

A Feeling Theory of Feelings

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in  
the Graduate School of the Ohio State University

By

Jeremy Weiss, M.A.

Graduate Program in Philosophy

The Ohio State University

2016

Dissertation Committee:

Sigrún Svavarsdóttir, Advisor

Justin D'Arms, Co-Advisor

Declan Smithies

Copyright by

Jeremy Weiss

2016

## Abstract

My dissertation is an investigation of positive and negative feelings, which I call ‘affective experiences.’ More specifically, this dissertation is an attempt to answer the following questions: *what is it* to feel positively or negatively about something? *In what sense* are positive and negative feelings positive or negative? After a bit of background and some previewing of what is to come in Chapter 1, in Chapter 2 I explain and begin my defense of an answer to these questions that I call a ‘phenomenal attitude view.’ Such a view claims, roughly, that what makes positive feelings positive is a similarity in how the attitudes in which they consist feel, and what makes negative feelings negative is a similarity in how the attitudes in which *they* consist feel.

Then, in Chapters 3 and 4, I consider the most plausible alternative answers to the questions above. In Chapter 3 I assess ‘evaluative content views,’ which have it, roughly, that what makes positive feelings positive is that they represent things as good in some way, and what makes negative feelings negative is that they represent things as bad in some way. In Chapter 4 I assess ‘motivational attitude views,’ which have it, roughly, that what makes positive feelings positive is that they dispose us to act positively towards what they are about and what makes negative feelings negative is that they dispose us to act negatively towards what they are about. I argue that neither alternative works: they are each either extensionally inadequate or they give the wrong *kind* of accounts of the sense in which affective

experiences are positive or negative orientations towards what they are about (or both).

Further, I try to show that the best arguments *for* these alternatives are not particularly good.

In Chapter 5, I argue that even if these alternatives join forces, so to speak, they are still inadequate accounts of affective experiences. I argue for this claim by showing that feeling is necessary for love, and that motivation and evaluation (even together) are not sufficient. But not only do we learn here that evaluative content views and motivational attitude views are implausible: we also learn that affective experiences are very important (since they are necessary for love and love is very important), and that can help to justify paying closer attention to feelings as such (instead of bulkier things like emotions, desires, etc., perhaps).

Finally, in Section 6, I conclude by pointing out some ways in which the discussions and arguments of this dissertation bear on a number of philosophical debates, and I make some tentative suggestions about promising areas for future research on these topics. The discussions and arguments of this dissertation promise, in some small way, to further our understanding of the nature of consciousness, intentionality, desire, moral judgment, practical reason, and lives worth living.

## **Dedication**

Dedicated to my brother

## Acknowledgments

There are a great number of people who deserve limitless gratitude for all they have done for me and for the world over the years it took me to write this dissertation: my family and friends, my cat, the OSU Philosophy Department and its wonderful students, faculty, and staff, and many others. But I would like to single out the members of my committee for special praise and thanks here. Justin D'Arms, Declan Smithies, and *especially* my advisor, Sigrún Svavarsdóttir: thank you all for your brilliance and inspiration, and for your aid and your patience. I am still constantly in awe of what each of you do, and I could not imagine having had better people to work with on this project.

## **Vita**

2000 ..... Oak Park and River Forest High School

2004 ..... B.A. Political Science, University of  
Southern California

2008 ..... M.A. Philosophy, University of Wyoming

2008 to present ..... Graduate Teaching Associate, Department of  
Philosophy, The Ohio State University

## **Field of Study**

Major Field: Philosophy

## Table of Contents

Abstract.....	ii
Dedication .....	iv
Acknowledgments.....	v
Vita .....	vi
Table of Contents.....	vii
Chapter 1: Introduction.....	1
Section 1: The Basic Concern .....	1
Section 2: Valence .....	3
Section 3: Phenomenology and Intentionality .....	12
Section 4: Coming Attractions .....	17
Chapter 2: Affective Experience and Phenomenal Valence .....	21
Section 1: Introduction.....	21
Section 2: Phenomenal Valence .....	22
Section 2.1: P and N .....	22
Section 2.2: The Heterogeneity Objection .....	26
Section 3: Affective Experiences are Intentional States .....	33
Section 3.1: Overview.....	33
Section 3.2: Initial Reasons to Think that Affective Experiences Are Intentional States .....	33
Section 3.3: Objections .....	37
Section 4: The Case for Impure Intentionalism .....	46
Section 4.1: Introduction, a Complication, and a Gift .....	46
Section 4.2: The Case for Pure Intentionalism .....	49
Section 4.3: Objections to Pure Intentionalism.....	50



Section 4.4: The Virtues of Impure Intentionalism .....	56
Section 5: Conclusion .....	63
Chapter 3: Affect and Evaluation .....	65
Section 1: Introduction.....	65
Section 2: Preliminary Challenges to Evaluative Content Views .....	69
Section 3: Correctness .....	75
Section 4: Enrichment.....	90
Section 5: Epistemic Justification .....	97
Section 6: Conclusion .....	108
Chapter 4: Affect and Motivation.....	110
Section 1: Introduction.....	110
Section 2: Motivational Attitude Views .....	113
Section 2.1: Positive and Negative Motivational Profiles .....	113
Section 2.2: The Attractions .....	118
Section 3: Failing the First Criterion .....	124
Section 3.1: Wishes and Hopes.....	125
Section 3.2: The Past .....	127
Section 3.3: Action-Thwarting Contents.....	128
Section 3.4: The Weather Watchers .....	130
Section 3.5: Feeling Motivated vs. Being Motivated.....	138
Section 4: Failing the Third Criterion .....	139
Section 4.1: A Brave New World.....	139
Section 4.2: Representing Motivation .....	140
Section 4.3: Mad Affect.....	144
Section 5: Conclusion .....	153
Chapter 5: Affect, Love, and Importance.....	156
Section 1: Introduction.....	156
Section 1.1: Overview.....	156
Section 1.2: Love Requires Affective Dispositions.....	159
Section 1.3: Friends and Foes .....	160

Section 2: From the Beloved's Perspective .....	163
Section 2.1: Fred and Wilma .....	163
Section 2.2: Partiality and Strangers .....	167
Section 2.3: Objections .....	170
Section 3: To Love and to Feel .....	173
Section 3.1: Affect and the Lover .....	173
Section 3.2: Objections Again .....	176
Section 4: Conclusion .....	180
Chapter 6: Conclusion .....	182
References .....	191

## Chapter 1: Introduction

### Section 1: The Basic Concern

Sometimes we “feel positively” or “feel negatively” about something or other. For example, one might feel positively about spending time with friends or family, or about finishing an assignment, or about the dish one is currently eating. And one might “feel negatively” about that driver’s behavior, or about that movie’s ending, or about the callousness of certain social policies. The questions that motivate this dissertation are the following. *What is it* to feel positively about something? *What is it* to feel negatively about something? *In what sense* are positive (or negative) feelings *positive* (or *negative*) stances or attitudes or orientations towards what they are about?

As we will see, these questions are good ones. First, they have not received adequate attention from philosophers (or at least not from recent philosophers in the analytic tradition). There has been plenty of work on potentially related phenomena, e.g. on emotions, evaluative judgments, desires, and pleasures and pains, but it is at best unclear how such work bears on our questions above. What *are* the relations between positive and negative feelings, on the one hand, and emotions, evaluative judgments, desires, or pleasures and pains on the other?

Second, insofar as the existing work of philosophers can – with a bit of creativity and charitable license – be used to answer the questions above, the answers suggested are often bad ones. Let us focus on the following question from above: what is it to feel *positively*

about something? One answer, suggested by some influential philosophical work on emotion and desire, has it, roughly, that to feel positively about something is to represent that thing as *good* in some way.<sup>1</sup> We should reject this answer, as we will see in Chapter 3 below. A different answer, suggested by some influential philosophical work on desire and pleasure (and emotion and evaluative judgment, too, to a lesser extent), has it, roughly, that to feel positively about something is to be disposed to *act* in positive ways towards it.<sup>2</sup> We should reject this answer, too, as we will see in Chapter 4 below.

I will defend a different approach. The view I favor has it that to feel positively about something is to have an *attitude* towards it that *feels* a certain distinctive way, and one can have such an attitude without representing the object of the attitude as good in any way and without one's being positively motivated with respect to the object of the attitude.<sup>3</sup> So *what it is* to feel positively about something is to have such an attitude, and it is *not* to represent value or to be motivated in any particular way. I will explain this idea further in Chapter 2 below. Its defense will begin in Chapter 2, too, and will last the rest of the dissertation.

The results of these discussions will be interesting in their own right, at least insofar as they force certain choices on us with respect to our theories of the mind. But they are also important because positive and negative feelings are important. They seem to be no less

---

<sup>1</sup> The views I have in mind here include e.g. “judgmentalist” theories of emotions which have it that emotions are, at least in part, evaluative beliefs, and “guise of the good” theories of desires which have it that desires are, at least in part, representations of value. For further discussion of such theories, see Chapter 3 below.

<sup>2</sup> The views I have in mind here include e.g. what Timothy Schroeder (2004) calls the “standard theory” of desire according to which, roughly, to desire that p is to be disposed to do what you believe will bring it about that p, and “motivational theories” of pleasure according to which, *very* roughly, an experience is a pleasure in virtue of the fact that the subject of the experience is disposed to do what she believes will lead her to have experiences of that kind. For further discussion of such theories, see Chapter 4 below.

<sup>3</sup> This point about the relations between affect and motivation will need to be complicated a bit, and it will be in Chapter 4 below.

than a necessary constituent of a life worth living. Imagine going a *day* – let alone a *lifetime* – in which you did not feel positively or negatively about anything! So a better understanding of such feelings will enable us to achieve a better understanding of worthwhile lives. At any rate, in Chapter 5, I will argue that positive and negative feelings are at least necessary for love and that we could not get all that we want out of loving or being loved from motivation or evaluation (individually or in combination). So a better understanding of affect will at least enable us to achieve a better understanding of love.

But the bulk of this dissertation will consist in the attempts in Chapters 2 through 4 to figure out just what in the world positive and negative feelings *are* and *in what sense* they are positive or negative. In the rest of this introductory chapter I will lay out the very general features of positive and negative feelings that will structure the discussions of later chapters and I will preview coming attractions in a bit more detail.

## **Section 2: Valence**

There seem to be a variety of ways in which someone can be “for” or “against” something (in some respect, to some extent). For example, you might think that reading philosophy is a *good* way to spend one’s time; your friend might be *motivated to play* soccer on the weekends; I might *enjoy* drinking a pint of beer on a hot day. In one way or another, you seem to be *for* reading philosophy, your friend seems to be *for* playing soccer, and I seem to be *for* drinking beer. On the other hand, you might think that taking another card in blackjack when your existing cards add up to seventeen is *irrational*; your friend might be *motivated to avoid* very loud concerts; I might *feel ashamed* of having put off my work until the last minute. In one way or another, you seem to be *against* taking another card, your friend seems to be *against* attending such concerts, and I seem to be *against* having put off my work.

Alternatively, you might *see* a coffee cup on the table; your friend might *believe* that Napoleon lost the Battle of Waterloo; I might *imagine* what it would be like to take up knitting. But it seems as though you might be entirely indifferent to the coffee cup being on the table, your friend might be entirely indifferent to Napoleon's defeat, and I might be entirely indifferent to taking up knitting. That is, we might not be *for* (or *against*) these things in any way, even though we are mentally related to them.

In this dissertation I will be directly interested in the former kinds of mental states<sup>4</sup> – e.g. evaluations, motivations, and positive or negative feelings – and only indirectly interested in these latter kinds of mental states – e.g. (non-evaluative) perceptions, beliefs, and imaginings. To fix terminology, let us say that those former kinds of mental states – those that amount to a subject's being for or against something (in some way, to some extent) – are “valenced” mental states. Further, let us say that mental states that amount to a subject's being *for* something are “positively valenced” mental states (or are “pro-attitudes”), and mental states that amount to a subject's being *against* something are “negatively valenced” mental states (or are “con-attitudes.”) If there are mental states that amount to a subject's being both for and against something, then we can call those “ambivalent” mental states. Those states that do not amount to a subject's being for or against something (in any way, to any extent) – e.g. (non-evaluative) perceptions, beliefs, and imaginings – are “non-valenced” mental states.

---

<sup>4</sup> I intend the expression ‘mental states’ in a broad, minimally metaphysically committal way, e.g. so that attitudes, stances, processes, events, etc. might conceivably count as mental states.

Valence, then, is a property that mental states can have or lack.<sup>5</sup> As I am thinking of it, valence is a non-normative property (even if it is a normatively relevant property, as we will discuss momentarily and in future chapters). Or at least, the concept of valence is a non-normative concept. Someone who attributes a positively valenced state to a subject while denying that there is *anything* good or reasonable or obligatory about that subject's being in that state need not be contradicting herself (and ditto, *mutatis mutandis*, for negatively valenced states and disvalue). In this way, a state's being or not being valenced is more like a state's being or not being conscious, or intentional, or vivid, or arrived at on a Tuesday than it is like a state's being or not being wise, or admirable, or praiseworthy.<sup>6</sup> To attribute a valenced state to a subject is just to claim that that subject is not entirely, thoroughly indifferent to everything.<sup>7</sup> We can, alas, be positively oriented towards horrible things and we can be negatively oriented towards things of great value.

There is a serious initial stumbling block, however, in the way of thinking about ourselves as being “positively” or “negatively” oriented towards things. The objector I have in mind would claim that categorizing our feelings (or, probably, our motivational tendencies, or even our evaluative practices) in terms of whether they are “positive” or “negative” is unhelpful (to put it mildly). One might worry, that is, that categorizing these

---

<sup>5</sup> In one clear sense, it is *the person herself* who is pro or con something, favorably or unfavorably disposed towards it, for or against it. But the idea in the main text is that the person is for or against something *in virtue of* having certain valenced mental states.

<sup>6</sup> See Robert Solomon & Lori Stone's (2002) excellent discussions of many of the more normatively-loaded uses of “valence” in the history of philosophical thought about the emotions and for some of the less normative uses, too, in both philosophy and psychology.

<sup>7</sup> We can distinguish “positive indifference” from “negative indifference”. One is negatively indifferent towards something when she has no positively or negatively valenced states towards it. For example, I am negatively indifferent towards the fact that one faraway galaxy is a particular distance from another faraway galaxy. One is positively indifferent towards something when she has both positively and negatively valenced states towards it of roughly equal “strengths.” For example, at the time of writing I am positively indifferent towards my doing laundry today. The point in the main text then can be understood as saying the following: “to attribute a valenced state to a subject is just to claim that that subject is not entirely, thoroughly *negatively* indifferent to everything.”

features of ourselves as either “positively” or “negatively” valenced encourages a *drastically* over-simplified account of our emotional or desiderative or even evaluative lives: “you’re for it or against it” sounds dangerously similar to “you’re with us or against us”. One might worry that it is like “going to an opera or an art museum, and being told that your only critical response can be a “boo” or a “bravo!”.” (Solomon & Stone (2002: p. 430))

Anyone who has reached adolescence, let alone adulthood, should feel the force of these worries. Our emotional and desiderative and evaluative lives are always, or almost always, messy and ambiguous. Nevertheless, we can accommodate these reasonable points without giving up on valence. For example, consider a character who we can call Paul. Paul is, by all, or almost all, accounts, a great artist. His work is beautiful, sensitive, and insightful, and he is prolific. He demands a great deal from himself, from those around him, and from other artists, and he often meets – and inspires others to meet – these demands. He is bright, confident, funny, and engaged, and on certain days can be shockingly thoughtful to those close to him. Through gestures by turns subtle and grand, he can make the people in his life feel like they are ten feet tall, whole, and capable of anything. But those days are as much the exception as the rule. Oftentimes, Paul is a jerk, at least to those close to him. He can be mercurial, self-absorbed, and extraordinarily cruel. He has been known to betray confidences out of boredom and to say viciously cutting, even degrading, things about those close to him so as to ingratiate himself with strangers he admires. As such, his past is littered with people who despise him, but whenever Paul loses a friend he gets over it immediately, always assured of his own righteousness, oblivious to – or even actively, perversely, proud of – the emotional carnage he has wrought.



Knowing what we know about Paul – or, better, knowing what we know about people like Paul in our own lives, or in history or literature – are we *positively* oriented towards Paul, or are we *negatively* oriented towards him? Solomon & Stone are right: this *is* a silly question. But here is why that point does not cast any doubt on the value of thinking about ourselves as being positively or negatively oriented towards things. When I think about people I have known who are more or less Paul-like, I find that I am *ambivalent* towards the relevant person or people (though here I will focus only on Paul), in a number of ways. Calling to mind his artistic talents, his adventurousness, and his (rarely used) capacity for great kindness, I find myself evaluating him positively, wanting to spend time with him, and feeling pleased about the prospect of seeing what he will do next. But when I focus on his callousness, his volatility, and his superficiality, I find myself getting angry at him, thinking that his talents are wasted on such a detestable character, and wishing to avoid him altogether.

That is, I am *both* favorably *and* unfavorably disposed towards him – or, perhaps, I am “for” some of his features *and* “against” others. This seems, to me, like exactly the right thing to say about my orientation towards Paul. It seems much better than saying, for example, that I am *neither* positively *nor* negatively oriented towards him or any of his features. As Solomon & Stone (2002: p. 418) describe their own project:

Our argument is not that there is no such thing as valence or no such polarity or contrasts, but rather that there are *many* such polarities and contrasts.

With this I agree. And if we can keep this in mind, we need not worry that appeals to valence will lead to an over-simplified picture of ourselves. None of this marks a retreat from valence, but rather progress in our thinking about it.

Further, valence might be very important. It is quite plausible (though not universally accepted), for instance, that what is in your *interest* is largely determined by your “pro”- and “con”-attitudes and that what it is *rational* for you to do is largely determined by what you favor or disfavor. How it is *morally* permissible for you to treat others is plausibly constrained by what *they* favor or disfavor. And what we are for or against plausibly determines, at least in large part, what really *matters* to us and, perhaps, what *kind of person* we are. A number of philosophical approaches can agree on these general points (although they disagree about the details) and the notion of valence in general allows us to capture *what it is* about which they agree.

