like moods, attitudes, motives and character traits there is a gray area in which it is not clear, or rather there is simply no fact as to whether a condition is an emotion or a mood, attitude, motive or a character trait.

We need only to review some of the examples I have used in discussing the traditional theories.

1. When we see a snake on our path we may undergo a variety of physiological changes associated with fear and flee or even get paralyzed, without feeling anything at the time.

2. John can be angry at Smith for years, without constantly feeling the anger or being in a physiological state associated with anger. The anger might be manifested only in John’s unintentionally shunning and ignoring Smith.

3. When we stand on the edge of a cliff we may feel fear, even if we know perfectly well that we are in no danger.

4. An antelope, incapable of conceptual understanding and propositional beliefs, fears an approaching predator.

5. One can fear a collapse of the stockmarket and take the appropriate actions, without undergoing any physiological changes, having any feelings of fear and without any typical fear reactions and expressions.

Examples one, two and five show that feeling is not a feature that unifies all emotions; examples two and five show they are not unified by physiological changes, and; examples
three and five show that beliefs and conceptual understanding are not necessary for all emotions to occur.

Someone might suggest that all these examples involve evaluations of some sort. One is afraid of the snake because one would not like to be bitten; John is angry at Smith because he thinks Smith behaved badly; we feel fear on the cliff, because we would not like to fall down; the antelope would not like to be eaten by a predator, and; the financial speculator likes to own money and therefore fears a collapse of the stockmarket.

In some cases this is surely right. But in the case of instinctive emotions, such as in examples one, three and five, it is questionable. The antelope's fear of the predator requires only a perception of a scary object, and an automatically triggered response. One might say that evolution has provided the antelope with an innate dislike of the predator, or even that evolution has evaluated an object and built the evaluation into an innate emotional response. This is a nicely metaphorical way of describing the evolutionary process (it has also been described as a calculator), but that is precisely what it is and all it is, a metaphor. The emotional response of the antelope needs not involve an act of evaluation. We can at best describe the antelope as behaving as if it disliked the predator and the prospect of being eaten. But that would only be a quasi-evaluation, and that would only give us a quasi-essence of emotions.

The search for a necessary and constitutive essence or a defining characteristic of emotions seems to come to an inevitable conclusion, namely that there is none. There appears to be no single necessary element and no significant description applicable to every
instance of every type of emotion, unless we are willing to bring in such impoverished essences as unfelt feelings/affects or unconscious cognitive perceptions or representations. Thinking of emotions as having a family resemblance seems a viable option. We can have emotions in a number of different ways and emotion words refer to a number of different elements and properties, some of which may be absent in individual cases and none of which is necessary in all cases of emotions.

IV. 3. Family Resemblance and Cognitive Emotions.

Thinking of emotions as having a family resemblance does not exclude the possibility that among the emotions there might be types with a necessary element or property. Wittgenstein points out that although numbers in general have only a family resemblance, there are rigorously defined types among them, such as the cardinal numbers, the rational numbers, the real numbers, etc. Among games there are indeed clearly defined types, such as some ball games, specified in terms of objectives and rules. The interesting question about emotions is whether there are clearly defined types among them, i.e. whether an essentialist model fits some emotion types.

In looking for types among the emotions we should avoid drawing arbitrary lines. We can in principle carve the emotions up in any way we like, but might end up with useless categories, like the zoological category of gray animals. It may serve a specific purpose, perhaps an aesthetic purpose, to draw a line between gray animals and non-gray animals, but in drawing that line we would not have discovered a type of animals. In emotion theory we
should restrict ourselves to trying to discover actual and significant lines among the emotions, but not postulate arbitrary ones.

Among the questions worth asking is whether each, or perhaps just some, of the folk psychological emotion types (i.e. emotions picked out by the same term in common language) have essences? Or are there even larger categories of emotions with an essence which sets them apart from other emotions? If so, do the larger categories cut across the folk psychological lines? I am not going to attempt to answer all these questions, but in the light of the current debate and the surge of cognitivism, the most interesting questions that arise here concern the place of cognitive emotions within the family resemblance model. Are there emotion types which necessarily involve cognitions of some kind, and if so, do they together belong to a genus of cognitive emotions?

The key to arguing against the necessity of cognitions in emotions was the empirical fact that we can identify and ascribe certain types of emotions without a belief and even contrary to explicitly avowed beliefs. Some emotions occur faster than we can form a corresponding belief or a thought, such as my fear of a loud bang. Other emotions last longer than the corresponding belief, and often emotions can easily be identified in the face of an opposing belief.

But how do we identify these emotions? Earlier I ruled out the possibility that emotions might be identified by introspection of feelings. Introspectively, different emotion types often feel the same and separate instances of the same emotion type often feel different. Introspection would therefore be too drastically at odds with common sense taxonomy of the
emotions, and would have to be taken as a classification of something else. Identifying particular emotions by physiological differences has also been ruled out. Experiments have shown for instance that the physiology of fear is, at times, indistinguishable from that of anger.

We have no reason to grant people first person authority for identifying their emotions. We need not accept a self attribution of guilt to someone who says: "I have no guilt beliefs, and there is nothing in my history, no past beliefs or values, that explain what I am feeling now as being (residue) guilt. But I know it is guilt.” How does one know, if there is nothing in the emotional experience or the outward expressions that distinguishes it from some other emotions, like shame or regret?

We seem to be left with only two possibilities; to identify emotions by beliefs or cognitions involved, or, as in the case of Tess and her fear of spiders and in a case of one jumping at a loud unexpected bang, by stereotypical spontaneous expressions or behavioral patterns. These expressions must be clearly and unambiguously identifiable as marks of particular emotions. A claim that an emotion can occur without or contrary to a belief, therefore seems to depend on its having an identifiable spontaneous expression or a behavioral pattern. Emotions that do not have such expressions can not be identified by anything but beliefs or judgments. Which emotions can credibly be said to occur without or contrary to a belief is going to depend on empirical evidence about which emotions do have a stereotypical expression.
To ascribe a non-cognitive emotion we need expressions or behavioral patterns that are sufficiently unambiguous, distinct and convincing, even when they are countered by an equally convincing refusal of a corresponding belief. This means that if, for instance guilt and shame, or compassion and grief cannot be identified and distinguished by their expressions or behavioral pattern, their identification must involve reference to beliefs or judgments. In addition to being unambiguous the expressions and behaviors must also be fairly involuntary. If an emotion is manifested, even only partly, in highly voluntary actions, a refusal of belief will not be credible. Someone who apologizes or begs for forgiveness (not only for the sake of etiquette or appropriateness, but from feeling guilty) cannot be taken seriously when he claims not to believe that he has done something wrong. His guilt clearly involves evaluative beliefs.

By these criteria I draw the line between what I call primary and secondary emotions. Primary emotions are emotions that can be identified on the basis of expressions and behavior. Secondary emotions are emotions that do not have distinct, spontaneous expressions. They are emotions which can only be identified by reference to beliefs or whose associated patterns of behavior involve highly voluntary actions, so that a refusal of a belief cannot be taken seriously. I have so far only given a stipulative definition of primary and secondary emotions, but have not yet contended that there exists such a thing as a primary or secondary emotion.

It might be argued that my definition of secondary emotions is empty. In response to my distinction someone might claim that any emotion can conflict with explicitly avowed beliefs, as is certainly the case with, for example, both guilt and shame. This is true, in a
way. But it does not follow that all emotions can be explained and identified without reference to beliefs. To use an example provided by John Rawls, a person brought up in a strict religious sect may have been taught that going to theaters is a sin. But even after giving up the beliefs of the sect he still feels guilt\(^{73}\) whenever he watches a play. Whether or not we are right to call such emotions guilt (or whatever secondary emotion is in question) is really a futile question and perhaps merely a matter of linguistic convention. Rawls chooses to claim that guilt feelings of this kind are not proper guilt feelings. The emotion of the former sect member when he sees a play is not proper guilt, says Rawls, “since he is not about to apologize to anyone, or to resolve not see another play.”\(^{74}\)

But if we choose, as I would be inclined to do, to think of a residue emotion like this as an incident of guilt, we must ask how it is identified as such? In Rawls’ example the answer is clearly that the emotion is identified as guilt by reference to previously held beliefs, i.e. even if the emotion can occur contrary to explicit, currently held beliefs, the emotion still requires reference to a belief for its explanation. The onus is on the “utterly-non-cognitivist” to identify, or to justify the attribution—whether self-attribution or attribution to others—of emotions which do not have a distinct expression, without reference to beliefs.

My cognitivism about the secondary emotions is thus not a claim that all instances of these emotions involve actual, currently held beliefs. It is not essentialism about a particular category of emotions in the strong sense that they must have a particular tangible property, but rather in a weaker sense that a full description and an explanation of their occurrence

\(^{73}\) I will assume for the time being that guilt does not have a distinct spontaneous expression.

\(^{74}\) Rawls, 1971, p. 482.
must involve a description of a particular type of beliefs and evaluations. I claim only that
emotions which do not have a distinct spontaneous expression, must refer to beliefs,—either
actual beliefs or a history of past beliefs--, and a notion of a full-fledged emotion of that
type, for their explanation and identification. Identifying a secondary emotion as being *as if*
one believed such-and-such, would not suffice, because if there is no actual belief or a past
belief, then there is nothing in the reaction or condition that is distinctly *as if* one, for
instance, had guilt-beliefs as opposed to *as if* one had shame-beliefs.

There are thus instances of secondary emotions where a person does not hold the
relevant beliefs. These residue, secondary emotions are significantly different from full-
fledged secondary emotions and they merit their names only by standing in certain kinds of
relationship to the full-fledged ones. The rationale for calling them guilt, shame, envy etc.,
is not that they feel somehow special or are manifested in a special response or a facial
expression, but that their occurrence is most plausibly explained with a story involving a
reference to previously held beliefs.

So far I have used the term "secondary emotion" as a collective noun for emotions that
must be taken to involve cognitions. But do they collectively form a genus or do they only
contingently share a property (like all gray animals share a property, without forming a
genus of gray animals) although that property may be essential to each of them?

Although I draw the line in epistemic terms, I am inclined to think that the division
between primary and secondary emotions is largely the same as that between natural and
cultural emotions. Beliefs and thoughts require language, and language is acquired through
acculturation. Emotion types which are always accompanied by thoughts and beliefs, or require reference to beliefs, seem therefore inherently tied to culture, they mark individuals as socialized and beneficiaries of cultural and moral education. Primary emotions, on the other hand, are detectable in both animals and infants and across very different cultures, and seem therefore to be (potentially) independent of acculturation. The experience of them does not presuppose the cultural transmission of language.

It is important to notice that my notion of cultural emotions does not entail that they are culturally relative or variant, i.e. that they might exist in one culture and not in another. Some cultural emotions are no doubt tied to particular cultures (consider Japanese *amae*), but there might be other cultural emotions which must be found in *any* culture. They might be necessitated by culture as such, or they might be necessitated by the very capacities which enable humans to have a language and a culture. They would thus be universal cultural emotions. People might want to call such emotions natural, but they would then be natural in a sense significantly different from the sense in which I call emotions natural.

My idea of natural emotions closely resembles the James-Lange model of emotions, although I believe we do not need to identify a natural emotion with any particular aspect of the emotional response. Evolution has endowed us with mechanisms that trigger certain responses to certain perceptions. We react to perceptual properties and do not need conceptual understanding of the circumstances. Natural emotions do not require acculturation and are not marks of cultural or moral education. This is consistent with Allan Gibbard’s account of emotions as adaptive syndromes and does not commit us to essentialism about natural emotions, because, as Gibbard wrote,
it may turn out, moreover, that more than one mechanism stands behind a syndrome, normally working together. In that case there may be no fact of the matter which of these constitutes the emotion: when all work together the animal has the emotion; when all are absent it lacks it—and otherwise there is no clear fact of the matter.\textsuperscript{75}

Between these most primitive responses and purely cultural emotions there is not a gap, but a continuum, most of which falls within my stipulative definition of primary emotions. Natural emotions can be learnt. Animals can, through repeated experience, learn to fear things they are not hardwired to fear. Cats, for instance, learn to fear garden hoses and volleyballs. But this does not necessarily mean that the animals are learning concepts or acquiring beliefs, i.e. these emotions are not signs of culture, but show only that animals can learn to perceive new things as being scary.

But natural emotions can also to be tied to culture and require conceptual understanding. We can learn to have natural emotions, such as fear, about things we know to be dangerous, although we are not naturally wired to perceive them as being scary. Fear becomes a response, not only to scary things, but to things that share the abstract property of scary things, namely dangerousness (in an extended sense of the term, which can include not only dangers of death and physical harm, but also of financial losses, diminished prestige, etc., i.e. losses of things we value by culture, but not by nature). We learn the common, abstract quality of scary things, namely their dangerousness. We then expand fear to respond to things which share the abstract quality of dangerousness, without having the concrete, perceptual quality of being scary. A cat's fear of a garden hose is an example of how an

\textsuperscript{75} Gibbard, 1990, pp. 133-134.
animal can learn to perceive new things as scary, without having a conceptual understanding. But a speculator's fear of a plummeting stockmarket is more than that. It requires a conceptual understanding of the abstract property of dangerousness, as well as a network of beliefs about the way the financial market works. These emotions thus have important features of cultural and secondary emotions. They require conceptual understanding and often guard or promote learned cultural values. But they are counted among natural emotion types because their characteristic objects share abstract qualities with things whose perception triggers purely natural emotions. An approaching predator, on one hand, and conditions which might cause the stockmarket to plummet, on the other, share the abstract quality of dangerousness, which explains why both tend to be occasions for fear, although one requires conceptual understanding while the other does not.