But what is it, really, for someone to be for or against something? Is there some feature that all positively valenced states share in virtue of which each is positively valenced? Is there some feature that all negatively valenced states share in virtue of which each is negatively valenced? At the beginning of inquiry, we should be open to any number of answers to these questions (including a “no” answer to each question). But here is one possibility: positively valenced states are those that are *correct* (or fitting or merited or appropriate) to have in response to things of value, and negatively valenced states are those that are *correct* (etc.) to have in response to things of disvalue.<sup>8</sup> For example, if X is admirable, then it is correct (etc.) to believe that X is admirable, to feel admiration towards X, and to act in admiring ways towards X. If X is shameful, then it is correct (etc.) to believe that X is shameful, to feel ashamed of X (if X is something you have done or some way that

---

<sup>8</sup> This is not to take a stand on whether the correctness of valenced states is ultimately to be explained in terms of value, or whether value is ultimately to be explained in terms of the correctness of valenced states (or whether there is “no priority.”) The proposal mentioned in the main text is neutral among these options. See e.g. Richard Y. Chappell (2012) for recent discussion of these options.

you are) or to feel contempt towards X (if X is something that someone else has done or some way that someone else is), and to act in ashamed or contemptuous ways towards X.<sup>9</sup>

This initial account of what all positively or negatively valenced states have in common seems quite plausible and we will have cause to return to it many times below. But it might not quench our investigative thirst. Even if we accept this initial account, we might still wonder whether e.g. evaluation, motivation, and feeling have something *deeper*, or more *descriptive*, in common in virtue of which they are all ways of favoring or disfavoring what they are about, and in virtue of which they are all beholden to value or disvalue in the ways just mentioned. Think about it like this. Let us suppose that some states – e.g. beliefs that something is good or bad, right or wrong, reasonable or unreasonable, etc. – have what we can call *evaluative valence*. For example, a state has positive evaluative valence with respect to X when that state represents X as having some positive normative or evaluative property, e.g. goodness. Next, let us suppose that at least some states have what we can call *motivational valence*. For example, a state has positive motivational valence with respect to X, roughly, when that state would motivate the subject of that state (to some extent, in the right conditions) to pursue, protect, or promote X.<sup>10</sup> Finally, let us suppose that some states have what we can call *phenomenal valence*. For example, a state has positive phenomenal valence with respect to X when that state is a way of *feeling* positively towards or about X.

On the one hand, in the ordinary course of events, when one is for (or against) something in one of these ways, one is for (or against) that thing in all of these ways. I do not just enjoy playing soccer on weekends, I also think that it is a valuable way to spend my

---

<sup>9</sup> As we can see in the case of Paul (see above), the idea in the main text does not imply that things are either of value or disvalue (or neither), but not both. Thanks are due to Justin D’Arms (p.c.) for encouraging me to clarify this point.

<sup>10</sup> We will complicate our account of positive and negative motivation in Chapter 4 below.

time (in several respects) and I am disposed to go and play. I do not just think that honesty is good, but I am motivated to be honest and I feel upset when I encounter dishonesty. And so on. And these connections do not seem accidental. We feel that there *have* to be some interesting systematic connections between these kinds of valence that mental states can have or lack, or between the kinds of mental states that “carry” these kinds of valence. And it seems like these connections have to consist in more than *just* the normative similarity mentioned above.

There are three very ambitious approaches one might pursue here. First, one might think that all valence is, at bottom, *evaluative* valence. One might claim, say, that to be motivated to  $\Phi$  is to desire to  $\Phi$ , and that to desire to  $\Phi$  is to represent  $\Phi$ -ing as good, or as what one ought to do. In this way, we might try to explain motivational valence in terms of evaluative valence. Similarly, one might claim that to feel positively towards X is to represent X as good in some way, and in that way one might try to explain phenomenal valence in terms of evaluative valence, too. In these ways, one might attempt to explain the systematic connections between the putatively distinct kinds of valence by showing that, at bottom, they are all really manifestations of evaluative valence.

The second very ambitious proposal puts *motivational* valence at the center. One might claim, say, that to represent X as good, or to feel positively about X, is to be positively motivated with respect to X. In these ways, one might hope to show that evaluative valence and phenomenal valence are really just manifestations of motivational valence.

The third very ambitious proposal would have us explain motivational and evaluative valence in terms of *phenomenal* valence. For example, perhaps what it is to be positively motivated with respect to something is to feel positively about it, or is to be disposed to act

in ways that someone who felt positively towards it would act, all things being equal. And maybe representing something as good just is, in one way or another, to feel positively about it, or maybe it is to think about it in the way that someone who felt positively towards it and possessed evaluative concepts, say, would think about it, all things being equal. By these means we might try to show that feeling, instead, can play the starring role in a unified account of valence.

But there are *prima facie* reasons to worry about each of these very ambitious proposals. Even though there do seem to be interesting systematic connections between evaluations, motivations, and feelings – and between evaluative valence, motivational valence, and phenomenal valence – it can also seem as though such states and such kinds of valence can come apart. It seems like we can imagine someone being for something in one of these ways but not in the others. For example, I might be motivated to put off my work until the last minute, even if I do not feel positively about putting off my work or think that it is a good thing to do. In my darker moments, I might feel pleased when a rival stumbles, even if I am not motivated to make him do so and even if I do not value his failures, not even in those periods of darkness. And so on.

It is not nearly obvious whether valence of one (let alone all) of these kinds can be explained in terms of valence of another kind, or whether all three kinds of valence can be explained in terms of some other kind of thing. So, at the beginning of inquiry, we should be open-minded about whether all of these kinds of valence, or the states that carry them, have something deeper in common, i.e. deeper than the fact that they are all correct responses to value or disvalue in the ways discussed above.

### Section 3: Phenomenology and Intentionality

Although I am quite sympathetic to one of the very ambitious proposals discussed above – the one which ultimately attempts to explain all valence in terms of feeling – I will not defend it in this dissertation. My aim here is less ambitious. I will only attempt to defend a particular view of *phenomenal* valence, a view I will call a *phenomenal attitude view*. To explain what such a view amounts to, and what the alternatives to it might look like, a bit of background information will be helpful.

First, the phenomena in which I am interested, positive and negative feelings – which I will call *affective experiences* – are, one and all, “phenomenally conscious” mental states (Ned Block (1995)).<sup>11</sup> As I am using the phrase ‘affective experience,’ it refers only to states such that there is “something it’s like” (Thomas Nagel (1974)) for the subject to be in them.<sup>12</sup> Equivalently, this is to say that affective experiences as such have *phenomenal properties* or that they have *phenomenal character*. With these points in the background, consider the following two groups of affective experiences:

---

<sup>11</sup> Block introduced the expression “phenomenal consciousness” to pick out the felt, subjective, first-personal quality of conscious experience and to contrast it with other things philosophers or psychologists might mean when they talk about consciousness. The primary distinction Block wants to draw is between phenomenal consciousness and what he calls “access consciousness.” A state is access-conscious, as Tim Bayne & David Chalmers (2003: p. 28) helpfully put it, “if by virtue of having the state, the content of the state is available for verbal report, for rational inference, and for the deliberate control of behavior. When I look at a red book, I can report the presence of the book (“there’s a red book”), I can reason about it (e.g., concluding that I must have put there when reading yesterday), and I can use its presence in deliberately directing my behavior (e.g., picking up the book and putting it back on the shelf). So my perception of the red book gives me the relevant sort of access to information about the red book. So my perceptual state here is access-conscious.” Typically, phenomenal consciousness and access consciousness occur (or fail to occur) together, e.g. in the example discussed by Bayne & Chalmers (and in actual cases of blindsight, as Block (1995) points out: both phenomenal consciousness and access consciousness seem to go missing in such cases). But *if* they can come apart, then the claim in the main text is only that affective experiences are, as such, phenomenally conscious. Whether they must, *therefore*, be access-conscious as well is a question about which I will remain neutral, at least at this point.

<sup>12</sup> So, for example, whatever e.g. Kent Berridge & Piotr Winkielman (2003) are talking about when they discuss “unconscious emotions,” it cannot possibly be affective experiences. None of this is to deny, of course, that we can have affective *dispositions* – tendencies to have affective experiences – that are not always manifested. Nor is it to deny that affective experiences may be tips of the icebergs of more complicated mental states – e.g. desires or full-blooded emotions – not all of whose aspects are conscious at any given time (or ever). It is not even to deny that things like “unconscious emotions” can be very similar in a variety of ways to affective experiences. But affective experiences *as such* are phenomenally conscious.

Group 1:

- *feeling attracted* to drinking the first cup of coffee of the day
- *feeling pleased* about the result of the game
- *feeling giddy excitement* about the upcoming concert
- *feeling schadenfreudic joy* at that hypocritical blowhard's electoral defeat
- *feeling mildly amused* at a co-worker's anecdote
- *feeling hopeful* about the novel you just started
- *feeling relieved* that your mom's getting better, finally
- *feeling proud* of your willingness to contribute a substantial portion of your insubstantial pay to Oxfam
- *feeling grateful* that your friend has forgiven you for canceling a planned lunch
- *feeling happily liberated* by God's non-existence
- *feeling awestruck* by the Grand Canyon at sunset
- *feeling drawn towards* that charismatic charlatan at the end of the bar

Group 2:

- *feeling averse* to the prospect of seeing one's 'ex' at a party
- *feeling pained* by the existing level of income inequality
- *feeling regretful self-loathing* about having let yourself get so fat
- *feeling wistful* about not having become a well-paid (or at least highly employable) lawyer
- *feeling grim annoyance* that so much of today has to be taken up with trivia
- *feeling envious resentment* of your hipster colleague for his weekend plans
- *feeling disgusted* by the texture of the mushrooms in that dish
- *feeling ashamed* about having treated a friend so
- *feeling despair* about your inability to make amends
- *feeling anxious fear* about your lack of a "plan B"
- *feeling indignant frustration* about your students' desires to have the material spoon-fed to them
- *feeling exasperated* by the irrationality of  $\pi$

The first general point to note about all of these feelings is that they are all phenomenally conscious (i.e. they all have phenomenal properties, they all have phenomenal character).

Second, in the terms used earlier, all of these feelings, of each group, have phenomenal

valence. More specifically, all of the feelings in Group 1 seem to have *positive* phenomenal valence: that is, each seems to be a way of feeling positively towards what it is about. And all of the feelings in Group 2 seem to have *negative* phenomenal valence: that is, each seems to be a way of feeling negatively towards what it is about. Hopefully these initial thoughts are not overly controversial, since I will just assume in what follows that some thoughts more or less along these lines are correct.<sup>13</sup>

To see what phenomenal attitude views and the alternatives amount to, let us now be a bit more careful about what we are, at the beginning of inquiry, taking positive and negative phenomenal valence to be. Let us say that, as a matter of definition, a state has positive phenomenal valence if and only if that state has one (or more) of a certain set of phenomenal properties (call these the *positive phenomenal properties*), and a state has negative phenomenal valence if and only if that state has one (or more) of a certain set of distinct phenomenal properties (call these the *negative phenomenal properties*).<sup>14</sup> This is intellectually unsatisfying, of course: what *unifies* the relevant sets of phenomenal properties? In virtue of what are the members of the former set “positive” phenomenal properties? In virtue of what are the members of the latter set “negative” phenomenal properties?

On one way of seeing things, the competing views of phenomenal valence to be considered in this dissertation are precisely competing attempts to provide intellectually satisfying answers to these reasonable questions. A key claim of phenomenal attitude views

---

<sup>13</sup> Consider also Chris Heathwood’s (2006: p. 559) intuitive motivation for thinking that enjoyment is a positive orientation towards what it is about: “If [someone] is enjoying [something], he must be liking it. If he is liking it, he must be “into it”. But this is just to say that he is for it, that he’s “pro” it.” (As it happens, Heathwood takes these points to support a motivational theory of pleasure. In Chapters 2 and 4 below I will, in effect, argue that they support no such thing.)

<sup>14</sup> It might be, of course, that there is only one kind of positive (negative) phenomenal property, and so only one member of the relevant set. Also, I want to thank Justin D’Arms (p.c.) and Declan Smithies (p.c.) for independently bringing to my attention issues that led me to characterize phenomenal valence in this way.



is that what unites the “positive” phenomenal properties is some *phenomenal* property that they all *share*, and ditto, mutatis mutandis, for the “negative” phenomenal properties. In this way, a phenomenal attitude view has it that all positively phenomenally valenced states *feel alike* in a certain way, and ditto, mutatis mutandis, for negatively phenomenally valenced states.

The major competitor views deny this, or remain neutral about it. According to *evaluative content views* of affective experience and phenomenal valence (more details below), what unites the “positive” phenomenal properties is, rather, the fact that they all have some shared *evaluative content*, and ditto, mutatis mutandis for the “negative” phenomenal properties. According to *motivational attitude views* of affective experience and phenomenal valence (more details below), what unites the “positive” phenomenal properties is the fact that they all come along with tendencies to *act* in certain ways, and ditto, mutatis mutandis, for the “negative” phenomenal properties. These three accounts of phenomenal valence – phenomenal attitude views, evaluative content views, and motivational attitude views – would each, if successful, offer an intellectually satisfying answer to the questions posed above. Each promises to tell us what all positively (negatively) phenomenally valenced states have in common in virtue of which each is positively (negatively) phenomenally valenced.

To understand the other major feature of the view to be defended here – the “attitude” part of phenomenal attitude views – a few preliminary words about the (potential) *intentionality* of affective experiences would be helpful. Let us assume for the moment that affective experiences are *intentional* mental states (we will defend this assumption in the next chapter). For something to be intentional, in the sense at issue, is for that thing to be *about* or *directed towards* things, it is for it to *represent* something or other. Let us assume, then, that

affective experiences are like that. Further, perhaps, let us assume that affective experiences are composed of both an “attitude” and a “content” or “intentional object” towards which that attitude is directed (this assumption, too, will be defended in the next chapter). The basic idea here is straightforward, at least initially. Take the paradigmatic intentional states: beliefs and desires. A belief that *p*, for example, is composed of an attitude – belief – and a content – that *p*. A desire that *p* has that same content but a different attitude. A belief that *q* has that same attitude but a different content. The assumption now in effect has it that affective experiences are like that: all affective experiences are composed of both an attitude and a content, and one affective experience can differ from another (or from another kind of mental state) either by having a different content, a different attitude, or both.

With these assumptions in place, we can distinguish between two approaches one might take to “locating” phenomenal valence. First, one might claim that the phenomenal valence of an affective experience is a feature of (or is reducible to) the special (i.e. valenced) kinds of *contents* that affective experiences have. Call these views “content views.” The evaluative content views mentioned above are the most promising species of this genus and we will discuss them in Chapter 3. Alternatively, one might claim that the phenomenal valence of an affective experience is a feature of (or is reducible to) the special (i.e. valenced) kinds of *attitudes* in which affective experiences consist. Call these views “attitude views.” Phenomenal attitude views are views of this kind. We will consider a distinct kind of attitude view – a motivational attitude view – in Chapter 4.

As we learned above, phenomenal attitude views have it that all positively (negatively) phenomenally valenced states feel alike in a certain way. Now we can say, more specifically, that phenomenal attitude views have it that all positively (negatively)

phenomenally valenced states consist of *attitudes* that feel alike in a certain way. The phenomenal valence of affective experiences is to be explained, then, in terms of the fact that affective experiences' proprietary attitudes have certain phenomenal properties. The attitudes involved in all "positive" affective experiences have a phenomenal property in common, a property that I will call "P", and the attitudes involved in all "negative" affective experiences have a phenomenal property in common, a property that I will call "N." To answer the questions from the beginning of this chapter, then, what it is to feel e.g. positively about something is to have an attitude that feels certain distinctive ways (i.e. it is to feel P-ly about such-and-such). Further, I will argue, one could have such attitudes without representing anything as good or bad, and without being motivated in any particular way, contrary to what evaluative content views and motivational attitude views claim.

#### **Section 4: Coming Attractions**

In Chapter 2 I will motivate my positive proposal about the nature of affective experiences and phenomenal valence. On my view, as seen, all states with positive phenomenal valence have something phenomenological in common and all states with negative phenomenal valence have something phenomenological in common. This view is subject to a very serious analogue of the heterogeneity objection to hedonic tone views of pleasure, and to respond I avail myself of resources developed by Timothy Sprigge (1988). Next, having hurdled this major initial stumbling block, I go on to defend a view according to which affective experiences are "impurely intentional" (Tim Crane (2009), David Chalmers (2004)), i.e. that such experiences are composed of both an attitude and a content, and each of these "parts" can make a difference to the phenomenal character of such experiences. Finally, I make the case for thinking that phenomenal valence "lives in" or is "carried by" the unique,

proprietary attitudes that affective experiences involve, and I say a bit more about those attitudes.

In Chapters 3 and 4 I consider alternative ways of understanding affective experience and phenomenal valence, each only slightly less ambitious than the very ambitious proposals considered in Section 2 above. In Chapter 3 I consider the idea that we can explain phenomenal valence in terms of *evaluative* valence. The idea here is that affective experiences are valenced in virtue of being representations of value. The best arguments for such views are based on the real, and important, connections that exist between affective experiences, on the one hand, and value or evaluative thought and judgment on the other. I will respond to three such arguments by showing that, even if we deny that affective experiences represent values, we can still explain (or at least allow for) the relevant connections between affective experiences and value or evaluative thought or judgment. I will also point out a few ways in which evaluative content views of affective experience and phenomenal valence are particularly implausible. As a result, I conclude, at least, that we should reject evaluative content views.