Now we reach the other end of the continuum, the secondary, necessarily cultural emotions. We learn through acculturation, possibly the mere acquisition of a language, to react emotionally to situations which do not correspond to triggers for any natural emotions, in the way that a handgrenade and conditions in the stockmarket can correspond to the dangerousness of a charging predator. Situations which we need conceptual understanding to grasp and which we by culture or by some cognitive aspect of human nature (which would explain the existence of universal cultural emotions), regard worthy of emotional reactions are somehow tied up to our basic emotional capacities and thus come to be triggers for emotional responses.

I have still not offered an argument to show that there actually are any secondary emotions. The best support I can give for my belief that there are such emotions is to cite
results of research into recognition of emotional expressions. Investigators, cited by Paul Ekman, do not fully agree on the number of the number of emotion types which are recognized across different cultures. There does seem to be general agreement among them about a few categories. They all found a happiness category (whether or not happiness can be considered an emotion), as well as categories for anger, surprise, sadness, and fear. All but one put disgust and contempt in one category and there is near agreement on a category for interest or attentiveness. Conspicuously absent from this list are emotions such as shame and guilt, envy and jealousy, as well as countless other.

I leave countless questions about cognitive emotions unanswered, among them some of the most pressing questions in contemporary emotion theory, much of which is carried out under the assumption that all emotions belong in this category. One question, to which I have no answer, is how emotional beliefs are tied up to our emotional capacities. Another question, more interesting from a philosophical point of view, is what, if anything, characterizes the beliefs essentially involved in secondary emotions? Are they, as Martha Craven Nussbaum claims, “fully accepted beliefs about what is valuable and important”? I believe they typically are about what is valuable and important, but that can hardly by itself make a belief an emotional belief. We can have beliefs about what is valuable and important without tying them to emotions at all. Are emotional beliefs then, as Robert Solomon suggests, “self-involved and relatively intense evaluative judgments”? Solomon’s suggestion seems to me rather implausible as a characterization of the beliefs as such.

---

76 Ekman, 1972, pp. 60-61.
77 Nussbaum, 1990, p. 293.
78 Solomon, 1976, p. 188.
because the intensity of emotional beliefs seems to me to be provided by the tie to the basic emotional capacities. That seems to suggest the possibility that perhaps emotional beliefs are characterized by nothing but the fact that they are somehow tied up to emotional capacities and tend to arouse emotional responses.

My contention is that there is not much more to be said about emotional beliefs than that they are typically about what is valuable and important, which in turn explains why they get tied up with emotional capacities. But to understand what is gained by tying them up with these capacities we need to understand what emotions typically do. That is what I turn to in the next and final chapter.
CHAPTER V.

FUNCTIONS OF EMOTIONS.

"Now if that being's preservation, welfare, or in a word its happiness, were the real end of nature in the case of a being having reason and will, then nature would have hit upon a very poor arrangement in having the reason of the creature carry out this purpose. For all the actions which such a creature has to perform with this purpose in view, and the whole rule of his conduct would have been prescribed much more exactly by instinct; and the purpose in question could have been attained much more certainly by instinct than it ever can by reason."

(Immanuel Kant, Grounding for the Metaphysics of Morals, Ak. 395)
V. 1. Introduction.

I have suggested that we suspend the search for a constitutive essence of emotions. A significant and informative description, applicable to all emotions, and defining of them as emotions, seems out of reach, and the search for such a description may even have inhibited understanding of some aspects of emotions. Perhaps Nico Frijda was right when he (in discussing the desire for revenge, i.e. wrath) wrote:

The fact that discussion is possible on whether desire for revenge is or is not an emotion is worth a moment's reflection. It shows the primitive state of our science that grapples with "substance concepts" in a true Aristotelian fashion, rather than fully going over to the side of function concepts, as in Galilean science. 79

I do not intend to analyze or define all emotions in terms of functions. I only wish to suggest that by looking at the functions of emotions we will gain understanding of an extremely important aspect of at least some emotions. I will exclude neither the possibility that non-emotional conditions could have the same functions as emotions, nor that there could be emotions which do not have a function that resembles my account in any way. A study of emotional functions should furthermore not blind us to the intrinsic qualities of emotions, i.e. qualities that they have regardless of their relations to other things. Functionalism in the philosophy of mind encourages us to ignore whatever intrinsic qualities mental states may have, and to think only of how they relate to other mental and non-mental

79 Frijda, 1994, p. 265.
states. It does not deny that mental states may have intrinsic properties, but as far as strict functionalism is concerned there are no limits to what the intrinsic properties can be. They could be feelings, brainstates or states of mechanical detectors. I do not want to take this step with regard to emotions. It is not impossible that non-emotional states could serve the same functions as emotional states, and that is something we would not be able to make sense of if emotions were to be understood solely in terms of functions, regardless of their intrinsic qualities.

So perhaps Frijda was not entirely right to suggest that we move fully over to the side of function concepts. Doing so could turn out to be yet another form of essentialism (broadly construed), if it assumes that there is to be found one description of functions, applicable to all emotions, such that conditions which did not have such a function were not emotions, and all conditions which did have such a function were emotions. Functionalist essentialism runs the risk of being just as Procrustean as substance essentialism; we might find ourselves describing emotions in an equally abstract, stilted and uninformative way as before, as well as being forced to include the unlikeliest things on our list of emotions. I, therefore, intend my account of emotional functions to find its place within the family resemblance model of emotions, rather than to replace it, and to add to the fabric of overlapping similarities and relations between the various emotions.

Looking for functions of emotions leads to several questions. One is what kind of functions we should look for; whether we should look for Darwinian functions, or functions
given to the emotions by the intention of human creators (like the function of a knife). In the traditional understanding, a function of a thing is what the thing is meant to do or the purpose it is designed to serve. This idea is easily applied to manmade objects. Their functions are given by their creators' intentions. A Darwinian function is different. Given that evolution did not take place according to a creators design or intention, things that have come to exist as a result of evolution have not been attributed functions by intention. The evolutionary surrogate for functions given by intention of a creator, is that which a characteristic of an organism has been chosen for in evolution, i.e., that which it contributes to a being's capacities for survival and the reproduction of its genes.

In the light of the previous chapter it is reasonable to assume that, on some level, we will find both of these basic types of functions among the emotions. Our basic emotional capacities are certainly natural, and we have a good prima facie reason to believe that the most natural emotions have Darwinian functions, but are not merely produced by chance evolutionary drifts. Insignificant features, such as the color of eyes or of skin, may have come about by chance. They do not affect behavior or survival capacities to any significant extent, and they are therefore not selected against in evolution. The susceptibility to emotions, on the other hand, has such a significant influence on actions and behavior, that if it did not in fact somehow contribute to survival and the reproduction of genes, and thus have a Darwinian function, it would most likely never have risen and survived natural selection.