In Chapter 4, I consider the idea that we can explain phenomenal valence in terms of *motivational* valence. The idea here is that affective experiences are valenced in virtue of the relations they stand in to positive (or negative) motivation. The potential virtues of such views include their significant immediate intuitive appeal and their natural fit with attractive functionalist pictures of the mind. Unfortunately, however, such views are incorrect. It is at least conceivable that one could feel positively or negatively about something without being relevantly motivated with respect to it, as a number of cases and a brief argument in Chapter 4 will make clear. However, there are at least two fallback positions that someone

sympathetic to a motivational attitude view could adopt, each of which is consistent with the cases and the argument just mentioned. One is inspired by some things that David Lewis (1980) says about “mad pain,” and the other is inspired by some points that Timothy Schroeder (2004) makes about the relations between pleasure and desire. In responding to these fascinating proposals, I argue that even if views of these kinds were perfectly extensionally adequate, across all worlds, they would still get phenomenal valence wrong and so should still be rejected. We cannot explain phenomenal valence in terms of motivational valence.

Here is one way of understanding the thesis of this dissertation, in a nutshell: *to feel positively towards something is a novel way of being for that thing, of favoring it – it is not to be explained as a special case of being positively motivated with respect to it or of positively evaluating it.* This thesis will come out most powerfully, although perhaps most subtly, in Chapter 5. In Chapter 5 I will argue that feeling is necessary for love. Add all the positive motivations and positive evaluations you like: unless they come with feeling *as such, in particular*, they are not enough for love. This gives us conclusive reason to believe, I think, that evaluative content views and motivational attitude views cannot be the correct accounts (even if they work as a team) of affective experience and phenomenal valence.

Another, perhaps more important lesson of Chapter 5 is that it shows, clearly and unambiguously, that paying closer attention to affective experiences as such is (and has been) worth our while. Affect is necessary for love, and what could be more important than love?

Finally, in a brief concluding chapter, I point out some of what seem to me to be the most important questions that are raised but left unanswered in this dissertation. These include normative questions having to do with the relationships between desires and reasons

for action and the relationships between normative judgments, affect, and rationality, for example. I suspect that our account of practical reason quite generally should put affective experience at its core and my hunch is that the discussions of this dissertation will help provide some of the groundwork for such an account.

Other important unanswered questions have more to do with philosophy of mind, e.g. there are further questions about phenomenology, intentionality, and the place of mind in nature that are particularly pressing. I believe that a closer focus on affective experience can provide crucial new data for these debates (which have largely focused heretofore on *perceptual* experience), and that the account of affect defended in this dissertation might be especially helpful in these regards. In these ways and others, I suspect that the understanding of affective experience embraced in this dissertation might help us to make progress on a number of points on the continuum from ethics to metaethics and beyond.

## Chapter 2: Affective Experience and Phenomenal Valence

### Section 1: Introduction

In this chapter I will explain in a bit more detail what phenomenal attitude views amount to and I will draw attention to their attractions. First, I will attempt to motivate the idea that all positively phenomenally valenced affective experiences share a phenomenal property in common (P) and that all negatively phenomenally valenced affective experiences share a phenomenal property in common (N). But such claims are subject to an analogue of the heterogeneity objection to hedonic tone views of pleasure and pain. I will argue in Section 2.2, however, that this objection need not lead us to give up on P and N.

Having discussed the phenomenology of affective experiences – and having made my preliminary case for the “phenomenal” part of phenomenal attitude views – we then turn to a discussion of the intentionality of affective experiences. Although I will regard the claim that affective experiences are intentional mental states as an assumption in this dissertation, more or less<sup>15</sup>, I will attempt in Section 3 to make this assumption palatable. Then, in Section 4, I will defend the claim that affective experiences are *impurely intentional* mental states. (David Chalmers (2004), Tim Crane (2009)) As we will see, it is only if such a view is correct that phenomenal attitude views have any hope of being true. These discussions will constitute my case for the “attitude” part of phenomenal attitude views. Along the way, I

---

<sup>15</sup> I will explain this hedge in Section 3 below.

will note some crucial respects in which affective experiences are like desires. Finally, I conclude. My goal in this chapter is to show that phenomenal attitude views should be our default account of affective experiences and phenomenal valence, the account out of which we would have to be argued.

## **Section 2: Phenomenal Valence**

### **Section 2.1: P and N**

Think back to the examples of affective experiences listed in Group 1 and Group 2 in the previous chapter. As these examples show, we can feel many different ways about many different things. Our feelings can be purely positive or purely negative stances towards what they are about, but can also be ambivalent (e.g. when one *feels ambivalent* about having children). They can be directed internally or externally (or both), at future, present, or past events, or at the timeless. They can be silly or profound (or both), can be about relatively simple or massively complicated things. They can arise spontaneously and unbidden or as the result of habituation or training or reflection. They vary widely in their typical effects, too, e.g. on where and to what degree your attention is drawn, on what other thoughts or feelings you have, on what you are disposed to do or forebear, etc. They can be more or less stable features of our lives: they can help constitute our character or can be fleeting, even disturbing, outliers.

    Their phenomenal character (how they feel) can vary along several dimensions, too. The felt intensity of one's guilt or longing, for example, can vary, obviously enough, and this includes variation in how much of one's attention a state takes up. Perhaps distinctly, how



positive or negative one's feeling feels might vary, too.<sup>16</sup> Felt arousal varies as well: compare a child feeling tempted by the unattended cookie jar to someone's feeling of mild contentment with her current state.<sup>17</sup> (Relatedly, perhaps, affective experiences might feel more "innervating" or "enervating," as Spinoza had it (Solomon & Stone (2002: pp. 420-421).) And there are complicated and important relationships between affective experiences and feelings of *pleasure* and *pain* as well.

Here is one natural way to carve up the affective terrain: the *positively* phenomenally valenced affective experiences are, as such, the ones that are *pleasant* to undergo and the *negatively* phenomenally valenced affective experiences are, as such, the ones that are *painful* or *unpleasant* to undergo.<sup>18</sup> But I have my doubts about this natural proposal. On the one hand, it is undeniable that it is often true that when we feel good *about* something, we feel good, period. And it is hard to imagine someone experiencing e.g. a pleasure while not being positively phenomenally oriented towards anything.<sup>19</sup> But are feelings of attraction *pleasures* (or pains)? Are feelings of anger *pains* (or pleasures)? These seem like they might be cases where someone feels positively about something without experiencing pleasure, or cases where someone feels negatively about something without experiencing pain or displeasure. As such, I want to leave it open as to whether the categories of affective experiences

---

<sup>16</sup> Here is a potential example of this kind of variation from a semi-recent review of a Hollywood movie about Julian Assange, the founder of 'Wikileaks': "Mr. Assange, who has published hundreds of thousands of classified documents exposing the secrets of governments, banks and other powerful institutions, is a figure who inspires strong feelings, including ambivalence." (A. O. Scott (2013)) Also, one might think that one can feel frustration about such-and-such very intensely, but angry about thus-and-so less intensely, and one might still think that even "mild" feelings of anger are "more negative" than "strong" feelings of frustration.

<sup>17</sup> We will return to such claims in Section 3.5 of Chapter 4 below.

<sup>18</sup> For some potential reasons to distinguish pain from displeasure, see Timothy Schroeder (2004: Ch. 3).

<sup>19</sup> The following seems very plausible: if A is experiencing pleasure at t, then A is having a positively phenomenally valenced affective experience at t. But as we will see in the main text, it is not nearly obvious whether the following is true: if A is having a positively phenomenally valenced affective experience at t, then A is experiencing pleasure at t.

(positive and negative) will map neatly onto our commonsense hedonic categories (pleasure and pain/displeasure, respectively).<sup>20</sup> And if there is space between positive affect and pleasure, and between negative affect and pain/displeasure, then maybe that is another way that the phenomenal character of affective experiences can vary, i.e. in terms of how pleasant or unpleasant they are to undergo.

Despite these numerous ways in which such experiences can differ from one another, even in terms of how they feel, I claim that all positive affective experiences have a phenomenological property in common (P) and that all negative affective experiences have a phenomenological property in common (N). But what *are* P and N? What are these phenomenal properties that all positively or negatively phenomenally valenced experiences share, respectively? What do they *feel* like?

Unfortunately, it might not be possible to give an incredibly helpful characterization of P or N. P and N are phenomenal properties and phenomenal properties are notoriously difficult to describe in illuminating ways. For example, if you were asked to describe to someone what it's like for something to look reddish, it is not clear what you could say other than (assuming that their "spectrum" is not "inverted") giving them a list of examples of red things and saying "it is like seeing those things." I have already given my list of examples above, so now I can say, for example, "to instantiate phenomenal property P is to have feelings that feel like those in Group 1."

Or one could be more ambitious and suggest e.g. that for something to look reddish is for it to look like fire trucks, tomatoes, etc. look to normal (or ideal) human beings in

---

<sup>20</sup> Put slightly differently, I shy away from the proposal that the positive (negative) experiences are, precisely, the pleasurable (painful) ones, because this identification seems poorly placed to capture the phenomenology of *favoring and disfavoring*, of being positively or negatively *oriented* towards such-and-such. Thanks are due to Justin D'Arms (p.c.) for suggesting this way of putting the point.

normal (or ideal) viewing conditions. Likewise, I might try saying, for example, that to instantiate P towards X is to be in a state in which you feel towards X like normal (or ideal) human beings in normal (or ideal) conditions feel about companionship, fair treatment, massages, sweet foods, etc. Maybe such accounts will approach extensional adequacy, in the case of visual experiences and affective experiences, if we are someday able to provide incredibly insightful accounts of the relevant subjects, the relevant conditions, and great lists of examples. Alas, I do not have such insightful accounts or lists at the ready.

Finally, we might try the following: someone instantiates P (or N) towards X if<sup>21</sup> the subject of that state could, by reflecting upon the phenomenal character of that state, “see” that it is a way of favoring or being for X (or disfavoring or being against X). This would at least distinguish affective experiences from things like conscious *visual* perceptions (reflection on their phenomenal character does not “reveal” any valence) and from things like unconscious motivations or unconscious evaluative beliefs (there is no reflecting on their phenomenal character).<sup>22</sup>

---

<sup>21</sup> But probably not “only if,” given e.g. the fact that non-human animals and very young children can presumably have affective experiences despite having limited reflective or conceptual capacities. We will return to such issues in later chapters.

<sup>22</sup> Justin D’Arms (p.c.) has posed to me the following reasonable and difficult question about the points just made in the main text: does the fact (assuming for the moment that it is a fact) that we can “see” upon introspection that certain classes of conscious experiences are ways of e.g. favoring their intentional objects imply that we see this by noticing some *single* phenomenal property (P) that such experiences share? Isn’t it possible that people could see upon introspection that a variety of conscious experiences were e.g. favorings without doing so by finding some one phenomenal property that they all share? Whether such a scenario is *really* possible will depend, at the end of the day, on what the correct *theory* of phenomenal valence is (i.e. whether a phenomenal attitude view, an evaluative content view, or a motivational attitude view is correct), and on what the correct account of the relations between introspection, justification, and consciousness is (see e.g. Declan Smithies (2014) for helpful recent discussion), and so space precludes a fully satisfactory answer. But two remarks are appropriate here. First, why *would* you conclude, *merely on the basis of introspection*, that all of those experiences are favorings if you did not see any phenomenal commonality among them? Second, I suspect that our responses to the heterogeneity objection below will help assuage these kinds of worries, and so I ask the reader’s patience.

It seems to me that when we think carefully about affective experiences and phenomenal valence, we can see that what it is to feel positively about something is to instantiate a certain phenomenal property (P), and we can see that what it is to feel negatively about something is to instantiate a different phenomenal property (N). Even though positive (negative) affective experiences differ widely from one another in a number of respects, even in phenomenological respects, I still think that we find some phenomenal similarity that unites them. What makes e.g. the positively phenomenally valenced states more than just a disunified hodgepodge of experiences is precisely the fact that they share P in common. There is some respect in which all of the kinds of feelings in Group 1 are phenomenally similar to one another, and there is some respect in which all of the kinds of feelings in Group 2 are phenomenally similar to one another, and these phenomenal similarities are what ultimately justify our grouping the feelings together in these ways.

## **Section 2.2: The Heterogeneity Objection**

But one might reasonably doubt that there is *any* way in which *all* of the members of either of these Groups are phenomenally similar to one another. The doubts here are analogous to those that give rise to the heterogeneity objection to hedonic tone theories of pleasure and pain. These theories claim that what makes all pleasures pleasures is something phenomenological – namely, a positive hedonic tone – and that what makes all pains pains is also something phenomenological – namely, a negative hedonic tone. But many philosophers have denied that there is *anything* phenomenological that all pleasures or all pains share. Here is Chris Heathwood (2006: pp. 25-26) motivating the problem<sup>23</sup>:

---

<sup>23</sup> Heathwood cites, among others (including Feldman) Richard Brandt (1998) as presenting the original version of this problem. Karl Duncker (1941: p. 407) calls it “the old dilemma.”

As has been widely observed, pleasure is a diverse and varied phenomenon. There are bodily pleasures, like those had from relaxing in a Jacuzzi tub, from sunbathing on a warm beach, or from sexual activities. There are gustatory and olfactory pleasures... There are what we might call “emotional pleasures,” such as the elation of receiving an ovation or the satisfaction of completing a difficult and worthwhile project. There are more “cognitive” pleasures, such as the pleasure derived from working on a crossword puzzle, from reading an insightful philosophy paper, or from listening to an amusing anecdote. There are aesthetic pleasures, like those derived from listening to beautiful music or from taking in a powerful sculpture.

It is hard to see what phenomenal similarities there might be between all of these disparate cases. We might well conclude, like Fred Feldman (2004: p. 79), with the following:

Each of these experiences involves a feeling of pleasure...yet they do not feel at all alike. After many years of careful research on this question, I have come to the conclusion that they have just about nothing in common phenomenologically.

Christopher Hill (2009: pp. 209-210) puts the point even more strongly:

Careful examination shows that there is no felt quality that is common to all experiences that are pleasant, or even to any large subset of those experiences. Nor is there a common qualitative component of the experiences that we find unpleasant.

And if these points apply when we are only considering *pleasures*, then they will apply with even more force to affective experiences, given that affective experiences seem to include not only pleasures (and pains), but other feelings, too, e.g. feelings of attraction and felt urges.

The responses we might give to the *affective* version of the heterogeneity objection can be illuminated by thinking about the responses that one might give to the original, *hedonic* version. One response we might offer to the original, hedonic version of the objection is to claim that it shows that there is *no such thing* as pleasure or pain. But this seems unattractive, as does the analogous claim about affect (i.e. that there is no such thing as phenomenal valence.) A second response to the original version would have it that, although there are such things as pleasures and pains, there is no one feature or set of features that unites all

pleasures or pains: an experience's being a pleasure is just a brute fact about it, perhaps, or there are a series of overlapping commonalities among pleasures, even though there is no one feature or set of features that they all share, say. While this response and the analogous response we might offer to the affective version of the objection are not as obviously unsatisfactory, I suggest that we try to do better.

In fact, I think that this is a particularly illuminating way to think about what phenomenal attitude views, evaluative content views, and motivational attitude views of phenomenal valence are all up to, and what they disagree about. Each offers a (competing) unifying explanation of what unites the disjunction of experiences that constitutes the positively phenomenally valenced affective experiences and of what unites the disjunction of experiences that constitutes the negatively phenomenally valenced affective experiences. A phenomenal attitude view says that they are united by P or N. An evaluative content view says that they are united by the values or disvalues they represent. A motivational attitude view says that they are united by the positive or negative motivations with which they come along. And one might imagine analogous ways of responding to the original, hedonic version of the heterogeneity objection, too: pleasures are pleasures because of how they feel, or because of the values they represent, or because of their attendant motivations.

Understood in this way, the *force* of the heterogeneity objection (in its affective and hedonic formulations) is precisely that it pushes us away from the first unifying accounts just mentioned – those that appeal to shared phenomenology – and towards the other accounts – those that appeal to evaluation or motivation. Each of these latter kinds of view can, in principle, accept that there is no single phenomenal feature that all positively (negatively)

phenomenally valenced states share. And given the phenomenal dissimilarities of the items on e.g. Heathwood's list, doing so can seem quite attractive.

Nevertheless, such a response would be unwarranted. Another perfectly adequate response to the affective version of the heterogeneity objection (hereafter just 'the heterogeneity objection') is analogous to Timothy Sprigge's (1988) underappreciated response to the original, hedonic version of the objection (see also York Gunther (2004)). Sprigge's basic response to the original heterogeneity objection goes by way of an appeal to the distinction between determinates and determinables. Consider visual experience. Presumably, (e1) a visual experience as of scarlet has much in common, *as a matter of phenomenology*, with (e2) a visual experience as of brick red (i.e. a different shade of red). But one could deny this phenomenal commonality claim in the following way. There is no aspect of the phenomenal character of e1 that is *identical* to any aspect of the phenomenal character of e2: after all, *each* "pixel" of the one, as it were, *looks different* from *each* pixel of the other.

But denying that there are any phenomenal similarities between experiences of scarlet and experiences of brick red for these reasons would be overkill. Surely there are some *more general* phenomenal properties that each experience shares, e.g. each is a determinate of the determinable: visual experience of reddishness. And there are yet more general phenomenal determinables. Take a visual experience as of pure yellow and a visual experience as of pure blue. In this case the pixel point from above is even more compelling. Nonetheless, of course there are phenomenal similarities: they are both visual experiences as of color. The phenomenal character of your visual experience of pure yellow is surely more

similar to the phenomenal character of your visual experience of pure blue than it is your auditory experience of middle C, at any rate.

Sprigge's idea (here translated into explicitly affective terms) is that affective experiences are like visual experiences in the above respects. Imagine two specific but distinct positive affective experiences, e.g. (e3) feeling pleasantly relaxed by the warmth of the sun on your shoulders while lying on the beach on vacation and (e4) a soldier's feeling of steely resolve not to let his comrades down while on night patrol in a hostile city. Perhaps the maximally specific (low-level, determinate) phenomenal properties that would be analogous to the "pixels" in the visual case would be the specific bodily feelings that each experience involves, or things like warmth/coolness, felt intensity, specific aspects of content represented, their arousal, specific effects on attention, etc. We can imagine that *all* of the "pixels" of (e3) are distinct from *all* of the pixels of (e4). The pixels of (e3) include warmth, drowsiness, inattentiveness, passivity, etc., while the pixels of (e4) might include coolness, alertness, activity, etc.

Nevertheless, it seems to me that concluding on this basis that there are *no* meaningful phenomenal similarities between these two cases is precisely as unjustified here as was the analogous conclusion reached about (e1) and (e2) above. What is the phenomenal similarity? Precisely P: both experiences *feel* like positive orientations towards their intentional objects.<sup>24</sup> P and N are, like phenomenal color (i.e. the phenomenal character of a visual experience as of color), very thin and very general phenomenal properties. And yet, despite their thinness and their generality, they are quite meaningful.

---

<sup>24</sup> Gunther (2004: p. 45) makes congenial points: "Feelings, like colors, are more fine-grained than the nouns most of us have at our disposal. In fact, like colors, feelings also have different "shades". Some manifestations of anger may be more intense than others, some joy more brilliant, and so on. While this fine-grained character of feeling is undeniable, it doesn't change the fact that attitudes are accompanied by distinctive feeling types."



They are, I suspect, what was implicitly guiding our sense that the feelings listed in Group 1 all belong together, and that the feelings in Groups 2 all belong together.

Nevertheless, one might still worry that this appeal to the distinction between determinates and determinables does not satisfactorily solve the problem. One might worry that this response merely points to a logical possibility that, as of yet, we have been given no reason to think is actual. And without further reason *to* embrace this Sprigge-like response to the affective version of the heterogeneity objection, we might think that its employment here amounts to special pleading.

To show that this is not special pleading, I would like briefly to consider motivational and evaluative valence. It looks like each of these will face their own heterogeneity problems. Begin with motivation. Suppose that Alfred and Barbara, brother and sister, were left a work of art by their departed grandfather. Alfred is motivated to protect it and to jealously guard it by keeping it locked in the basement. He has no interest in praising it or showing it to others. Barbara, on the other hand, is motivated to hang it on the wall at work so that others can enjoy it and so that she can talk about it. She is not at all motivated to keep it safe and secure.

In this story, it seems clear enough that both Alfred and Barbara are positively motivated with respect to their grandfather's artwork. But what *particular* motivations do they have in common with respect to that artwork? None, perhaps. But that would not show, it seems to me, that they are not both positively motivated with respect to their grandfather's artwork, or that positive (negative) motivational valence is just a disunified hodgepodge. What we should say about such a case is precisely analogous to what we said in response to the affective version of the heterogeneity objection: there is some more *general*

sense in which both Alfred and Barbara are positively motivated with respect to their grandfather's artwork.

Similar points apply to evaluative valence. Suppose that Carla believes that Einstein's general theory of relativity has impressive theoretical virtues and Donald thinks that chocolate tastes great. Each person applies positive evaluative concepts to something, and so each person is positively evaluatively oriented towards those things. But Carla and Donald's judgments presumably have very little, if anything, *else* in common.<sup>25</sup> The standards appropriate to evaluating physical theories are just entirely different from the standards appropriate to evaluating things for tastiness. But that does not show that there is no such thing as evaluative valence, or that positive (negative) evaluative valence is just a disunified hodgepodge. Despite the lack of particular, determinate overlap between Carla and Donald's judgments, there is still a *general* property that they share: namely, each person *positively* evaluates something.

Maybe there are special reasons to think that Sprigge's kind of response to the affective version of the heterogeneity objection is less powerful than that kind of response to the motivational or evaluative versions of the objection, but I do not see what those reasons might be. And so, since Sprigge's kind of response is compelling in the motivational and evaluative cases, and since those cases seem very similar to the affective case, I conclude that we have good reasons to embrace Sprigge's kind of response in the affective case, too, and not merely to save our favored theory. I have not shown, of course, that this kind of response to the heterogeneity objection is *better* than a response that unites the positively

---

<sup>25</sup> One might respond that, in this case, they do have something in common. Namely, they each think that certain responses to things would be fitting or correct. This might be correct, but it seems to leave something out, namely that they each think that certain *positive* responses to things would be fitting or correct.

(negatively) phenomenally valenced experiences by appealing to evaluation or motivation instead of to phenomenal similarity. But I believe that I have shown that phenomenal attitude views have some extremely plausible things to say in response to the objection, and thus that the affective version of the heterogeneity objection is not itself a particularly strong reason to embrace evaluative content views or motivational attitude views.

### **Section 3: Affective Experiences are Intentional States**

#### **Section 3.1: Overview**

Having now discussed the phenomenology of affective experiences, and having said a bit to motivate a particular account of phenomenal valence, I will discuss the intentionality of affective experiences for the rest of this chapter. As we will see, it is only if certain claims about the intentionality of affective experiences are true that phenomenal attitude views of affective experience and phenomenal valence have any hope of being true.

In this section I will offer some initial reasons to think that affective experiences are intentional states, and then I will respond to objections. I will not respond to arguments for thinking that *no* conscious experiences are intentional<sup>26</sup>, but I will attempt to respond to arguments for thinking that affective experiences *in particular* are not intentional.<sup>27</sup> Then, in Section 4, I will compare two different intentionalist accounts of feelings: *pure intentionalism* and *impure intentionalism*, and I will argue that we should adopt the latter approach. This choice ends up being quite important, as we will see.

#### **Section 3.2: Initial Reasons to Think that Affective Experiences Are Intentional States**

---

<sup>26</sup> This is the only sense in which I am *assuming* that affective experiences are intentional states.

<sup>27</sup> For a very helpful overview of arguments of the former kind, see William Fish (2010).

To say that an affective experience – or any other kind of mental state – is an intentional state is to say that it is *about* or *directed towards* something or other. Trees, crutches, planets, and televisions, for example, are not intentional: they are not about or directed towards anything. Beliefs and desires, however, are intentional: the belief that the cat is on the mat is about cats and mats, for example, or perhaps it is about mental items or abstract entities that themselves stand in some interesting relations to cats and mats. There are a number of complexities here that I will not address for reasons of space.<sup>28</sup> Let us just say that the belief that *p* and the desire that *p*, for example, are each about *p*: *p* is the *content* or *intentional object* of that belief and that desire.<sup>29</sup> For a state to be an intentional state is for that state to “have” a content or intentional object, i.e. something that it is about or directed towards.

The claim that I want to defend here is that affective experiences are intentional states, i.e. that they have contents or intentional objects, that they are, one and all, about or directed towards this or that. And this is not just a matter of theoretical preference. The – or one – guiding thought of this dissertation is that affective experiences are *valenced* states: they are ways of favoring or disfavoring what they are *about*, are pro- or con-attitudes *towards* things. Valenced states – pro- or con-attitudes, ways of favoring or disfavoring things – *have* to be intentional states. One cannot just have a pro- or con-attitude towards nothing at all. One cannot favor or disfavor nothing. No mental states are valenced in their own right, so to speak: they must be valenced attitudes or orientations or stances *towards* or *about*

---

<sup>28</sup> Again, for an excellent overview of the numerous options here and many of the strengths and weaknesses of each, see e.g. Fish (2010).

<sup>29</sup> I will most often use ‘content’ and ‘intentional object’ interchangeably here, simply to refer to that which an intentional state is about or directed towards. But I might sometimes speak as though contents are special *kinds* of intentional objects, intentional objects that are propositions (or proposition-like, as e.g. sentences or events or states of affairs might be). This is merely a stylistic choice: given the way philosophers often speak about content, it would be awkward at best to say that Miki – a *person* – is the *content* of my love for Miki, even if we want to allow that Miki is what my love is about or directed towards. But I mean for all of this to be as theoretically neutral as possible.

something. So if affective experiences are not intentional states, then this dissertation is hopelessly misguided.

One argument for thinking that affective experiences are intentional states is, in fact, suggested by the very points just made: affective experiences are clearly valenced orientations to this or that, if a state is a valenced orientation towards this or that, then that state is an intentional state, so, affective experiences are intentional states. Given the discussions of the previous chapter and the earlier sections of this chapter, it seems to me that the first premise is incredibly plausible (think back to the examples included in Group 1 and Group 2, for instance). Given the points of the previous paragraph, the second premise seems unimpeachable. And the reasoning from premises to conclusion is beyond reproach. As such, I think that this argument is quite powerful. Nevertheless, we will see reasons to worry about it, where these can be understood either as reasons to worry about the first premise, or reasons to worry about an equivocation across the premises. So it is worth our while to bolster this initial argument if we can.

The next reason for thinking that all affective experiences are intentional states is incredibly general: all mental states are intentional states and affective experiences are mental states, so affective experiences are intentional states. But while a number of philosophers (e.g. Franz Brentano (2002)) do see intentionality as the “mark of the mental,” I am not here in a position to assess the plausibility of anything as sweeping as the first premise. So even though this argument is potentially quite promising, I will not put a great deal of weight on it here.

The third reason for thinking that all affective experiences are intentional states is less general, but similar to the previous reason in being guided by the value of theoretical

unity: *some* affective experiences are intentional, so (probably) all affective experiences are intentional. In support of the first premise, think back to the examples listed in Group 1 and Group 2, and consider the following points made by Brentano (2002: p. 481 – emphasis added):

*Certain* feelings undeniably refer to objects. Our language itself indicates this through the expressions it employs. We say that we are pleased with or about something, that we feel sorrow or grieve about something. Likewise, we say: that pleases me, that hurts me, that makes me feel sorry, etc. Joy and sorrow, like affirmation and negation, love and hate, desire and aversion, clearly follow upon a presentation and are related to that which is presented.

And if we accept that some affective experiences are intentional states, then it would be at least a bit surprising if other affective experiences were not. Our discussions to this point suggest to me, anyway, that affective experiences are at least a fairly cohesive kind. And so, if we think that a mental state's being or not being intentional is a deep metaphysical fact about it, then it would be surprising if affective experiences differed from one another in such a deep way, given this apparent cohesiveness. We would need to say more fully to support the inference embodied in our most recent argument, but I hope these points are at least suggestive.

These three initial arguments constitute the beginnings of a cumulative case for thinking that affective experiences are, one and all, intentional states. But that case will be bolstered substantially if we have something plausible to say about problem cases. That is the goal of the next section. In that section we will also come across an important additional reason for thinking that affective experiences are intentional states. (It will become clear below why I defer discussion of that reason.) The overall conclusion favored by these

numerous considerations is precisely that affective experiences are, one and all, intentional states.

### Section 3.3: Objections

Headaches and toothaches are kinds of affective experience. They are quite clearly feelings and they are quite clearly aversive in some sense. But what are they *about*? What do they *represent*? One initially compelling answer is: nothing at all. More generally, pleasures and pains do not seem to be about or directed towards anything. Moods, too, seem to be affective through and through but are not obviously about anything at all: “I’m not sad (elated) *about* anything,” one might say, “I’m just sad (elated).”<sup>30</sup>

But most difficult here are cases of what I will call ‘Wandering Affect Syndrome’ (WAS). These are cases (sometimes involving moods) in which, it seems, we feel one way or another about we-have-no-idea-what (maybe nothing), and that feeling “goes looking” for an intentional object. You are in a bad mood, your friend gently teases you, and your anger or contempt *comes* to be directed towards your friend, where it seemed to be about nothing in particular before. You are in a good mood, the cashier at the grocery store is efficient and polite, and your delight *comes* to be directed towards man’s general humanity to man, where it seemed to be about nothing in particular before.

Perhaps most interesting here are cases of what Andrea Scarantino (2010) calls “blindfright.” Psychologist Michael Gazzaniga has studied this phenomenon in “split-brain” patients (patients in whom the neural connections between the two hemispheres of their

---

<sup>30</sup> See e.g. John Searle (1983) and Nico Frijda (1994) for this non-intentional line on moods. Ned Block (1990, etc.) makes the case for something close to this line on pleasures and pains. Katalin Farkas (2009) might also embrace similar views.

brain have been severed.) The key features of such cases are nicely summarized by William Seager (2002: p. 671):

In this experiment, the patient was (via a clever optical device) shown an emotionally charged film exclusively to her right hemisphere. The patient then reported experiencing disturbing emotions despite being completely unaware of their source...[T]he patient went on to attempt to account for the emotional response, with such remarks as “I don’t know why, but I feel scared. . . I know I like Dr. Gazzaniga, but right now I’m kind of scared of him” (Gazzaniga, 1985, p.77).

In such cases, the (contents of the) “emotionally charged film” were clearly the *cause* of the patient’s fear, but the cause of a mental state need not be its intentional object. For example, it seems possible that a blow to the head (or some neuroscientist’s more precise neural intervention) could cause one to believe that Socrates was a lizard, or to hope that one’s feet grow three sizes overnight. But the blow to the head (or the surgeon’s scalpel) are not the *contents* of these mental states, even though they caused them. And once we remind ourselves of this distinction, it looks very likely that, at least for a moment, Gazzaniga’s patient felt afraid, but of nothing at all.

It should be acknowledged that for many affective experiences, e.g. those just listed, there does seem to be logical space for the claim that they are not intentional states. Beliefs and desires, on the other hand, have their intentionality built right into their grammar: the claim that someone has a belief or a desire with no content is, quite literally, nonsense. But the claim that someone feels pleased or pained, but that his feelings are not about or directed towards anything at all, is not nonsensical.



Nevertheless, such experiences are intentional.<sup>31</sup> My responses to the cases above are not novel but I think that their force has been under-appreciated. First, *localized* pains and pleasures (e.g. headaches and toothaches) do in fact seem to have some kind of intentional component. A pain in one's hand, for example, is, at a minimum, directed towards one's hand (see e.g. Michael Tye (1995) and Tim Crane (2009)).<sup>32</sup> That is all an intentionalist needs.<sup>33</sup> More general or non-localized pleasures or pains can be dealt with along the lines of moods.

Moods, instead of being about *nothing*, might, with equal plausibility, be said to be, in a sense, about *everything* (see e.g. Tye (1995).) Restricting our focus to phenomenology, to feel depressed is to feel negatively towards (practically) every *particular* thing one considers (or at least more negatively than one usually does), or it is to fail to feel positively towards things (or at least to feel less positively than one usually does). Slightly differently, moods might just have very *general* kinds of intentional objects (the world, all of this, etc.). The idea, then, is that when we say things like "I'm not depressed about anything, I'm just depressed", we can interpret that in something like the following ways: "I don't know what in particular is *causing* these grim feelings that I'm having about most everything, and my sadness or hopelessness or despair about any *particular* thing, at any rate, isn't all that my depression amounts to (it's more general than that.)" An account along these lines is at least as true to

---

<sup>31</sup> A more conciliatory approach to these cases goes as follows: *these* experiences are not intentional states, and so *they* are not ways of favoring or disfavoring anything, but, nevertheless, *other* feelings *are* intentional states, and so might be ways of favoring or disfavoring things, and those latter experiences are what this dissertation is about. I am disinclined to go this route – affective experiences seem like a more cohesive kind than this suggests (see above) – but the skeptical reader is encouraged to keep this fallback position in mind.

<sup>32</sup> We have known since Descartes (or *at least* since experimental psychology really got going in the nineteenth century) that a "pain in one's hand" is not really *in* one's hand. There can be tissue damage in one's hand, for example, without pain and there can be "pain in one's hand" without one's even *having* a hand (as we see e.g. in "phantom limb" cases.) Pain (or pleasure) in one's hand should, rather, be understood as pain in one's mind/brain *directed towards* (putative) goings-on in one's hand.

<sup>33</sup> This will come to seem more plausible after I sketch impure intentionalism below.

the phenomenology of moods as is the non-intentional line sketched above. And this kind of view can also make good sense of general, non-localized pleasures and pains (“I don’t know, my *whole body* just feels...ugh, yuck”).<sup>34</sup>

The cases of Wandering Affect Syndrome, including blindfright, are more challenging, but here, too, there are things to say. First, we could treat them exactly like we have just treated moods. After being shown the “emotionally charged film”, Gazzaniga’s patient might have begun to feel afraid of every particular thing he thought about, or about the world at large. Either of these kinds of account could explain why the patient felt afraid of Gazzaniga. These views would have it that the “emotionally charged film” was merely the cause, but not the intentional object, of the patient’s fear.

But we might worry that this leaves something out. What the patient is *really* afraid of, we might think, even if he *also* comes to feel afraid of Gazzaniga, etc., is *that film* (or what that film represented). But if we want to say this, then given that the patient was not consciously aware of that film (or of what it represented), we would have to try a second response to WAS cases: affective experiences are all conscious and intentional, but we are not always consciously aware of their intentional objects. That is, we might have utterly no idea, sometimes, what our feelings are about, even though they are always conscious and are always about something.<sup>35</sup>

---

<sup>34</sup> I would like to thank Declan Smithies for numerous discussions of these issues and for first getting me to see the attractions of intentionalist accounts of moods.

<sup>35</sup> A different response to split-brain cases, at any rate, is suggested by Bayne & Chalmers (2003: pp. 38-39). (Thanks are due to Declan Smithies (p.c.) for bringing this work to my attention.) On their view, what happens (or at least what might happen) in such cases is that the subject is *phenomenally* conscious of e.g. the scary film but is not *access* conscious of it. The scary film is registered in one’s phenomenal consciousness, but one does not believe that it is and one is not in a position to use this experience in reasoning or action. I have no arguments that this is impossible, but I would just like to register my skepticism about it. In any case, if this response could be made to work, then we would not necessarily need to allow the other curious possibilities mentioned in the main text.

This seems odd, for example, because when we know that we have a particular belief, we know *what it is* that we believe. When we know that we have some particular desire, we know *what it is* that we desire...or do we? It seems commonplace to realize that although we thought we wanted X, what we *really* wanted was Y. Sometimes, such claims seem to be *true*.<sup>36</sup> So in general, something can be the intentional object of one of our mental states even though we do not know that it is.

But affective experiences, unlike desires, perhaps, are essentially conscious states. And how could one be in a *conscious* intentional state without knowing what *it* is about? Here is my proposed answer: in just the same way that we could be in *any* intentional mental state without knowing what it is about. I will defer further discussion of how this might work until Section 4.4 below.