---

80 I will not consider Aristotelian, natural teleology, nor functions given by a divine creator.
It is however possible that emotions were chosen for and served a function at an earlier stage in evolution, whereas now all or some emotions may have become obsolete. I.e. that they no longer serve the purpose for which they were chosen, but have not yet been eliminated. This would put them on par with the appendix; useless and a potential nuisance, but still there. This possibility raises some interesting questions, particularly with regard to emotions in humans. It is hard to see what might have superseded the emotions in animals, but humans are a different story; they have reason. It is a plausible assumption that emotions preceded reason in human evolution. Could it be right, as countless philosophers have claimed, that reason provides guidance superior to that provided by emotions, and that the emotions are therefore to be suppressed by reason? Are emotions in humans obsolete? Among the things I will argue in what follows is that they are not. Emotions still serve functions in humans; they do things reason is not capable of. And furthermore, they do not only have Darwinian functions, but also cultural functions.

Nature seems not only to have endowed us with a general susceptibility to emotions, but also with separate types of emotions. One reason to believe so is that specific emotion types can be recognized in both very young children and in animals. Another reason is that a number of emotion types seem to be universal and to have stereotypical, spontaneous expressions, which can be identified across very different cultures. Investigators do not agree on the number of separate cross-cultural emotion types, but they do seem to agree on several categories. They all found a happiness category, as well as categories for anger, surprise, sadness and fear. All but one put disgust and contempt in one category and there is
near agreement on a category for interest or attentiveness. If these are truly natural types of emotions, it is reasonable to assume that they each serve a specific function, or did so at an earlier stage in evolution.

But what could the functions of emotions be? The problem is that the emotions often do not seem to be up to much good. Countless instances of emotions are clearly destructive, such as the desire for revenge when one knows that one can expect retaliation. Other emotions are altruistic, they motivate us to make sacrifices for others, with no apparent personal gain. This seems to go against the Darwinian principle that the traits that get chosen for in natural selection are those that contribute to survival and to the reproduction of genes, whereas destructive traits get eliminated. The obvious question therefore must be, what emotions are chosen for in evolution. What do they contribute to make up for their apparently harmful effects?

In chapter IV it was argued that emotions could be placed on a scale between, on the one hand, natural, instinctive, and non-cognitive emotions, and on the other hand, cultural emotions, identifiable only in terms of beliefs held either at the time the emotions occur or at some earlier time. I submit that the types of contributions specific emotions make correspond closely to their location on this scale. Natural emotions contribute to the reproduction of the genes, by, for instance, inciting an animal to protect and care for themselves and their offspring; cultural emotions guard or promote what we, by acculturation or by our power of reflection, find valuable, and they can, and frequently do, go against our gene-promoting instincts. But I do not intend to focus on the functions of

---

81 Ekman, 1972, pp. 60-61.
specific emotions. In what follows I will argue that there is an account of functions (abstract, but yet informative), widely applicable to emotions along this continuum, both to natural and cultural emotions. But once again I stress that I do not intend the account to be exceptionless or to describe a defining feature of all emotions.

V. 2. Emotions and Instrumental Rationality.

When we look at human emotions, as they appear in real life, as well as in literature, one of their most conspicuous and notorious characteristics is their apparent irrationality. They keep us from thinking straight, they make us judge and act irrationally at times, and they ignore and hinder the processes of cool, calculative reasoning. I believe this gives a good hint in the search for functions of emotions. It is tempting to assume that this is in some way the function of at least some emotions, i.e. that they are "meant" to override or supplement rationality and to control us in ways that rationality is, for some reason, ill equipped to do.

It is worth noting that I am here concerned with rationality in a different sense from the one cognitivists used as a basis for the claim that emotions must involve beliefs. Then I was concerned with the way in which it does make sense to cry over spilt milk, which closely resembles epistemic warrant. Now, on the contrary, I am concerned with instrumental rationality; rationality which finds means to given ends. That is the way in which it does not make sense to cry over spilt milk. It does not do any good; it is instrumentally ineffective.
Given what I have argued in earlier chapters, this approach seems to face an immediate objection. I argued against cognitivist essentialism, and claimed that creatures incapable of beliefs and reasoning can have emotions. But it might seem then, that the conclusions this approach to the search for functions of emotions, will yield, will be relevant only to emotions in humans, because only humans are capable of reasoning. This does not count as rebuttal of my approach per se, but if this were right, the application of my results would be severely limited. But as I said, I am still working within a family resemblance model of emotions and am ready to accept that the account I come up with will not be applicable to all emotions. One virtue of approaching the issue from this angle is that, if I succeed, I will have given a reason to believe that emotions in humans are not obsolete, as I earlier suggested might be the case. It will show that the emotions have not been superseded by the evolution of reason. Besides, I think it is not unreasonable to hope that an account of the functions of one group of emotions may have a wider application. It is possible that a description of the way in which emotions supplement reason can also be applicable to emotions of animals incapable of reasoning. When it is applied to animals, it will of course not involve a description of faulty reasoning, but supplementing reason may very well be an instantiation of a more general function of emotions.

Two questions arise here: What are the faults of reason? and how do the emotions supplement the faults? I will discuss three theories which attempt to answer these questions. Discussing the first two theories, Ronald de Sousa’s and Nico Frijda’s, and their shortcomings is primarily meant to give added plausibility to the third theory. Their failures
are easily avoided by the third theory, Robert Frank’s theory of emotions as commitment devices.

V. 2. 1. Ronald de Sousa: Emotions as Patterns of Salience.

Ronald de Sousa suggests answers to both questions. De Sousa asks what it would be like to lack the capacity for emotions, and suggests two possible answers: Such a creature would be either an angel or a machine. I am here only interested in the angel. By “angel” de Sousa means a being that is purely rational, some sort of a “Kantian Rational Will”. De Sousa formulates the rationality he has in mind rather vaguely, but the essence of the formulation is that an angel’s goals follow from the principles of reason alone and it has a perfectly determinate system for fixing on the means towards the goals. For an angel nothing can be random, and nothing can be determined merely by crass material contingency.\(^\text{82}\) All the actions and beliefs of an angel are determined by the principles of reason.\(^\text{83}\)

But the angel is, in fact, a myth, argues de Sousa, because there is no such thing as a fully determinate rationality, whether cognitive or strategic. He points out several problems with pure reason. I will not concern myself with the question how real the problems are, and describe them only very briefly, because they all boil down to one simple point. One problem is that logic does not always give a unique prescription for the choice of beliefs.