A third possibility might be even more radical. On this kind of account, we should move the intentional objects of affect inward. One might – in certain respects following e.g. William James (1884) and Antonio Damasio (1994) – think that what affective experiences represent are changes *in* one's mind or body. We might say, for example, that the intentionality of affective experiences is *exhausted* by what they tell us about such internal goings-on, e.g. how things are going with your viscera, whether your heart-rate is increasing or decreasing, whether your muscles are contracting or relaxing, etc. On this view, blindfright and other WAS cases are not particularly troubling. Gazzaniga's patient's feelings of fear represent what *all* feelings of fear represent, namely, internal changes in the fearful person.

---

<sup>36</sup> See e.g. Michael Smith's (1994) examples of people who (a) believe they desire to  $\Phi$ , but do not and (b) desire to  $\Phi$ , but believe that they do not. We will return briefly to Smith's account of the epistemology of desire in Chapter 4.

While such views might be onto something important, they seem to leave something out, something worth capturing. First, it typically seems from the first-person perspective that affective experiences *also* represent *external* (i.e. outside of one's mind or body) things. Here is James (1884: p. 203) on the phenomenon as quoted by Peter Goldie (2000: pp. 54-55 – emphases added):

An object falls on a sense-organ and is apperceived by the appropriate cortical center; or else the latter, excited in some other way, gives rise to an idea of the same object. Quick as a flash, the reflex currents pass down through their pre-ordained channels, alter the condition of muscle, skin, and viscus; and these alterations, apperceived like the original object, in as many specific portions of the cortex, *combine with it in consciousness* and transform it from an object-simply-apprehended into an *object-emotionally-felt*.

On this picture, we first perceive or imagine an object. This perception or imagining causes a variety of bodily changes in us, bodily changes that are registered or represented by affect-like feelings. But here is the key part that James (and Goldie<sup>37</sup> and many others<sup>38</sup>) and I want to capture: those bodily feelings, on the one hand, and that perception or imagining, on the other hand, combine (with each other) *in consciousness*. It more or less immediately *feels* or *seems* as though those feelings are *about* that (perceived or imagined) object. From one's first-person perspective, it seems as though affective experiences very often have outside-of-the-skin intentional objects, even if it also seems as though, in some sense, they have internal ones, too. And it is not clear how the view under consideration can capture these data.

---

<sup>37</sup> Here is Goldie: "When we talk, taking James's own example, of a grieving person feeling a pang in the breastbone, we want to say that the pang is a pang *for* the one who is being grieved over; although it is undoubtedly a feeling of something bodily, and can be pointed to as being in the chestbone, what makes it a pang of *grief*, rather than any old pang in the breastbone, is surely that it has been, as James says, 'combined in consciousness' with the object of the emotion." (2000: p. 55)

<sup>38</sup> Though I end up disagreeing with him about the intentionality of affect at a number of key points, I would like to direct the reader to Jesse Prinz's (2004), especially with respect to the issues presently being discussed in the main text. Prinz has *many* insightful and challenging things to say about these issues (and others).

Second, there are numerous and crucial normative *and* non-normative relationships between feelings and their (putative) outside-of-the-skin intentional objects. Consider a feeling that we would pre-theoretically be inclined to say is a feeling of attraction towards going for a walk. Why might we think that such a feeling really has *going for a walk* as its intentional object (instead of something inside-of-the-skin, say)? Here are some reasons. Such feelings tend to be caused e.g. by thoughts about *going for a walk* and also to include a disposition *to go for a walk*. All else equal anyway, such feelings also justify *going for a walk*. At a minimum, at least for those with the relevant concepts, such feelings tend to cause and justify e.g. the belief that one feels attracted to *going for a walk*. But unless such feelings are somehow directed towards going for a walk, these causal and normative relationships seem mysterious.

Now return to Gazzaniga's patient. We can well imagine that his feeling of fear would stand in these kinds of normative and non-normative relations to a number of external-object-involving behaviors, thoughts, etc. For example, his feeling would (prima facie) justify him in being, or at least would tend to cause him to be, wary of *every particular thing* he comes across. If we accept that external things can feature as the intentional objects of his feelings of fear, we seem well-placed to explain these phenomena, whereas the view inspired by Damasio et al. seems not to be. For these reasons, we should accept that affective experiences very often have outside-of-the-skin intentional objects, even in WAS-type cases.

And there is a more general – and ultimately more important – lesson to draw from our discussion of these normative and non-normative relationships between feelings and their putative contents or intentional objects. Namely, the existence (and the stability and

richness) of such relationships give us good reasons to think that affective experiences are, in fact, intentional states. This is the additional reason for that claim that I promised above.

There are multiple ways we might go here. On the one hand, we might claim that affective experiences are intentional states, and they have the (outside-of-the-skin) contents or intentional objects that they do, precisely in virtue of the fact that affective experiences play these functional roles. This approach is embodied by Robert Kraut's (1986) in which he develops (but potentially rejects) the idea that

even if feelings lack "intrinsic" intentional content, they may nonetheless acquire such content by virtue of their position in a causal-counterfactual and normative network that ties them to environmental input, to action, and to one another. (1986: p. 648)<sup>39</sup>

The idea here is that we can “get” (outside-of-the-skin) contents for affective experiences in the same way we seem to be able to “get” contents for many states, i.e. *via* their functional roles. Why is a particular belief a belief *that p*? Because it (normally or necessarily) plays certain roles in the believer's mental economy. Affective experiences might warrant a similar treatment, and if they do, then they will very often have (outside-of-the-skin) contents or intentional objects, as we have seen.<sup>40</sup> Let us call this the ‘functionalist proposal.’

Second, we could try reversing the order of explanation. That is, we might try claiming that affective experiences (normally or necessarily) play these roles in the subject's

---

<sup>39</sup> One general view of intentionality that Kraut considers has it that “an internal state-type S has the intentional content that-P if and only if S nomically covaries under optimal conditions with the state of affairs that-P.” (1986: p. 648) (See also our discussion of Robert Stalnaker in Chapter 4.) But as Kraut points out (1986: p. 649), if that is our model, then there is no in principle barrier to feelings “acquiring” their contents in just this way. Why is this feeling of fear about or directed toward *lions*? Because feelings of this type would tend, in optimal conditions, to be set off by lions. And so on. See Prinz (2004) for a seminal development of a view of the intentionality of the emotions of roughly this kind.

<sup>40</sup> I have to tread carefully here. In Chapter 4, I will argue, roughly, that affective experiences need not have any particular *motivational* profile. So I cannot appeal to the idea that affective experiences have the contents they do in virtue of having any specific motivational profile. (Cf. Michael Smith (1994: pp. 107-116), for example, who seems to embrace an account according to which *desires* have the propositional contents they do in virtue of their motivational profiles.)

mental economy in virtue of the fact that affective experiences are intentional states, and the fact that they have particular (outside-of-the-skin) contents or intentional objects.<sup>41</sup> Perhaps the most promising way for a defender of a phenomenal attitude view to implement this proposal, at any rate, would be to embrace some of the key claims made by philosophers friendly to the *phenomenal intentionality* research program.<sup>42</sup> According to at least some such views, the content – of at least one kind<sup>43</sup> – of a conscious experience is determined by its phenomenal character, and not by e.g. its functional role.<sup>44</sup> And then, the idea continues, the functional role of e.g. an affective experience would be (in some sense) determined by its

---

<sup>41</sup> I would like to thank Declan Smithies (p.c.) and Sigrún Svavarsdóttir (p.c.) for independently encouraging me to consider alternatives to the earlier functionalist proposal more seriously, given other points I want to make later (see e.g. the previous note).

<sup>42</sup> Jerry Fodor (1998) – *not* a subscriber to phenomenal intentionality, as far as I know – argues that the “inferential role” of a concept is determined (at least in part) by its content, and *not* the other way around. Fodor’s reasons are complex ones having to do e.g. with a desire to defend “informational semantics” and “atomism” about concept possession. I would like to remain as neutral as possible about these particular issues, and so Fodor’s reasons would not be my reasons here. The point of mentioning Fodor is, rather, to make it clear that there are a number of possible reasons to deny that content – of a concept or a conscious experience or whatever – is determined by its functional role. Thanks are owed to Declan Smithies (p.c.) for encouraging me to think about Fodor’s view in this context.

<sup>43</sup> As Terence Horgan & John Tienson (2002: p. 521) put it: “there is a kind of *narrow* intentionality that is pervasive in human mental life – a form of intentional directedness that is built into phenomenology itself and that is not constitutively dependent on any extrinsic relations between phenomenal character and the experiencer’s actual external environment.” (emphases added) A note on terminology: “narrow” content, if there is such a thing, is, as such, the kind of content that the mental states of intrinsic duplicates share. “Wide” or “broad” content, on the other hand, if there is such a thing, is the kind of content that the mental states of intrinsic duplicates might not share. “Internalists” about content believe that there is such a thing as narrow content (and they often think that it is quite important), though they might accept that there is also wide or broad content, too. “Externalists” about content believe that there is such a thing as wide or broad content (and they often think that it is quite important), and they either deny that there is any such thing as narrow content, or they at least argue that it is not as important as it might have seemed. For the seminal externalist thought experiments, see e.g. Hilary Putnam (1975) and Tyler Burge (1979). Ultimately, to adjudicate the issues being raised in this section (and elsewhere in this dissertation), we will probably have to take sides in the debates between internalists and externalists and between e.g. functionalists about content and e.g. phenomenal intentionality views of content. But these must be projects for another day.

<sup>44</sup> See e.g. Charles Siewert (1998), Horgan & Tienson (2002), Katalin Farkas (2008), and many of the essays in Uriah Kriegel’s (2013) anthology for helpful discussions of the phenomenal intentionality program.

phenomenologically-given content.<sup>45</sup> Let us call views of this general kind ‘intentionality-first proposals.’

At this stage anyway, I would like to remain neutral between functionalist proposals and intentionality-first proposals. The important point of our recent discussions is, rather, that according to a number of highly influential views of mental content, the fact that there are the relevant rich and stable causal and normative connections between affective experiences and their (putative, outside-of-the-skin) intentional objects gives us very good reasons to believe (a) that affective experiences are intentional states and (b) that their intentional objects are often outside-of-the-skin kinds of things. Given the other points made in this section, I conclude that we should accept both (a) and (b).

## **Section 4: The Case for Impure Intentionalism**

### **Section 4.1: Introduction, a Complication, and a Gift**

As we have now seen, there are a number of important debates *among* those who believe that affective experiences are, one and all, intentional states: some think, for example, that the intentional objects of affective experiences are all inside-of-the-skin, while others deny this. In the rest of the chapter, I will discuss another such internecine debate. This coming debate sets the stage, in certain ways, for the rest of the dissertation.

---

<sup>45</sup> Things will get complicated here when we throw attitudes into the mix. When we conjoin this kind of phenomenal intentionality approach with a phenomenal attitude view of phenomenal valence, for example, - apologies for any confusions caused by this terminological morass! - we should say that not only is the *content* of an affective experience determined by its phenomenology instead of by its functional role, but that the *attitude* in which such experiences partly consist is, too. Then, we might try saying that, on this approach, the functional roles of affective experiences are determined by (or grounded by or explained by) what we might (following Tim Crane (2009: p. 1)) call the *entire intentional nature* of affective experiences: affective experiences play the causal and normative roles they (normally or necessarily) do at least partly in virtue of *both* their (outside-of-the-skin) contents *and* the attitudes in which they partly consist.



This debate is (more or less – see below) that between *pure intentionalists* and *impure intentionalists* (Tim Crane (2009)).<sup>46</sup> We might follow Tim Schroeder & Ben Caplan (2007: p. 591) in thinking about *pure* intentionalism as the view that:

Sameness of qualitative character is to be explained [entirely] by sameness of the contents of experiences and...differences in qualitative character are to be explained [entirely] by differences in the contents of experiences.<sup>47</sup>

Or we might think about it like Crane (2009: p. 1) does, given that we are using his terminology: “[pure intentionalism is the view that] the conscious character of a state of mind is determined [entirely] by its *intentional* or *representational content*.” Then, we could contrast pure intentionalism with *impure* intentionalism in similar terms: “[impure intentionalism] is the view that the conscious character of a state of mind is determined by...its *entire intentional nature*.” (Crane (2009: p. 1)) Impure intentionalism will be distinct from pure intentionalism, then, if the ‘entire intentional nature’ of a state can include more than its content (e.g. an attitude).

While these ways of presenting these views provide helpful initial glosses on the relevant ideas, they are actually needlessly ambitious for our purposes, so I would like to weaken them a bit. Earlier, we learned that those friendly to the phenomenal intentionality program claim that the *phenomenology* of an experience determines the content (or entire intentional nature) of that experience. But Schroeder & Caplan and Crane seem to claim here that, according to intentionalism, the order of explanation is reversed: the *content* (or the entire intentional nature) of an experience determines the phenomenology of that

---

<sup>46</sup> See also David Chalmers (2004) who gives the opposing sides the names “pure representationalists” and “impure representationalists.”

<sup>47</sup> See e.g. Jesse Prinz (2011: p. 176) and William Seager & David Bourget (2007: p. 263) for slightly different versions of the claim.

experience. But for our purposes, as we will see, we need not (and should not) understand intentionalism in this ambitious way.

Instead, I would like to focus on the more general claims about which these seemingly conflicting approaches could agree, leaving order-of-explanation questions to the side.<sup>48</sup> In these more general terms, *pure intentionalism* is composed of two claims: (a) necessarily, if two experiences have the same phenomenal character, then they have the same contents; and (b) necessarily, if two experiences differ in phenomenal character, then they have different contents. And *impure intentionalism* will have two parts as well: (a) necessarily, if two experiences have the same phenomenal character, then they have the same entire intentional nature (e.g. same content and same attitude); and (b) necessarily, if two experiences differ in phenomenal character, then they differ with respect to some feature of their entire intentional nature (e.g. with respect to their contents or attitudes or both).<sup>49</sup> It is the disagreement between *these* views that will be especially, surprisingly important in what follows.

Finally, also in the spirit of trying to force us to focus only on the most crucial features of the views to be considered, I would like to grant to pure intentionalists the most favorable account of content that they could ask for. The important idea embodied by pure intentionalism is the idea that if experience (e1) and experience (e2) feel different in any way, then they *must* have different contents. We will have to embrace a *very* fine-grained view of

---

<sup>48</sup> David Chalmers's (2004: pp. 155-156) account of 'pure representationalism' (which he contrasts with 'impure representationalism') is presented in a congenial way, a way that does without the explanatory asymmetries mentioned above: "a *pure representational property* is the property of representing a certain intentional content (or the property of having a certain intentional content [...]). Intuitively, this involves representing things as being a certain way in the world...[And] *pure representationalism* is the thesis that phenomenal properties are identical to pure representational properties."

<sup>49</sup> One might be reluctant to call these ecumenical views "intentionalist," but no matter. The concerned reader should just add asterisks where she finds appropriate.

content to make this plausible. For example, I might consciously judge that Hesperus is bright while you consciously judge that Phosphorus is bright. But these conscious judgments might feel quite different, despite their having the same content in a coarse-grained sense – ‘Hesperus’ and ‘Phosphorus’ both refer to the planet Venus. But I am happy to allow, in assessing the prospects for pure intentionalism, a maximally fine-grained notion of content, where even objects or properties that *necessarily* occur together (e.g. being three-sided and being three-angled,  $4^2$  and  $2^4$ , etc.) can be or can be represented by different contents. Further, perhaps, I will also assume that any differences in modes of presentation (Gottlob Frege (1892)) or character (David Kaplan (1989)) or aspectual shape (John Searle (1983)) also count as differences in content for present purposes. I am happy, that is, to grant pure intentionalists a *maximally* fine-grained notion of content.

What matters for a phenomenal attitude view of phenomenal valence is whether *attitudes* can *as such* be phenomenal difference-makers. Impure intentionalism allows that they can be. Pure intentionalism, however, denies that they can be (our attitudes might *cause* us to undergo phenomenal changes, even according to pure intentionalism, but so can eating a steak.) Only the contents of a mental state (at least when conceived in a maximally fine-grained way), can, *as such* make a phenomenological difference, according to such views. This is the claim that I will challenge below.

## **Section 4.2: The Case for Pure Intentionalism**

First, however, I would like to consider arguments for pure intentionalism. The best argument for pure intentionalism appeals to the “transparency” or “diaphanousness” of

experience.<sup>50</sup> Suppose that someone, Eloise, is having a visual experience as of a tree before her. Here's how Gilbert Harman (1990: p. 39) puts the transparency point:

When Eloise sees a tree before her, the colors she experiences are all experienced as features of the tree and its surroundings. None of them are experienced as intrinsic features of her experience. Nor does she experience any features of anything as intrinsic features of her experience. And that is true of you too...Look at a tree and try to turn your attention to intrinsic features of your visual experience. I predict you will find that the only features there to turn your attention to will be features of the presented tree, including relational features of the tree "from here."

The basic idea is that, at least in the case of visual experience, the phenomenal character of the experience is exhausted by what the experience is an experience is *of*, its content or intentional object. Once we take into account the phenomenal character of the represented content, there is no phenomenal remainder.

Alex Byrne (2001) supplements these thoughts with an argument, one helpfully summarized by Katalin Farkas (2009: p. 44):

Whenever a subject notices a change in the phenomenal character of her experience, the 'way things seem to her' changes. And the way things seem to the subject is nothing but the content of her experience.

In Harman's Eloise case, this might be right. Any change in Eloise's visual phenomenology (i.e. any change in "the way things seem to her") seems to entail a change in how her experience represents things as being (e.g. *the tree* would seem to be a different color or shape or distance away.) And these are, precisely, changes in the contents of her experiences.

### **Section 4.3: Objections to Pure Intentionalism**

But do these thoughts hold up when we try to apply them to *affective* experience? In this section I will argue that affective experiences are *not* transparent or diaphanous in the ways

---

<sup>50</sup> See G.E. Moore (1903), Gilbert Harman (1990), Fred Dretske (1994), William Lycan (1996), Michael Tye (1995, 2000), etc. For criticisms of such transparency arguments see e.g. Christopher Peacocke (1983), Tim Crane (1998), Amy Kind (2003), David Chalmers (2004), Charles Siewert (2004), Daniel Stoljar (2004), and especially Ned Block (1990, 2003).

that visual experiences might be (according to Harman et al.).<sup>51</sup> In the process I will attempt to show, more generally, that pure intentionalism about affective phenomenology is implausible.

First, try to attend closely to a case in which you come to *have* or *lose* an affective experience. For example, find and focus on photographs of the ‘My Lai Massacre’.<sup>52</sup> Presumably, there will be a (brief) time period during which you go from not feeling any way at all about the photographs (or what they represent) to feeling, eventually, quite negatively towards them. Does the dawning and strengthening of that feeling literally change how you represent the world as being, literally change the *contents* of your experiences? Is that feeling itself, somehow, a *new part* of the content of your experience? One might reasonably doubt that it is. Perhaps more persuasively, keep looking at the photograph for awhile (if you can stomach it). Presumably, there will be a time period during which your strongly negative feelings weaken and, likely, dwindle more or less to nothing, through acclimation or immunization. Does the weakening and eventual absence of your feeling literally change how you represent the world as being, change the *content* of your experience? One might reasonably doubt that it has. And if this is right, then we have good reasons for thinking that affective experiences are not transparent or diaphanous.<sup>53</sup>

---

<sup>51</sup> It is not obvious even that any change in *visual* phenomenology entails a change in content. Christopher Peacocke (1983), Crane (1998), and Chalmers (2004) – among others – present what seem to be compelling counterexamples, e.g. involving blurry vision. But these cases seem to me to rely on something less than a maximally fine-grained notion of content, so I am reluctant to push these objections here. In any case, Byrne’s particular argument above might rely on an equivocation. (See Crane (2009: pp. 11-12) and Farkas (2009: p. 44) for discussion.) Second, even if it were true that any change in *visual* phenomenology entails a change in content, there is no special reason to think that that will generalize to encompass every change in phenomenology whatsoever. In fact, there is very good reason to think that it *will not* so generalize (see below).

<sup>52</sup> This is the horrible photograph I have in mind: [https://en.wikipedia.org/wiki/File:My\\_Lai\\_massacre.jpg](https://en.wikipedia.org/wiki/File:My_Lai_massacre.jpg)

<sup>53</sup> See Michelle Montague (2009: p. 173) for an account of such cases that *might* conflict with that offered here.

Interpersonal cases might make the point even more vividly. Think of things that you like and that your friends do not (or *vice versa*). Derek Parfit has some wonderful examples here:

[There are] some sensations that some people love and others hate, such as the sensations that we can give ourselves by eating milk chocolate, taking strenuous exercise, and having cold showers. Some of these likings or dislikings are odd. Many people hate the sound of squeaking chalk. I hate the feeling of touching velvet, the sound of buzzing house-flies, and the flattening, deadening effect of most overhead lights. (2011: p. 53)<sup>54</sup>

It seems possible and, moreover, perfectly ordinary, for two people to feel different ways *about the very same thing*, even things as determinate and immediate as present sensations. This seems even more obviously possible in cases involving contents or intentional objects of a less immediate and determinate nature.<sup>55</sup> Some people feel strongly attracted to the prospect of a Democrat winning the next U.S. presidential election, other people feel strongly averse to such a prospect.<sup>56</sup> As the objects of attraction or aversion become more abstract and phenomenally “thinner,” the possibility of two people having feelings with distinct phenomenal characters towards the very same object becomes practically undeniable. As Julien Deonna & Fabrice Teroni (2012: p. 77) put an analogous point about the *emotions*:

If it makes sense to say that what frightens Julie is what John is amused by...then we have reason enough to think that the difference between their two emotions is not to be located at the level of their respective contents.

And it does make sense to say this.

---

<sup>54</sup> Parfit (2011) calls these “hedonic” likings and dislikings.

<sup>55</sup> For reasons for thinking that phenomenology can be less than fully determinate, see Roderick Chisholm’s (1942) famous ‘speckled hen’ case.

<sup>56</sup> Here it might seem to matter what the nature of “cognitive phenomenology” is (see Tim Bayne & Michelle Montague (2011) for discussion). But I take it that whether it’s *sui generis* or sensory, the phenomenology of the occurrent consideration of the possibility *that a Democrat wins the next U.S. presidential election* is often, as David Hume would say, less vivid or lively than the content (of a present visual experience, let us suppose) *that there is a fire raging before me*.

If we are taking seriously the idea that affective experiences can have outside-of-the-skin contents (that is, contents like *that I go to the beach* instead of inside-of-the-skin contents like *that my heart is racing* – see Section 3.3 above), then it does not only seem possible, but commonplace, for two affective experiences to differ in phenomenal character without differing in content. Nevertheless, one might have the lingering thought that in all of these cases there *are* some differences in content. For example, in the case of the My Lai Massacre, in coming to feel negatively towards the photograph or towards what it represents, one might have come to represent it as *bad* in a way one did not before. (See Chapter 3 below for extensive discussion of this proposal.)

The idea that there are some differences in content in these cases is further bolstered by an argument from York Gunther (2004). Gunther has us consider a case where someone, let's call her 'Emma', hears a joke (the same joke, told in the very same way) three times. Naturally, Emma is (A1) very amused the first time, (A2) less amused the second time, and (A3) hardly amused at all the third time. Gunther argues that despite the fact that her different levels of amusement are directed towards the same joke told in the same way each time, Emma's different feelings – (A1), (A2), and (A3) respectively – should not be attributed the same content. Here is why:

The problem is that, by suggesting that the amusement's content remains constant...one is attributing to the individual a kind of content that doesn't capture her viewpoint. (2004: p. 50)

The idea seems to be that unless we attribute different contents to (A1), (A2), and (A3), Emma will end up looking silly when she is patently *not* being silly. Why in the world would one's feelings of amusement change so dramatically in such a case unless there was *some* change in content? That seems arbitrary or capricious and we have no good reason in this

case to think that Emma is even remotely arbitrary or capricious. We need to attribute different contents to the three episodes of amusement for (roughly) the same reason that we need to attribute different (fine-grained) contents to Gottlob's belief that  $a=a$  and Gottlob's belief that  $a=b$  (even though, as a matter of fact but unbeknownst to Gottlob,  $a=b$ ).<sup>57</sup>

I am not inclined to deny anything that Gunther says here. But nothing he says here commits us to claiming (i) that *every* change in affective phenomenology entails a change in content, let alone (ii) that any such changes in content account for the *entire* change in phenomenology.

First, and most importantly, sometimes we *are* arbitrary and capricious! If Gunther's entire case for something like pure intentionalism relies on the *impossibility* of our being silly, unreasonable, and whimsical, then his case will be very difficult to make.<sup>58</sup> Second, at least in the *interpersonal* cases discussed above, there need not be *any* question of rationality or intelligibility. I might love vanilla ice cream while you hate it but this difference need not reflect badly on either of us. In these cases, there is no pressure from charity to attribute different contents.

Finally, I agree with Gunther that, in this particular case, the idea that the (fine-grained) content changes from (A1) to (A3) is fairly plausible. Maybe by the second and third tellings, the additional content of Emma's (reduced) amusement is something like

---

<sup>57</sup> Gunther (2004) explicitly endorses this analogy.

<sup>58</sup> Imagine cases where a neuroscientist induces a new feeling in you via direct electrical stimulation of your brain. Also, this claim does not *in any way* threaten the intuitive response(s) to Frege puzzle cases. The claim here would be analogous to claiming that it is *possible* for one to drift between e.g. belief, doubt, and disbelief of a given proposition without there *necessarily* being any difference in the content of these attitudes, or even any difference in evidence that would "explain" the drifting.



...and I've heard this joke before or ...and I've heard this joke a bunch.<sup>59</sup> But even if that is right, and even if something like that is *always* right, these content changes will often not be all there is to the change in phenomenology.<sup>60</sup> Even if the content of (A1) is *The Joke* and the content of (A2) is *The Joke which I've heard before*, and this difference in content helps to make Emma's different levels of amusement *intelligible*, the phenomenological difference between uproarious amusement and mild amusement amounts to more than just *these* differences in content. Maybe e.g. going from liking to disliking something always involves (perhaps subtle) changes in (represented) content, and if this is correct, it is to the pure intentionalist's credit that they bring this to our attention. But I maintain that, nevertheless, those changes in content often do not *fully* account for the change in phenomenology.

Furthermore, there are other reasons to worry about pure intentionalism about affective experience. First, as we saw in Section 3.3 above, we might want to allow for the possibility that one can have an affective experience but be entirely unaware of what the content of that experience is. Pure intentionalism cannot allow this but impure intentionalism can, as we will see momentarily. Second, and much more importantly, pure intentionalism would force us to tell an awkward story about phenomenal valence. We will return to these issues at the end of this chapter and in the next chapter, at least. At any rate, as we will see presently, another problem with pure intentionalism is that, when you put it side-by-side with impure intentionalism, it just does not hold up.

---

<sup>59</sup> Presumably, however, we should not attribute to Emma differential representations of *funniness* in each case, at least if we are interested in making sure that she comes out as sensible. It would be strange for one to think that *The Joke itself* became less funny just because one's heard it before.

<sup>60</sup> I take this to be Ned Block's (highly plausible) response to Michael Tye's pure intentionalist account of the phenomenal character of orgasm. It is not necessarily that the representational contents of orgasms that Tye points to are not there, it is that those contents do not account for the full phenomenal richness of orgasm experiences. See Block (2003) for further discussion.

#### Section 4.4: The Virtues of Impure Intentionalism

The general argument for impure intentionalism about any kind of conscious experience is given nicely by Tim Crane (2009: p. 8):

The difference between feeling one's leg to be damaged and seeing it to be damaged is just the difference between *feeling* and *seeing*. In other words, it is a difference in what Searle and I call *mode*, and what others would call *attitude*. We already know that sameness of content does not suffice for sameness of mental states in general; a belief and a hope might have the same content. So why should we expect that it suffices for sameness of phenomenal states, states which are distinguished by their phenomenal character?

This additional, *attitudinal* logical space thus seems like an important resource to be able to draw on in our theorizing about the mind quite generally. Here is how I would like to draw on it in the present case. Compare (a) the phenomenal character of feeling attracted to going to the beach to (b) the phenomenal character of feeling averse to going to the beach and to (c) the phenomenal character of feeling attracted to going to the circus. Although the experiences referenced in (a) and (b) have the same content or intentional object, their overall phenomenal characters are distinct. Taken at face value, this implies that (1) the phenomenal character of an affective experience is determined at least in part by affective *attitude*. Second, although the experiences referenced in (a) and (c) share an attitude, their overall phenomenal characters are distinct. Taken at face value, this implies that (2) the phenomenal character of an affective experience is determined at least in part by their *contents* or *intentional objects* (at least when such contents or intentional objects are conscious). The impure intentionalist takes these thoughts at face value and so accepts that affective phenomenology is content-specific and attitude-specific (Declan Smithies (2012)), that it consists of a phenomenology of intentional content and a phenomenology of attitude-type

(Terence Horgan & John Tienson (2002)), that intentional modes and intentional contents can each have (distinct) phenomenal properties (Crane (2009)).

This kind of view has several virtues. First, appealing to conscious attitudes can help us give intuitive accounts of the cases where transparency seemed to fail. The differences in phenomenology in those cases do imply intentional differences, but they are differences in attitude instead of differences in content (like going from believing that *p* to doubting that *p* instead of like going from believing that *p* to believing that *q*).

Second, an appeal to affective attitudes can finally help us explain what is going on in cases of blindfright. Although the *content* or intentional object of e.g. Gazzaniga's patient's fear was unconscious, his (fearful) *attitude* towards that content was conscious. This might be possible, and it seems to be a positive feature of a view of the intentionality of affect that it would allow us to interpret these cases in this way.

Finally, once we have these phenomenally conscious attitudes on the table, we are well positioned to give a compelling account of phenomenal valence.<sup>61</sup> There are some telling analogies between desires and affective experiences that I would like to exploit. As the examples of affective experiences given throughout this chapter show, the contents of affective experiences can, like the contents of desires, be anything at all.<sup>62</sup> And the *valence* of

---

<sup>61</sup> On one way of understanding pure intentionalism, pure intentionalism denies that there is *any such thing* as a conscious attitude. The reasoning is as follows: if the only phenomenal difference-makers are contents, then how could there be conscious attitudes? But this would be too quick, as Declan Smithies (p.c.) and Sigrún Svavarsdóttir (p.c.) have each independently brought to my attention. One alternative possibility that Smithies suggested to me has it that, according to pure intentionalism, there is exactly one kind of conscious attitude that one can have, and every conscious experience includes that attitude. On this approach, there would be conscious attitudes, even according to pure intentionalism, but such attitudes could not be phenomenal difference-makers.

<sup>62</sup> Well, more or less. Maybe G.E.M. Anscombe (2000) is right that one cannot just desire a saucer of mud. Maybe the contents of desires have to be things with a sentence-like structure, e.g. propositions, states of affairs, events, etc. Maybe the same applies to affective experiences, but maybe not. I am neutral about these

desires is not, intuitively, a feature of their contents, but rather of the attitudes that they involve: call these “desiderative attitudes.” If you desire that *p*, you are not positively oriented towards *p* in virtue of representing *that p*. Rather, you are positively oriented towards *p* in virtue of *desiring* that *p*, i.e. in virtue of having a desiderative attitude towards *p*. We should, it seems to me, say something similar about affect. If you feel positively about *p*, you are not positively oriented towards *p* in virtue of representing *that p*. Rather, you are positively oriented towards *p* in virtue of *feeling positively* about *p*, i.e. in virtue of having an “affective attitude” towards *p*. In the case of both desires and affective experiences, valence seems to “live” in the attitude and not in the content.

And affective experiences seem to be like desires in another way. When you desire that *p*, you need not represent that *p is true*: you might know that *p* is false and wish it weren't! Desires are intentional states, but they are not in the business of “telling you” that the world actually is one way or another.<sup>63</sup> They are in some other line of work entirely. Here is G.E.M. Anscombe's (2000: p. 56) helpful overview of the kind of distinction that might be at issue:

Let us consider a man going round a town with a shopping list in his hand. Now it is clear that the relation of this list to the things he actually buys is one and the same whether his wife gave him the list or it is his own list; and that there is a different relation where a list is made by a detective following him about. If he made the list himself, it was an expression of intention; if his wife gave it him, it has the role of an order. What then is the identical relation to what happens, in the order and the intention, which is not shared by the record? It is precisely this: if the list and the things that the man actually buys do not agree, and if this and this alone constitutes a mistake, then the mistake is not in the list but in the man's performance (if his wife

---

issues. The point in the main text is only that aside from the very general, abstract restrictions just mentioned, anyway, the contents of desires and affective experiences can be anything at all.

<sup>63</sup> This is why I have been using Crane's (im)pure *intentionalism* instead of Chalmers' (im)pure *representationalism*. The latter, rightly or wrongly, suggests to me that all conscious experiences are in the business of telling you how things are, but the former does not. But I take there not to be a *substantive* difference between the relevant views.

were to say: “Look, it says butter and you have bought margarine”, he would hardly reply: “What a mistake! we must put that right” and alter the word on the list to “margarine”); whereas if the detective's record and what the man actually buys do not agree, then the mistake is in the record.

The idea is that some mental states – e.g. beliefs and ordinary perceptions – are relevantly like the detective's list: they are in the business, as I put it earlier, of telling you how things are. That is, they have a ‘mind-to-world’ direction of fit. But other states – e.g. desires and intentions – are relevantly like the shopper's list: they are in the business, instead, of telling you *to do* something or other.<sup>64</sup> That is, they have a ‘world-to-mind’ direction of fit.

Now although I am quite confident that affective experiences do *not* have a mind-to-world direction of fit (see below), I am not at all sure that affective experiences *have* a world-to-mind direction of fit (see Chapter 4). For these reasons, I quite like Michael G.F. Martin's (2002) more permissive discussion of non-belief-like representations:

On one way of talking about representation, beliefs and judgments both count as representational, while such states as hopes and desires do not...But in talking of representational or intentional content, one might have a broader sense of the notion in mind. One on which desires, hopes, and non-indicative sentences all count as representational as well, since they are all about (or of, or involve reference to) objects, properties and states of affairs, *even though they do not present anything as being the case*. Let us call this the *semantic* conception of representation, and the narrower conception of representation we can call the *stative* conception. (2002: pp. 386-387 – first emphasis added)

In these terms, desires are *semantic* representations, and are not *stative* representations, as “they do not present anything as being the case.” As I will often put the relevant idea, desires are not “truth-apt.” They favorably or unfavorably dispose you towards things being

---

<sup>64</sup> Michael Smith (1994: p. 115) attempts to draw the relevant distinction in less metaphorical terms: “The difference between beliefs and desires in terms of direction of fit can be seen to amount to a difference in the functional roles of belief and desire. Very roughly, and simplifying somewhat, it amounts, *inter alia*, to a difference in the counterfactual dependence of a belief that p and a desire that p on a perception with the content that not p: a belief that p tends to go out of existence in the presence of a perception with the content that not p, whereas a desire that p tends to endure, disposing the subject in that state to bring it about that p.”

or not being certain ways, but they do not themselves “tell you” that things are or are not those ways.

Intuitively, affective experiences are exactly similar. To feel attracted to  $\Phi$ -ing does not entail that you represent that you really are  $\Phi$ -ing: you might know that you are not and wish that you were! Also, it just sounds bizarre to say that e.g. positive feelings are true or false or veridical or not, although it sounds perfectly fine to say that beliefs or perceptions are true or veridical. If I tell you that I like nachos, and you respond that my feeling misrepresents reality, I will think that you are joking, and not just because nachos are so tasty.