\(^{83}\) It is questionable whether this is really what it is like to be a rational being without emotions. But that is irrelevant to my discussion here.
"Don’t believe an inconsistent set," but "Believe the consequences of your beliefs" are not always compatible in their application, de Sousa claims, nor do they prescribe their own ordering. Another problem concerns inductive reasoning. It is that reason cannot tell us how probable it must be that a hypothesis is true, on evidence, and how improbable that it is false, before it is rational to accept it. The third problem is that reason or logic does not determine salience, i.e. what things to notice, pay attention to or to inquire about. The fourth problem on the cognitive level, and the one de Sousa pays most attention to, is what he calls the philosophers’ frame problem. We need extensive knowledge about “how the world works” to interpret simple instructions and to disambiguate simple sentences. We have a vast amount of such knowledge, not only semantic but also encyclopedic. We do for instance have the general knowledge required to know that “snow-shoes, alligator-shoes, and horse-shoes and not respectively made of snow, worn by alligators, or used to walk on horses.” Some of the knowledge needed can hardly be placed under a topic in an encyclopedia. The knowledge required to disambiguate the sentence “I left my raincoat in the bathtub, because it was still wet,” can hardly be labeled under a topic. “Assume all the powers already listed-logic, induction, and more than encyclopedic knowledge: the philosophers frame problem, roughly, is how to make use of just what we need from this vast store, and how not retrieve what we don’t need.” Reason is also insufficient on the strategic level. We are sometimes faced with equally good choices of courses of action, i.e.

---

84 De Sousa, 1987, p. 192.
choices which reason cannot decide between. Making a decision, therefore, requires additional principles and an angel, whose every decision must follow from the principles of reason alone, is stuck in the position of Buridan's ass, who died from having to make a choice between two equally distant carrots.

De Sousa poses the angel's problems as five separate problems. But they have a common core, namely that reason is, in each case, an insufficient or an inconclusive guide to making choices, whether between belief candidates, between things to attend to, knowledge to make use of, or courses of action to take. And he suggests that emotions are one source of the necessary supplemental principles that can help making these choices.

Given that the problems are real, how do the emotions supplement reason? To answer that de Sousa looks briefly at conditions where the philosophers' frame problem does not arise. Clearly it does not arise unless there is a body of knowledge to choose from and use selectively, i.e. it arises "only when we consider what to do with information interpreted and stored in an intentional system." For a completely mechanistic creature, such as an ant, the problem does not arise. The motor system of an ant, de Sousa claims, is controlled by a limited range of information to which its sensors are attuned. They are sensitive only to particular aspects of the world, and anything that might be called a belief is either a triggering cue for a specific response, or a condition that fixes a limited range of possible

---

86 Some rational choice theorists argue that throwing dice would be the rational thing to do in such a situation. But de Sousa's stipulated angel does not have that choice. He would be leaving his choice to material contingency, which he is not allowed.

responses. Humans are to some extent attuned to respond mechanically to particular aspects of the world. But in situations which are not dealt with in this way, the philosophers’ frame problem might arise, and the role of emotions is to supply reason in such situations by imitating, what de Sousa calls the encapsulation of perceptual modes, by which I take him to mean the attunement of the senses to a limited range of information. For a limited time an emotion limits the range of information an organism will take into account and the set of options to choose from. A terrified man focuses on the danger, and ignores other possible action than fleeing. De Sousa states his (biological) hypothesis thus:

The function of emotions is to fill gaps left by (mere wanting plus) “pure reason” in the determination of action and belief, by mimicking the encapsulation of perception: it is one of Nature’s ways of dealing with the philosophers’ frame problem.\(^89\)

Now if the inconclusiveness is the only problems with pure reason, it is hard to see why we need emotions to solve them. Every other moment we deal quite efficiently and unemotionally with all these problems. We continuously upgrade our belief system without consulting hard logic; we seem to apply standards of soft- or hard-headedness with little worry and quite unemotionally; we constantly notice, pay attention and inquire about things

---

\(^{88}\) De Sousa, 1987, p. 15 and p. 195. The aspects of the world, to which their sensors are attuned, are presumably ones that are common in their natural environment, and to which it is important for their survival, that they react appropriately. Some characteristics of dangers, food and potential mates, would probably be among them.  

\(^{89}\) De Sousa, 1987, p. 195. De Sousa has another way of putting his hypothesis, which he seems to think of as equivalent to the first one. He says: “Emotions are species of determinate patterns of salience among objects of attention, lines of inquiry, and inferential strategies.” (p. 196)
with little or no emotion; the philosophers' frame problem is also a minor obstacle, we
disambiguate the bathtub example coolly, and the same goes for Ziff's duck example.90
Problems on the strategic level seem soluble with the same ease. We often choose among
apparently equally good options is a very calm way. If reason does not determine which of
two roads we should take to get from place A to place B, we simply take one or the other.
All these things could indeed be done in an excited, emotional way, and occasionally are,
but if the problems can be so easily and coolly dealt with, isn't emotion quite superfluous?
Emotions, even calm ones and even if they are (sometimes) patterns of salience, seem
excessive as supplements for these particular gaps in pure reason.

I believe therefore that, even if emotions are capable of solving the problems de Sousa
brings up (just like a walnut can be cracked with a sledgehammer), we need something more
to account for their functions. The account should give some rationale for their force, and
explain why they, not only help us direct our attention and choose a course of action, but
often do so with a painful and irresistible intensity.

My suggestion is that this will be accounted for by showing that the function of emotions
is not merely to fill gaps in reason, but to actually go against it, to override the results of
calm deliberation (even when it comes to a determinate conclusion), if not to hinder
deliberation altogether. To justify this as a function of emotions I need to show pure reason
to be counterproductive, or at least less than optimally productive, as opposed to being

90 "I saw her duck . . . when they were throwing rotten eggs . . . and then I saw it
swim out to the middle of the lake." (De Sousa, footnote, p. 192)
merely insufficient or indeterminate, and that the emotions can procure some tangible gains. One such account is suggested by Nico Frijda in “Lex Talionis: On Vengeance”.

V. 2. 2. Frijda on Vengeance.

Frijda argues that, in spite of the apparent irrationality and destructiveness of vengeance, it brings some real gains. Desire for revenge is “as ‘rational’ as any other emotion, and functional in the same way as are these emotions.”91 He does not tell in what way other emotions are functional and rational, so I shall look at what he has to say about vengeance.

The problem with desire for revenge is clear from the definition Frijda offers. He defines it as the desire to,

harm someone else, or some social group in response to feeling that oneself has been harmed by that person or group, whereby the act of harming that person or group is not designed to repair the harm, to stop it from occurring or continuing in the immediate confrontation, or to produce material gain.92

The immediate purpose of revenge is thus only to inflict suffering on the perceived enemy, and it is not meant to repair the harm or to produce material gain. Further striking features of vengeance, although not a part of the definition, are, the frequently disproportionate cruelty of revenge (the Law of Talion, “eye for an eye, tooth for a tooth”, is meant to restrain the vengeful person, to keep him from taking more than an eye for an eye), the incredible persistency of vengeance over a long period of time, and perhaps the most

striking of all, the self-destructive measures a vengeful man is willing to take to harm his enemy.

A classical model of the way emotions work attributes their driving force to the seeking of pleasure and the avoidance of pain. According to this model one seeks the allegedly pleasurable emotions and avoids the painful ones. I believe Frijda is right when he claims that this is a false picture of the mechanism of emotions. For one thing, emotions can go unfelt without losing their motivational powers, and people may even seek allegedly painful emotions or regret not having them, because they believe them to be appropriate. Frijda sets out to give a new and more plausible account of the mechanism of vengeance, one that is better suited to "satisfy our current cognitive view of emotions," and according to which feelings do not add up or subtract, but respond to patterns of appraisals.

These patterns of appraisals are structured patterns. They result from the multiple concerns touched upon by an emotional event and its implications; they include the multiple ways in which an event can be emotionally relevant. It is not that different affects add up or subtract, but that different relevances summate, conflict, or keep each other in balance.

An offense can have a number of painful implications. One of them is that one is harmed by someone who goes unharmed and even enjoys the offense or gains from it. Generally, Frijda claims, pleasures and pains are measured, not in absolute terms, but by comparison with other people's pleasures and pains or one's own pleasures and pains at an earlier time. This, coupled with the knowledge of an offender possibly enjoying having

inflicted suffering makes the pain more acute or poignant. A second implication, noted by Frijda, is that the offense is a manifestation of power-inequality. The offender has the power to offend, “he is the actor, you are the object”\(^95\). The third implication is that being offended or used to further other people’s own interests damages one’s self-esteem or social prestige. It can attack an individual’s sense of personal value and identity because one is treated as an object. These implications of offense can, according to Frijda, all give rise to a vengeful impulse and the pain of each of them can be alleviated by a successful revenge. A successful revenge restores the balance of suffering, it equalizes power, restores self-esteem and it provides an escape from pain by removing the added sting of the offenders gain.

Frijda believes these gains help making sense of the desire for revenge, in the face of its apparent irrationality. But there is something lacking from this account. Revenge, on this account, provides satisfaction and relief from pains and desires which are a part of, or produced by the emotion itself. It may describe well the immediate emotional reward of a revenge and lay out the structure of the immediate, explicit motivation for revenge. But if this were all there was to the function and rationale of having emotions\(^96\), it would not be very different from saying that itching provides a tangible gain, namely the satisfaction we get from scratching. I agree with Robert Frank that,

the satisfying feeling someone gets from having done the right thing [or taken revenge] is, in a very real sense, its own reward. But our task here, . . . , is to explain

---

\(^95\) Frijda, 1994, p. 275.

\(^96\) Frijda does not claim that it is. He points out that vengeance also serves to stabilize society. It is a tit-for-tat strategy, which Axelrod showed to be a superior strategy in bringing a rival in line for cooperation. (see Frijda, 1994, p. 271 and Richard Dawkins, 1989, ch. 12)
how such sentiments might have evolved in the material world. We can’t eat moral sentiments. For them to be viable in competitive environments, they must have a material payoff.\footnote{Frank, 1988, p. 54.}

The payoff described so far is merely psychological, although very real as such. But since emotions frequently appear to be destructive, we need to show how they make up for their harmful effects, and how they can survive and be chosen for in natural selection. For that, a psychological payoff or the satisfaction of desires internal to the emotions themselves, is not sufficient. To explain how emotions can provide a payoff in a harder currency, the currency needed to make them viable in a competitive Darwinian environment, I will use Frank’s theory of emotions as commitment devices.

V. 2. 3. Emotions as Commitment Devices.

Robert Frank describes emotions as commitments to strategies. The problem they solve is the commitment problem. The commitment problem will best be introduced by means of an example. A kidnapper suddenly gets cold feet. He wants to set his captive free, but is afraid he will go to the police. The captive, on the other hand, will, in return for his freedom, promise not to do so. The problem is that as soon as the victim is free it will no longer serve his interests to keep his promise. From a purely self interested point of view he will have no reason not to turn the kidnapper in to the police. Therefore the kidnapper concludes that he must kill his captive. That is a conclusion neither of them likes.
There would be a way out of this dilemma if they could find some method to commit the victim to keeping his promise. "If the victim has committed an act whose disclosure could lead to blackmail, he may confess it; if not, he might commit one in the presence of his captor, to create a bond that will ensure his silence." The problem is a commitment problem, and the blackmaileable act is a commitment device, which can solve the problem and ensure a more desirable outcome for both.

The kidnapper example exemplifies a central feature of the commitment problem: it describes a problem whose solution requires people to make *ex ante* commitments to behave in ways that will not serve their interests *ex post*. The classic example of problems of this kind is the so called prisoners dilemma. Each of two prisoners is given the following choice: To rat on the other and if, a) the other rats as well, get three years in prison, or if, b) the other keeps mum, get one year in prison; or to keep mum and if, a) the other keeps mum as well, get two years in prison, or if, b) the other rats, get four years in prison. The dilemma is created by the fact that each of the prisoners is better off ratting than keeping mum, regardless of what the other does, and it is therefore rational for both to rat, but they would both be better off if both kept mum than if both ratted. A commitment device of the sort described in the kidnapper example would help solve the problem.

The prisoners dilemma and the commitment problem are not merely abstract theoretical devices, problems with the same general features arise in countless contexts in daily life. Consider cooperative ventures. Jones, a talented cook, and Smith, a shrewd businessman, could both profit from cooperating in running a restaurant. But once the venture is off the

---

98 Frank, 1988, p. 4.
ground each of them will have a chance to cheat and thus get more than his fair share, at the others expense. But if both cheat, the business will go bankrupt and both will be worse off than they would be if they were both honest. If they knew each other to be purely calculative they would not trust each other, and the business would never get started. They would therefore enjoy the benefits of the cooperation only if they could commit to not cheating.

A similar story could be told about bargaining situations. By cooperating, Smith and Jones can make a thousand dollars. Smith badly needs money, whereas Jones doesn’t. Jones has an advantage. He can credibly threaten to walk away from the deal unless he gets eight hundred dollars of the profit. It would be in Smith’s advantage here to be committed to not accepting an unfair bargain, i.e. in this case, to walk away from a much needed two hundred dollar profit.

And about deterrence. Jones and Smith have adjacent plots of land. Jones is liable for damage his steers do to Smith’s wheat. Jones could prevent damage by spending 200 dollars on a fence. If he does not do that his steer will cause 1000 dollars damage. But it will cost Smith 2000 dollars to take Jones to court and get the damage paid. Therefore, if Smith is not committed to seek justice, even at excessive cost, Jones will have no good reason to prevent the damage.