Also, the most natural linguistic vehicles to use for directly expressing affective experiences are what York Gunther (2003) calls “expressives,” e.g. interrogatives, commands, requests, etc. and (most importantly) expressions like what Ayer (1952) calls “ejaculations” or what Jesse Prinz (2007: p. 17) calls “expletives,” e.g. “boo to having class on such a nice day!” and “hooray for class being cancelled!”<sup>65</sup> These linguistic constructions are not truth-apt and it stands to reason that the mental states they directly express are not, either.<sup>66</sup>

Impure intentionalism is fully compatible with these points. There is no reason to think – in advance, anyway – that all conscious attitudes are truth-apt representations of what they are about. On the other hand, it is very difficult to see how pure intentionalism

---

<sup>65</sup> The expressing/reporting distinction is key here. I *express* my desire for you to open the door by saying something like “please open the door,” while I *report* this desire by saying something like “I want you to open the door.” Likewise, I express my positive feeling about meeting Roger Daltrey by saying something like “woohoo: I’m meeting Roger Daltrey!” whereas I report it by saying something like “I am quite pleased to be meeting Roger Daltrey.” The “boo” and “hooray” language is inspired by Ayer (1952) and Blackburn (1984), but used to a different purpose here.

<sup>66</sup> This line of thought was inspired by Gunther (2003) who uses similar considerations to argue that linguistic expressions of emotion provide good reasons to deny that the force/content distinction applies across the board.

could accommodate these points. Such views suggest that *all* conscious experiences purport to tell you how things are. Suppose that I feel positively about Gauguin's paintings and you feel negatively about them. These experiences feel different, we can suppose, but the pure intentionalist cannot explain these phenomenal differences in terms of our having different conscious attitudes towards Gauguin's paintings. Instead, such phenomenal differences have to be explained in terms of our experiences having different contents. Suppose (to anticipate some of the discussion of the next chapter) that my experience represents that Gauguin's paintings are *good* but your experience represents that Gauguin's paintings are *bad*. These differences in content, let us suppose, are hypothesized to account for the phenomenal differences between our feelings.

But what do we say next? One option is just to leave things there: my experience has a *content* that tells me that Gauguin's paintings really are good, and so my *experience* thereby tells me that Gauguin's paintings really are good. On this approach, an approach that does without mention of attitudes altogether, my feeling would be truth-apt, and so this approach would conflict with the points recently made.

Another option is to say instead that I feel positively about Gauguin's paintings in virtue of (a) the fact that I have a mental state that has *Gauguin's paintings are good* as its content *and* (b) the fact that I have some attitude towards that content. That seems better, but what is the relevant attitude? After all, I might *wonder whether* Gauguin's paintings are good, or *hope* that they are, but surely *these* kinds of attitudes are not a part of my positive feeling towards Gauguin's paintings, the feeling that we were trying to explain in the first place. The answer that the pure intentionalist should give, rather, is that the relevant attitude

is one of *acceptance*, of taking-to-be-true.<sup>67</sup> I accept that Gauguin's paintings are good, and you accept that Gauguin's paintings are bad.<sup>68</sup> But on this view, too, it turns out that my feeling is truth-apt (taking something to be true is *paradigmatically* a truth-apt mental state), and so this view, too, conflicts with the points made above. It still looks like, according to pure intentionalism, all conscious experiences are as such in the business of telling you that things are a certain way. Maybe there is some way around this, but I do not see it.<sup>69</sup>

This is an important point, one that goes beyond the narrow topic of this dissertation. Even if all consciousness is intentional, that *does not* imply that our conscious states, one and all, are only there to tell us how things are (or could or could not be), as pure intentionalism seems to have it. We are not just believing and perceiving creatures, but also acting and feeling creatures. Impure intentionalism fits well with this more expansive, and more plausible, view of the conscious mind. So we have very good reasons to take this approach to the intentionality and phenomenology of affective experience seriously and to see where it leads.

---

<sup>67</sup> J. David Velleman's (1992a) presentation (but not endorsement) of an analogous proposal about *desire* can help us get a better grip on such a view. Proponents of the view that Velleman (1992a: p. 6) calls "the evaluative conception of agency" attempt to capture the valence of desires in the following way: "[Proponents of the evaluative conception of agency] incorporate the valence of desire into its content, by describing desire, not as a favorable attitude toward the representation of some outcome, but rather as an attitude toward a favorable representation of the outcome. The agent who wants to know the time is said, not to be favorably disposed toward "I know the time," but rather to accept a proposition such as "My knowing the time would be good." (1992a: p. 6) The suggestion in the main text is that pure intentionalists should say something exactly analogous about affect.

<sup>68</sup> If pure intentionalists claim (see note 61 above) that there is exactly one kind of conscious attitude, then I think that we have now seen good reasons for thinking that that attitude needs to be one of acceptance. What else could it be?

<sup>69</sup> None of this implies that, according to pure intentionalism, *all* mental states are truth-apt. For example, pure intentionalists might want to claim that the belief that *p* and the desire that *p* are distinct mental states (and one is truth-apt while the other is not) in virtue of the fact that they play *different functional roles*, even if, as far as phenomenology is concerned, there need be no difference between them. This response is fine, as far as it goes, but it will not help one respond to the objection just presented (and that about to come) in the main text, as far as I can tell.



## **Section 5: Conclusion**

Affective experiences as such have phenomenal valence. At the most general level, what it is for a state to have e.g. positive phenomenal valence is for it to have one (or more) of a disjunction of phenomenal properties. The examples listed in Group 1 were supposed to give us a rough initial feel for what this set of phenomenal properties might be like. I suggested that the set of positively phenomenally valenced states is unified by the fact that the members of that set all instantiate phenomenal property P. The heterogeneity objection, however, gave us reasons to worry that there are no phenomenal properties that all positive phenomenally valenced states share. But I argued that this objection should not lead us to embrace such a conclusion. In these ways I attempted to motivate the idea that affective experiences are united by their phenomenology. This was the first part of my defense of a phenomenal attitude view of affective experience and phenomenal valence.

The next part of that defense had me arguing that affective experiences are intentional states. This was important since, if affective experiences are going to be valenced orientations towards what they are about, they need to be intentional states. The final part of my defense of phenomenal attitude views consisted in my arguments for thinking that affective experiences are impurely intentional states. This was necessary in order to allow for the possibility that phenomenal valence is a feature of the attitudes that affective experiences involve, as phenomenal attitude views have it. Lastly, I attempted to present a few key features of phenomenal attitude views, both to bring out their attractions and to give the reader a better grip on what they really amount to.

But the next two chapters will have us consider alternative theories of affective experience and phenomenal valence. It is only when we have considered such theories that

we will be in a position to assess phenomenal attitude views properly. But I hope that the points made in this chapter put phenomenal attitude views firmly on the front foot.

.

## Chapter 3: Affect and Evaluation

### Section 1: Introduction

In this chapter and the next we will look at alternatives to phenomenal attitude views of affective experience and phenomenal valence: evaluative content views (this chapter) and motivational attitude views (next chapter). I will not attempt, in advance, to provide an exhaustive list of desiderata that such accounts must satisfy in order to be successful. Each of these accounts has strengths and weaknesses, and we should be open to the possibility that we will have to resort to weighing these strengths and weaknesses, with all the imprecision and inconclusiveness that such weighing typically involves, in order to determine which account is best. But there are at least three conditions that any such account must meet in order to be even minimally adequate.

First, the properties, said by the relevant account to be the properties that unite the positive (negative) affective experiences, have to be properties that all positive (negative) affective experiences *actually share*. Second, the properties appealed to should enable us to *distinguish* the positive from the negative affective experiences. It would not do, for example, to appeal to a property that all positive *and negative* affective experiences share in explaining *positive* phenomenal valence. These first two criteria ensure extensional adequacy, the third criterion demands a bit more. Third, the properties, said by the relevant account to be the properties that unite the positive (negative) affective experiences, have to be properties the

possession of which is suitably related to (more strongly: explanatory of) the *valence* of the experiences. We have to be able to see, upon reflection, perhaps with some coaching, how it is that someone's instantiating *that* property amounts to that person's being *for* or *against* something.

Phenomenal attitude views, evaluative content views, and motivational attitude views all seem to satisfy this third criterion, and I am not sure how any *other* view of phenomenal valence *could*. As we learned in the previous chapter, affective experiences are intentional states that consist of both a content and an attitude towards that content. As such, it looks like we have two fundamental options for explaining phenomenal valence: appeal to the special (i.e. valenced) contents of affective experiences (this would be to embrace a *content view*), or appeal to the special (i.e. valenced) attitudes in which affective experiences consist (this would be to embrace an *attitude view*). Where else *could* phenomenal valence “live”?

In this chapter, our focus is entirely on content views. Among content views, *evaluative* content views are clearly best-placed to satisfy the above criteria, the third criterion in particular. Such views claim, roughly, that a feeling is a positively phenomenally valenced orientation towards something iff and because it is a way of *accepting* that that thing has some *positive* normative or evaluative property, and ditto mutatis mutandis for negative phenomenal valence.<sup>70</sup> For example, my positive feeling about going to the beach is a

---

<sup>70</sup> There are many positive and negative normative or evaluative properties that we could represent something as having: e.g. goodness, rightness, virtue, courage, ought-to-be-doneness, etc., and e.g. badness, wrongness, ugliness, sleaziness, cruelty, unreasonableness, etc. For present purposes, we should be as permissive as possible about which properties count as positive or negative normative or evaluative properties. In fact, the only restriction we should impose is that the properties appealed to should allow evaluative content views to meet our third desiderata from above. That is, we should be able to see, upon reflection, how someone's representing something as having the property in question would as such “count” as a way of favoring or disfavoring that thing. As long as the relevant property meets this criterion, we should happily count it as a positive or negative normative or evaluative property in what follows. (But I will typically just focus on goodness and badness below for the sake of simplicity.)

positive orientation towards going to the beach in virtue of the fact it involves my accepting that going to the beach is good, say. Right away, we can see that this promises to satisfy our third criterion. To accept that something is good, say, *does* seem to be a way to favor something. Alternatively, things get *much* murkier if we opt for (a) some other kind of contents (what *other* kinds of contents could there be such that my accepting a content of that kind amounts to my being for or against something?<sup>71</sup>) or (b) some other kind of attitude (what kind of attitude, other than acceptance, could I have towards the proposition that X is good that could be such that my having that kind of attitude towards that content is a way for me to be for X?<sup>72</sup>). For these reasons, evaluative content views of phenomenal valence are worth taking especially seriously.

In Section 2 I will raise some preliminary challenges for evaluative content views. I have two goals here: first, to show that evaluative content views shoulder the initial argumentative burden vis-à-vis (phenomenal) attitude views; and second, to have us home in on the strongest version of evaluative content views. The view that emerges has affective experiences coming out as analogous to *perceptions* of values.<sup>73</sup> Then, in Sections 3 through 5, I consider and respond to three arguments for thinking that we can account for a number of

---

<sup>71</sup> Suppose, for example, that I accept the proposition *that x is likely to harm me*. Does my representing x as a likely cause of *harm* to me amount to my being in some way negatively oriented towards x? What if I don't care about future harms? Or suppose I accept the proposition *that I promised to pick Jane up from the airport today*. Does my representing that action as something that I *promised* to do amount to my being in some way positively oriented towards doing it? What if I'm a utilitarian who sees promises only as contingently binding? Or suppose I accept the proposition *that I am motivated to Φ*. Does my acceptance of *that* proposition amount to my being in some way positively oriented towards Φ-ing? What if that proposition is false? About all of these cases I am inclined to shrug my shoulders and plead uncertainty. These considerations show, I think, that *evaluative* content views in particular are worth taking especially seriously.

<sup>72</sup> The positive phenomenal valence of my positive feeling towards X *cannot* plausibly consist of my having attitudes towards X's value like the following, for example: *wondering whether* X is good, *hoping* that X is good, *supposing* that X is good, etc. None of these amounts to ways of being positively oriented towards X. But attitudes of acceptance are not as such valenced: as discussed in Chapter 1, one can accept that the coffee table is brown while being entirely indifferent to that fact.

<sup>73</sup> Whether evaluative content views should understand affective experiences as *literally* a *kind* of perception is not a question that we will need to answer here, as I hope becomes clear, at least in passing, below.

important relations between affective experiences, on the one hand, and *values* or *evaluative beliefs* on the other, only if we accept that affective experiences, one and all, represent values. These arguments, respectively, concern the correctness of affective experiences (Section 3), the fact that affective experiences “enrich” our evaluative thought (Section 4), and the fact that affective experiences can justify evaluative beliefs (Section 5).

One way to respond to such arguments would be to deny that the relevant relations obtain. I will not take that tack, since I am inclined to think that these relations do obtain. Another way to respond to such arguments would be to point out that, really, all they show (if successful) is that affective experiences, one and all, represent values: they do not show *in addition* that phenomenal valence can be *explained* in terms of such representations of value. I will not take this tack either. If affective experiences, one and all, are representations of value, then – given how well placed evaluative representations seem to be to meet our third criterion from above – we might as well embrace an evaluative content view. This is why I will regard the arguments of Sections 3 – 5 as, in effect, arguments for evaluative content views.

Instead, I will respond to those arguments by showing that, in each case, we can account for the relevant phenomena without embracing an evaluative content view (or, really, *any* kind of content view). The resources necessary to account for the relevant phenomena are available to us all, even to adherents of attitude views. So none of the relevant arguments support evaluative content views over any other theory of affective experience or phenomenal valence. But we should not come away from the discussions of this chapter thinking that evaluative content views are *just as* plausible as attitude views, that we have arrived at a stalemate. I will point out a number of ways in which evaluative

content views seem to give deeply *wrong* accounts of relevant phenomena. These points, combined with the points made at the end of Chapter 2 and in Section 2 of this chapter should lead us, I think, to give up on content views of phenomenal valence altogether. We should choose among attitude views, and the next chapter will help us do that. At any rate, we will see here, I think, that affective experiences do not represent values. As Mark Johnston (2001: p. 189) antagonistically but ably puts the idea: “affect is never the disclosure or sensory presentation of the appealing.” Not only are evaluative content views not the *entire* story of phenomenal valence, they are no part of the story at all.

## **Section 2: Preliminary Challenges to Evaluative Content Views**

Think about a feeling of gratitude towards your co-worker for the lovely birthday gift they gave you. Doesn't the thoughtfulness of that gesture (or the gift itself, or your co-worker herself) *seem good* to you? And consider your feeling of anger at the careless way in which the driver of that car drove through that puddle and splashed that pedestrian (he saw the puddle coming, and he didn't need to drive in the right lane!). Doesn't the driver's behavior here *seem bad* to you? In these cases, and many others, evaluative content views seem to fit very well with our actual, lived affective experience. But in the rest of this section I will present reasons to think that such views are actually quite counterintuitive. As we will see, in response to some preliminary objections, evaluative content views should take a particular “shape.” But once they have taken that shape, then any initial intuitive support that such views might have had evaporates. If we are to continue to take such views seriously, their defenders need to marshal arguments in their favor. In the absence of such successful arguments, we should reject evaluative content views of phenomenal valence.

The first preliminary objection is based primarily on a particular account of how we should understand content, the attitudes of acceptance that, as we have seen, evaluative content views must invoke, and a plausible assumption about the ubiquity of affect:

- (1) To accept that such-and-such is good or bad is to *believe* that such-and-such is good or bad.
- (2) To believe that such-and-such is good or bad is to represent that such-and-such is good or bad by way of evaluative *concepts*.
- (3) Non-human animals and very young human children cannot represent that such-and-such is good or bad by way of evaluative concepts.
- (4) Many non-human animals and very young human children can feel positively or negatively about things.
- (5) So, feeling positively or negatively about such-and-such is not a matter of accepting that such-and-such is good or bad.

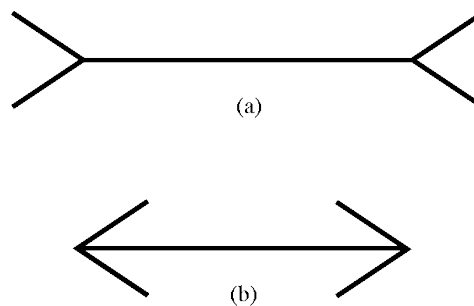
It would take significant work to make this objection fully cogent, and for reasons of space we will not get too deep into the philosophical-cum-empirical weeds with respect to the nature of beliefs and concepts and the mental capacities of infants and animals here.

Nevertheless, (2) seems almost definitional (of beliefs, concepts, or both) and (4) seems incredibly plausible. Even if one is inclined to deny that e.g. non-human animals can have (any or many) full-blooded *emotions*, it would be incredibly surprising if non-human animals were not capable e.g. of feeling *pleased* about this sensation, of feeling *attracted* to drinking that stuff, etc. And (3) would, I expect, at least secure a great deal of support in the philosophical community.

Although a defender of an evaluative content view might want to push on one or more of (2) – (4) to avoid commitment to (5), it seems much less risky for them to reject (1) instead. Luckily for defenders of such views, there seem to be very good independent



reasons to reject (1).<sup>74</sup> A common idea in the *emotions* literature in the past few decades has been that emotions represent things as possessing normative or evaluative properties, but not via belief. Instead, they represent such things by way of evaluative “perceptions,” or “construals,” or “non-conceptual representations,” or “quasijudgments.”<sup>75</sup> Perhaps adherents of evaluative content views could take some of these ideas on board for feelings as such.<sup>76</sup> The basic idea here is a fairly simple one that can be brought out by an analogy with visual perception and what we might call vision-based beliefs. Consider the Müller-Lyer lines:



The horizontal line in (a) looks longer than the horizontal line in (b). But suppose you have learned to your satisfaction that those lines are in fact the same length. Then, typically, you will believe that they are the same length and you will no longer believe that they are different lengths. Nonetheless, when you look at them, (a) still looks longer than (b). In some sense, your visual experience represents – in a truth-apt way – that the lines are not the

<sup>74</sup> I have become convinced that something like (1) is actually true, despite the points to follow. Unfortunately, however, I cannot make that case here (primarily because I have views about evaluative judgment that make that a *very* difficult case to make). But I hope to return to it in future work.

<sup>75</sup> See e.g. Justin D’Arms & Daniel Jacobson (2003) for a helpful overview of the literature on this topic. For a helpful overview of recent discussions of non-conceptual content more generally, see e.g. York H. Gunther (2003), especially the “General Introduction.”