It may seem excessive to give all these examples, but I want to make it clear that the commitment problem is not merely a contrived theoretical problem. It permeates every aspect of our lives. The common element in all these examples is that they describe
situations where it is “in a person’s interest to make a binding commitment to behave in a way that will later seem contrary to self-interest.” There are a number of ways in which the commitments can be made. One way is simply to eliminate beforehand, options that in the future might be tempting. Odysseus had himself tied to the mast of his ship so that he would not be lured by the Sirens. Altering the payoffs or the material incentives, and thus making particular future options less feasible, is another way. The solution in the kidnapper example is an example of that, whereas formal contracts specifying penalties in case of a breach are probably more common. These are what might be called external commitment devices. There are various problems connected with them, with actually making the commitments, with making it known that one is committed, and with enforcing the commitments. They can be inefficient and costly. We might very well be better off with internal commitment devices, i.e. something that makes us actually want to do things that are against our immediate interests, although we have the choices open to us and the material payoff matrix intact. What this amounts to is that we might be better off being partly irrational. That is where the emotions come into play. Frank suggests that emotions (and feelings, such as hunger) are a part of an internal, psychological reward mechanism, which make us want to do things which are sometimes against our immediate material interests.

If we look briefly at the examples of the commitment problems we will see that particular types of emotions would not only be suitable to solve these problems, but that they very frequently do. If Jones the cook and Smith the businessman, in the cooperative

99 Frank, 1988, p. 47. All the examples I give are from Frank, pp. 48-50.
venture, were both susceptible to strong feelings of guilt they would be much less tempted to cheat. They would be disposed to honesty. If Smith, in the bargaining situation, has a strong sense of justice, he will not only be concerned about how much, in absolute terms, he gets, but also that he gets a fair share of the total. He would be less inclined to accept an unfair bargain. If Smith, the wheat farmer in the deterrence situation, is susceptible to anger when he is offended he will be inclined to take Jones to court at any cost. That would deter Jones from letting his steer walk loose. It seems therefore that in all these examples it is profitable in material terms to be subject to emotional responses which override other considerations. On the account of the functions of emotions which stems from this, the emotions are meant to compete with and override calculative rationality. That provides a rationale for the aspect that was not accounted for by de Sousa. By viewing the emotions as competing with rationality, as opposed to merely filling gaps, we can make sense of their often painful and irresistibly intense motivating force. This way of viewing emotions furthermore fills the gap left by an account of immediate, psychological gains by showing that they also provide material gains which can help explain how emotions might have evolved in a competitive Darwinian environment.

Frank’s theory also explains another aspect of emotions which de Sousa’s theory fails to account for, namely that a number of emotions are associated with typical, recognizable expressions, which are at least partly free from direct control. If commitments are to serve their purpose it is not enough to be disposed to behave in certain ways, our potential cooperators have to know that we are so disposed. If a noncheater is to benefit in material terms, others must be able to recognize him as a noncheater, and he must be able to
recognize other noncheaters. The same goes for the commitment to seeking revenge. If
Jones does not recognize Smith as someone who will seek revenge at considerable cost he
will not be deterred.

The function of (some) emotions that emerges from this discussion is to serve as
commitments to strategies, which might otherwise be overruled by calculative self-interested
rationality. Thus stated the theory seems only to apply to emotions in beings which are
capable of calculative reasoning, namely humans. But earlier I expressed the hope that an
account arrived at by looking for defects in reason, might have a wider application. My
suggestion is that emotions have the function to override, not only reason, but also other
kinds of control mechanisms. We can imagine an antelope obeying a control mechanism
which could be interpreted as commanding it to eat grass when it is hungry. Then a predator
approaches and a fear mechanism is triggered and commands the antelope to flee. If the two
mechanisms were equally strong the antelope might continue eating and consequently be
eaten. But the fear mechanism is more intense and gives a stronger motivation. Its function
is to override other control mechanisms. Its command could be interpreted as “Flee, even if
you are hungry and there is grass around!” The intensity of fear commits the antelope to the
strategy of fleeing, no matter what, when it is approached by predators. A similar story
could be told about anger. One mechanism may command a dog to rest when it is tired. But
when another dog trespasses on its territory the “rest!” command is overridden by an
“attack!” or “threaten!” command. An anger mechanism thus commits the dog to
threatening when resting might otherwise be a tempting option.
Once we notice this aspect of natural emotions we can also understand better why some types of beliefs come to be tied up with basic emotional capacities. I suggested in the last chapter that perhaps there is nothing special to be said about emotional beliefs, as a class of beliefs, except that they are typically about what is valuable and important (which does not really set them apart from some non-emotional beliefs). Overriding calculative reason is thus only one way in which emotions carry out their more general function of checking other control mechanisms in situations which call strongly for specific responses.
A SUMMARY OF CONCLUSIONS.

Theories of the nature of emotions have typically been concerned to list necessary, and sometimes sufficient conditions of emotions, i.e. to describe the alleged essence of emotions. Essentialist theories have been mostly of two types, affect theories and cognitive theories. The negative side of this thesis was an attack on essentialism. I first attacked the particular essentialist theories, and then suggested that perhaps looking for necessary conditions for emotions was a misguided approach to the issue. The positive side of the thesis was to suggest treating emotions as having a family resemblance rather than an essence. I furthermore suggested that we might improve our understanding of emotions by looking, not only at what they may be, but also at what they do, i.e. by looking at their functions.

The central claim of affect theories of emotions is that an affect is necessary for an emotion to occur. They involve different ideas of what constitutes an essence, most consider feelings essential to emotions, other stress physiological changes. The strictest of the affect theories claim that each type of emotion is identical with a particular type of feeling. I offered four basic arguments against affect theories. Firstly, emotions seem sometimes to be independent of affects. Secondly, affects are not intentional states, at least not of the type needed to make sense of the intentionality of emotions. Thirdly, mere affects cannot be assessed as being either rational or irrational as emotions often can. And finally, affects
cannot be individuated into types parallel to the types of emotions. The conclusion to be drawn from these arguments is not that emotions have nothing to do with feelings or other affects, but only that particular types of emotions can not be associated with particular types of affects and that there can be instances of emotions which are not accompanied by affects.

Cognitivism arose as a response to these problems. The central claim of cognitivism is that thoughts or beliefs are necessary for an emotion to occur. Some of the most influential arguments for cognitivism have taken as their main premise the same observations as were used to reject affect theories. It has been argued that because emotions often seem independent of feelings, because they are typically intentional, because they can be assessed as being either rational or irrational and because they are divisible into types with only subtle differences they must involve beliefs or thoughts. Against these arguments for cognitivism I argued first that intentionality, in the sense of being directed at an object, does not entail belief or thought. There can be other ways to account for object directedness. Secondly I argued that there are instances of emotions about which it makes no sense to ask whether they are rational or irrational. Not all emotions are subject to assessment of rationality. The rationality of emotions does therefore not provide grounds for a general cognitivist claim about the nature of emotions. Thirdly, I argued that at least some emotions can be recognized without reference to beliefs, namely by reference to typical, spontaneous expressions. I raised the further objection that emotions can conflict with explicitly held beliefs. Taking the emotion as evidence for a belief would in such cases be to beg the question for cognitivism, since no other argument has managed to establish the necessity of
belief for emotion. The conclusion I draw from all this is that cognitivism as a general claim about the nature of emotions has not been given sufficient support.

Neither of the two traditional essentialist theories seem to work. I do not see other plausible essentialist alternatives and therefore suggest that we stop looking for an essence of emotions. I argued that even if we were to find a characteristic common to all emotions we would not necessarily have found an essence. Such a characteristic could be incidental to emotions. I suggested that we think of emotions as having a family resemblance. They would thus have overlapping similarities and relations, but not be unified by some essential characteristic. "Emotion" would thus be a clusterconcept referring to a number of different things, (e.g. beliefs, feelings, physiological changes and motivations) none of which is necessary for the occurrence of emotions.

But thus far I have only been concerned with essentialism as a general claim about the nature of emotions. Thinking of emotions as having a family resemblance does not exclude the possibility of there being particular types of emotions which have an essence. I argued that that is indeed the case about emotions which do not have distinct spontaneous expressions. If an emotion type is strongly associated with such an expression it makes sense to talk of it as occurring without a belief. I called these emotions primary emotions. But in cases of emotions which do not have a distinct spontaneous expression it makes no sense to talk of them as occurring without some reference to beliefs. These emotions I called secondary. Secondary emotions are necessarily cognitive in the weak sense that their occurrence can only be accounted for in terms of actually held beliefs or residues from previously held beliefs.
I argued that the division between primary and secondary emotions is largely the same as that between natural and cultural emotions. Secondary emotions involve beliefs, beliefs require language and language can only be acquired through acculturation. But that does not necessarily imply that secondary emotions are culturally relative. Something about culture, language or even humans might somehow necessitate particular types of secondary emotions so that they can be found in one form or another in all cultures.

I also argued that between natural and cultural emotions there is not a wide gap, but a continuum, which I divided roughly into four parts. On one end there are emotions which are purely instinctive reactions to particular aspects of the environment. Next there are emotions which are learned but do not require conceptual understanding. The third are emotions which are versions of natural emotions but require conceptual understanding of circumstances. Finally there are emotions which do not correspond to any natural emotions, but involve purely cultural beliefs and values which have come to be tied to basic emotional capacities.

This approach to emotion theory may seem to tie emotions together rather loosely as a class. Perhaps that is so, but I believe that the fabric of overlapping similarities and relations can be strengthened considerably by including in it not only intrinsic characteristics of emotion but also an account of what they do, i.e. of their functions. That was my task in the final chapter of the thesis.

The function of a thing is that which it is meant to do, the purpose it is supposed to serve. The paradigm example of functions is function given to a manmade object by its
creator. The Darwinian surrogate for such functions is that which a characteristic of an organism is chosen for in evolution. Now, one of the most conspicuous aspects of emotions is their apparent destructiveness. The challenge in finding the functions of emotions is therefore to explain how natural emotions came to be chosen for in evolution in spite of their destructiveness.

I approach the problem through the observation that emotions often seem to incite one to apparently irrational beliefs and actions, i.e. they override calculative rationality. I ask whether that could somehow be their function. For that to be so the rationality that emotions compete against needs to be shown to be inadequate in ways that emotions can supplement or do better.

I first explored Ronald de Sousa’s theory that emotions are meant to fill gaps left by pure reason, i.e. where pure reason is inconclusive about what to believe, pay attention to, or what course of action to take. I argued that the function de Sousa attributes to emotions could and typically were filled with much subtler means than emotions. His theory does therefore not make sense of the intensity of emotions. To make sense of that we need an account according to which emotions are not merely meant to fill in for reason where reason is inconclusive, but to actually fight against and override reason in particular circumstances. For that we need to show emotions to procure tangible gains which reason can not achieve.

Nico Frijda tries to give an account of how emotions can bring real gains which could not be achieved by reason. His example is of the desire for revenge. Being offended has, according to Frijda, various painful implications, and the pain can be alleviated by a
revenge. Revenge gives real, tangible satisfaction. But the satisfaction is only satisfaction of desires which are internal to the emotion and I argued that although there are real rewards internal to emotions it does not show what is gained by having emotions at all. The satisfaction following a revenge or an altruistic act is real, but does not explain what is gained by having emotions which dispose one to (potentially self-destructive) revenge or to altruistic acts.

To fill that gap I used Robert Frank’s theory of emotions as commitment devices and solutions to what he calls commitment problems. A commitment problem is a problem whose solution requires that one make a prior commitment to behave in ways which later will not seem rational. If one were known to be purely calculative one’s promises and threats would not be taken seriously (given that at a later time it would not be in one’s interest to keep a promise or carry out a threat). That means one would not enjoy the benefits of being able to deter or cooperate with people. Being known not to be purely calculative, but rather emotionally disposed to honesty and anger, can solve the problem and enable deterrence and cooperation. This shows how emotions can bring tangible gains which are not merely internal to the emotions themselves.

According to this emotions seem to have the function to suppress or override calculative reason. But then it seems to apply only to human emotions (given that only humans have calculative reason that needs to be suppressed). To give it a wider application I finally suggested that this was only an instantiation of a more general function, namely to override other control mechanisms and motivations in circumstances which strongly require a particular response. Fear thus overrides an animal’s desire to eat. This can make sense of
both the intensity of emotions and explain how emotions came to be chosen for in a competitive Darwinian environment.

Although I will not indulge here in speculation about the implication of my views it is clear that theory about the nature of emotions can have very wide ramifications. It matters to questions in ethical theory as well as about psychotherapy and various other issues. In this thesis I have been mostly concerned with the question whether or not emotions have essences. I rejected the transition from one form of essentialism to another, i.e. from affect theories to cognitivism. But we may still ask whether the rise of cognitivism has led to progress in emotion theory. I believe it has. I argued that although essentialism is untenable as a general theory of emotions there is a significant subclass of emotions which can only be understood in (weak) cognitive essentialist terms. Much of the work that has been done under the banner of cognitivism applies to emotions of this class. Much has been done to analyze the cognitive content of various emotions, and also the cognitive differences between emotions of this class. Shame and guilt and other (broadly) moral emotions are being carefully scrutinized, sometimes in the context of psychology and psychotherapy, but often in relation to various question about moral theory as well as meta-ethics, and there is little doubt that a shift towards cognitivism has improved understanding in these areas. But I believe that keeping the natural, non-cognitive basis of emotions and their functions in clear focus can only improve understanding even further.
BIBLIOGRAPHY


Kristján Kristjánsson, “Um Gedshaeringar,” Unpublished manuscript.