<sup>76</sup> This need not threaten e.g. “cognitivist” theories of the *emotions* (see e.g. Nussbaum (2004) and Solomon (2004)) according to which emotions are, at least in part, evaluative beliefs. Such views typically deny that emotions are, or are just, feelings (what I have been calling affective experiences). So even if affective experiences as such neither are nor entail evaluative beliefs, it might still be true – for all I say here – that emotions as such are or entail evaluative beliefs.

same length. What we see in such cases seems to suggest that to (seem to) see that *p* is not to believe that *p* – folk wisdom got this one wrong – but to (seem to) see that *p* is still to represent that *p*, i.e. visual experiences are non-belief, truth-apt representations.

The idea is that affective experiences are in these respects like our visual experiences. To feel attracted to something is not necessarily to believe that it is good, but it is to represent it as good – to *accept* that it is good – in a non-belief way. With this distinction, we can explain – in a principled and well-motivated way – how it is that feelings represent values without being or entailing evaluative beliefs, and so defuse the argument above.

It is difficult to know how to implement this strategy fully: are these non-belief representations different from beliefs in virtue of having different (e.g. nonconceptual) contents, or in virtue of being different in attitude (e.g. perhaps there is an element of endorsement present in belief but not in these other attitudes, etc.), or both?<sup>77</sup> For present purposes, we can allow evaluative content views to implement this strategy in a number of ways. The key idea, in any case, is that on this approach, the contents, or the attitudes, or both, that make up affective experiences are like (or analogous to) the contents or attitudes or both of perceptual experiences, at least in the respects recently discussed.

The next preliminary objection to evaluative content views that I would like us to consider is analogous to standard objections to “guise of the good” theories of *desire*. These theories try, as J. David Velleman (1992a: p. 6) puts it, to

incorporate the valence of desire into its content, by describing desire, not as a favorable attitude toward the representation of some outcome, but rather as an attitude toward a favorable representation of the outcome.

---

<sup>77</sup> Thanks are due to Sigrún Svavarsdóttir (p.c.) for encouraging me to think about these different possibilities.

The original versions of these objections appeal to cases where it seems as though someone desires that p, but where they do not represent p as good in any way. For analogous objections to apply to evaluative content views of affective experience and phenomenal valence, we would have to find cases where it seems as though someone *feels* positively towards X but where X does not seem good to them, or where someone feels negatively towards X but where X does not seem bad to them.

There do seem to be such cases. Consider Gary Watson's (1975: p. 210) case of "a woman who has a sudden urge to drown her bawling child in the bath" or Harry Frankfurt's (1971: p. 12) "unwilling addict, helplessly violated by his own desires." Or consider Ruth Chang's (2004: p. 66) very different case in which turning a cartwheel while walking down the sidewalk "simply appeals to" one. We seem to be able to imagine (versions of) each of these cases as cases where we feel positively about something or other (drowning our bawling child, doing heroin, turning a cartwheel) even though we do not represent it as good in any way. If you were to ask Frankfurt's or Watson's subjects about this, they would presumably tell you that *of course* doing heroin or drowning the child is not good. If you were to ask Chang's subject about this, she might well respond "I have no idea whether turning a cartwheel's good or not – those kinds of thoughts aren't even on my radar! – it simply appeals to me." And similar points apply in the other direction, too: consider feeling jealous of or spiteful towards someone who obviously, even by your own lights, does not deserve it.

One way to respond to such cases is precisely to appeal to the distinction between belief and non-belief evaluative representation made above. The characters just mentioned do not *believe* that this or that is good or bad, perhaps, but they might – given what we have seen above – still represent this or that as good or bad in a *non*-belief way. While this might

be a step in the right direction, I wonder how much ice is actually cut by this response. In the case of the Müller-Lyer lines, suppose that you ask me: I know that you don't believe that one of the lines is longer than the other, but doesn't one of the lines *seem* longer than the other? To this question I would answer "yes," without hesitation. But in the Watson, Frankfurt, and Chang-type cases, suppose that you ask me: I know that you don't believe that drowning your bawling baby/using heroin/turning a cartwheel is good, but doesn't it still *seem* good to you? To this question I would answer "no." Or, *at least*, a "yes" answer to this question is anything but the obvious response. So why *should* we accept that in all such cases we represent this or that as good or bad in a non-belief way?

Were you merely to insist, at this point, that I *must* represent the relevant thing as good in some way, given that I feel *positively* towards it, I would find your insistence idle.<sup>78</sup> The *only* sense in which Watson's mother, Frankfurt's addict, or Chang's cartwheeler seems to represent the relevant things as good is in what Justin D'Arms & Daniel Jacobson (2003: p. 136) call "the trivializing sense": they feel *attracted* to performing them, i.e. they have a *pro-*attitude towards them.<sup>79</sup> On the other hand, phenomenal attitude views, for example, need

---

<sup>78</sup> Ruth Chang (2004: p. 66) offers a similar response to an analogous move in debates about desires. Also, Peter Goldie (2000), a defender of an evaluative content view, seems actually to admit that such views approach idleness. He writes: "I appreciate that not everyone will be persuaded by my argument that the difference between thinking of X as Y with, and without, feeling is not just a difference in attitude but also a difference in content; it always seems possible for an opponent to force all the difference into the attitude, so that the debate degenerates into a matter of competing intuitions. For those who are unpersuaded, I would just point out that *the arguments of the rest of the chapter, including those concerning the intentionality of feelings, remain unaffected by this point.*" (2000: p. 60 – emphases added) The italicized bit is what is important here. If we can understand the "intentionality of feelings" without appeal to representations of value, then, given the plausibility of the remarks above and those to follow, we should.

<sup>79</sup> Philippa Foot's (1978) account of pride implies that one can only feel proud of something if one regards it as "in some way one's own." Against this, they (2003: pp. 135-136) discuss the case of the person who is proud of the achievements of their favorite football team: "What advance is really made by claiming that the fan construes the triumph as his own, or that he thinks of it that way? Surely he need not have any extravagant thoughts about his own role in the outcome. This is where some of the attraction of [Foot's kind of view] seems specious, inasmuch as it rests on an attribution that, in a pinch, can be made simply in virtue of the fact that someone has the relevant emotion...The sense in which the club's accomplishments belong to the fan is

not insist upon any such hidden, unobvious evaluative representations in these cases. They say that you can feel positively about something without representing it as good in *any* way. One might think, then, that phenomenal attitude views offer us a *more* promising way to account for the valence of the feelings at issue in these cases. But even if one does not want to go that far, it is at least fairly clear that intuition does not favor evaluative content views in such cases: it is either neutral/divided or antagonistic.

The final preliminary challenge for evaluative content views was discussed at length in the last chapter and I will not belabor the point again here. However evaluative content views understand what acceptance amounts to, and whatever the ultimate nature of the relevant evaluative contents is, it will still turn out that, according to evaluative content views, affective experiences are truth-apt mental states. And as we discussed in the last chapter, that seems deeply wrong. The distinction between belief and non-belief representation does not make this implication any less implausible.

### **Section 3: Correctness**

Evaluative content views, then, need additional argument. In the absence of such argument, given what we have seen so far, we should opt for a phenomenal attitude view. In this section and the two following we will consider arguments that attempt to fill this gap and to show that, really, evaluative content views are the way to go.

Our first such argument begins with the following thought: certain feelings are fitting or correct (or merited or appropriate) while others are not. It is fitting or correct to feel afraid of certain things, for example. Which things? The things that merit feelings of fear,

---

simply that he is able to be proud of them...Should [a defender of a Foot-type view] fall back on the claim that the fan feels as if the triumph were his own, we would suggest that the only sense in which this is true is the trivializing sense: he is proud of it.” The point that I am making here is analogous, as I imagine the reader can see.

e.g. things that are likely to harm you or yours. It is incorrect or unfitting to feel amused at certain things. Which things? The things that do not merit amusement, e.g. unfunny jokes or humorless anecdotes. And depending on your views of such things, it might also be fitting to feel sad about the fact that you will die someday (even though you cannot “do anything about it”), or fitting to feel pleased when justice is done. At any rate, you are making some kind of mistake or error when your feelings are unfitting responses to their intentional objects, or when you fail to have fitting feelings.

But this kind of mistake or error is special. For example, it need not be (but it can be) a *moral* failing, at least if e.g. Justin D’Arms & Daniel Jacobson (2000) are right that it can be unfitting or incorrect not to feel amusement at a funny but immoral joke.<sup>80</sup> And it need not be (but it can be) a *prudential* failing, at least if e.g. Wlodek Rabinowicz & Toni Rønnow-Rasmussen (2004) are right that it can be unfitting to feel admiration for someone who is not remotely admirable, even if a “demon” threatens to torture you unless you feel such admiration.

So why, then, are e.g. feelings of amusement fitting or correct responses to humorous jokes, but feelings of anger or pride are not as such fitting responses to those same jokes? Why are e.g. feelings of disgust fitting or correct responses to certain kinds of decay or contamination, but feelings of ecstasy or guilt are not as such fitting responses to those same kinds of decay or contamination? Here is a simple and seemingly powerful answer: feelings *represent* the presence of the relevant evaluative or normative properties (e.g. feelings of amusement represent humorousness, feelings of fear represent fearsomeness, etc.). And so feelings are correct or fitting, we might try saying, if and only if their

---

<sup>80</sup> Perhaps the “correctness conditions” for *some* feelings – e.g. guilt and anger (Gibbard (1990)) – are, however, essentially moral.

intentional objects actually instantiate the properties that the relevant feeling represents them as having. On this view – call it the “accuracy model of affective fit” – the correctness or fittingness of feelings is exactly analogous to the correctness of beliefs or perceptions. As with these states, fittingness or correctness of affect is a matter of accurate representation.

Here, then, is the Argument from Correctness:

- (1) Affective experiences are correct or fitting iff their intentional objects actually have<sup>81</sup> certain (dis)values (e.g. feeling admiration towards X is correct or fitting iff X is actually admirable, etc.).
- (2) Mental states are correct or fitting iff they are accurate truth- or veridicality-apt representations (e.g. a belief that a is F is correct or fitting iff a is F, etc.).
- (3) So, affective experiences are correct or fitting iff they are accurate truth- or veridicality-apt representations of things as having certain (dis)values (e.g. feeling admiration towards X is correct or fitting iff that feeling accurately represents X as being admirable, etc.).

There are at least two points in favor of this account. First, it appeals to the clearest understanding we have of what it is for a mental state to be “correct.” Alternatively, it is not at all clear, at least initially, what it would be for a mental state to be correct or incorrect if it is not an accurate or inaccurate representation of something or other. Second, this kind of explanation seems especially helpful here given (it seems) that having unfitting or incorrect feelings need not be a moral or prudential failing. Just as a belief or a perception can (in principle) be accurate, even if your having that belief or perception is morally problematic or

---

<sup>81</sup> But what about cases e.g. where I *believe* that I have just suffered a great loss (though I have not, in fact)? Wouldn't a feeling of sadness be fitting in such a case, given my belief? I am certainly not interested in denying this here. We could distinguish, say, between subjective fittingness (e.g. the case just mentioned) and objective fittingness (e.g. the case where you really have just suffered a great loss). With this distinction made, I will primarily focus on objective fittingness below. I will touch on issues related to subjective fittingness, however, below.

contrary to your interests<sup>82</sup>, so a feeling can (in principle) be fitting or correct even if your having that feeling is morally problematic or contrary to your interests.

Nevertheless, I will argue that the apparent promise of the accuracy model of affective fit is merely apparent. Premise (2) is false, and without (2), the accuracy model of affective fit has to fight its own battle. This is a battle it will ultimately lose, given the close analogies between feelings and motivational states that we will discuss below. However we go on to explain what affective fit amounts to, the accuracy model is not up to the task.

In general, what do we mean when we say that something – a mental state or anything else – is fitting or correct? For something to be fitting or correct is for it to meet some relevant standard, or to comply with some relevant requirement or demand. If this is right, then what accounts for the fact that beliefs, say, are correct or fitting iff true is the fact that truth is the (or a) relevant standard for beliefs. What accounts for the fact that perceptions, say, are correct iff veridical is the fact that veridicality is the (or a) relevant standard for perceptions. Maybe there is *logical* space for one to deny that beliefs or perceptions are correct iff they are accurate representations of what they are about. For example, maybe one thinks that the relevant standard for beliefs or perceptions is that they help their subject navigate their environment successfully, and one thinks that beliefs or perceptions could meet this standard without being accurate representations of what they are about. What accounts for this logical space, if it exists, would be the fact that accuracy of

---

<sup>82</sup> Examples of beliefs that are accurate, but not in one's interest, seem fairly easy to come by, at least if ethical egoism is false. For example, the belief *that other people matter as much as I do* might lead me to do things that go against my interest, but it still seems true for all that. It is more difficult, however, to think of cases of beliefs that, though immoral, are accurate. But think of a virulent racist's belief that a person *of race X* did a monstrous thing. People of any "race" can in fact do monstrous things, so such a belief might be true in a particular case, but maybe thinking about this particular monstrous thing *in this unjustifiably race-focused way* is morally problematic. I leave it to others to come up with perceptual examples of either kind.



representation might not, in principle, be the – or the only – relevant standard for beliefs or perceptions.

But I will simply accept that beliefs and perceptions are correct iff accurate, and I will not look further into the logical status of this fact. But if this general account of correctness, the one that says that for something to be fitting or correct is for it to meet some relevant standard, or to comply with some relevant requirement or demand, is on the right track, then we have no reason in advance to expect that things *other* than beliefs or perceptions will be correct iff accurate. There are many standards, requirements, and demands out there! Why would anyone think that correctness-determining standards, requirements, or demands only apply to truth-apt mental states, or that they only demand accuracy? They do not, and this will eventually undermine the Argument from Correctness, but I would like to work up to that slowly.

For now, consider a contractor who is attempting to build a building in accordance with blueprint B. Suppose that B requires the placement of a wall at position P1, but the contractor makes a mistake and puts a wall nearby, at P2, and puts no wall at P1. The placement of the wall is incorrect. Literally where the wall is is incorrect. It is supposed to be in another location, according to the standards, requirements, or demands set by B. The fact that the placement of the wall does not meet these standards, requirements, or demands is what makes its location incorrect. Further, the location of a wall – literally where a wall is – *represents nothing at all*. So, as we reasonably expected, correctness is not always, in general, a matter of accurate representation.

None of this on its own is a problem for premise (2) of the Argument from Correctness, of course – that premise only aspires to tell us about the fittingness or

correctness of mental states and says nothing about the locations of walls. Maybe it will turn out that the relevant standards, demands, or requirements that apply to mental states are all accuracy-involving. The next step on the way to seeing why that is unlikely has us consider the *actions* of our contractor, i.e. his building of the wall at P2, i.e. literally his *behaviors*. Or consider a dancer who is supposed to throw her arms out to the left at time t1 (according to the relevant standards set by the choreographer), but mistakenly throws her arms out to the right at t1, or an orchestra member who is supposed to play a middle C at t1 (according to the relevant standards set by the composer), but who mistakenly plays an E sharp at t1. In all of these cases, our characters' actions – literally, their behaviors, their physical movements – were incorrect. But in none of these cases do their behaviors, literally their physical movements, represent anything at all – at least not in a “mind-to-world” way. Also, the incorrectness of their behaviors does not seem to be reducible to the incorrectness of any of the mental states had by our characters. We can imagine our exasperated architect or impatient choreographer or frustrated composer saying “listen, it doesn’t matter what you *want* or what you *think* here, you need to put this wall at P1 (or: you’re in this piece and you need to throw your arms out to the left now, or: you need to play a middle C now)!”<sup>83</sup>

There is a marked disanalogy, of course, between the incorrectness that these cases involve and the incorrectness involved in e.g. believing a falsehood or feeling admiration towards someone who is not admirable. The cases just discussed each depend on one or another *contingent* standard or requirement or demand (i.e. a blueprint, a bit of choreography,

---

<sup>83</sup> And note further that in each of these cases, we can imagine that the incorrect behaviors need not have been a moral or prudential failing on the part of the relevant agent. Morally innocent mistakes are possible, as are ones that do not adversely affect your interests.

a musical composition).<sup>84</sup> Putting a wall at P2, throwing one's arms out to the right, and playing an E sharp are not, as we might say, incorrect "in and of themselves" or "without further ado," whereas believing a falsehood or feeling admiration towards someone who is not admirable *are* incorrect in and of themselves and without further ado.

The next step in the argument has us come closer to removing this disanalogy. Suppose that someone receives a great gift, a gift given respectfully and out of generosity, but the gift recipient acts – literally behaves, literally physically moves – in ways that display ingratitude, e.g. she shouts an obscenity at the gift-giver and storms off. Her actions – her behaviors – here are objectively, non-contingently incorrect, in and of themselves and without further ado. It is unfitting (unmerited, inappropriate) to behave in these ways towards a respectful, generous gift-giver (period, at least all else being equal).<sup>85</sup> But do these behaviors represent anything at all, e.g. that she has nothing for which to be grateful? I strongly doubt it.<sup>86</sup> If our contractor's, dancer's, or musician's behaviors represented nothing, then neither do our gift recipient's.

The obvious response here is to suggest that these ungrateful behaviors are incorrect or unfitting only insofar as they are (typically) the expressions of certain mental states (e.g. feelings of hostility, or of a lack of feelings of gratitude), and we have yet been given no

---

<sup>84</sup> I suspect that we can elaborate the cases in such a way that each involves the violation or flouting of a perfectly objective, non-contingent standard, and is rendered incorrect thereby. For instance, if placing the wall at P1, or throwing one's arms out to the left, or playing an E sharp violated objective, non-contingent *aesthetic* standards, then, in virtue of so doing, such behaviors would thereby be (non-contingently) incorrect.

<sup>85</sup> More complicated cases could arise, of course. For example, if this "gift" had been given by an obsessive stalker or by the man who killed her family, she could be entirely justified in her behavioral reactions. But these are not the kinds of cases I have in mind here.

<sup>86</sup> Here it is important to keep in mind that when we talk about X's representing Y, we mean that Y is an intentional object of X. There is another sense of representation, e.g. the sense in which a pollster might regard my views as representing, or as being representative of, the views of thirty-something philosophers. In this looser sense, our character's actions might well "represent" that she has nothing to be grateful for. But that is not the sense at issue.

reason to think that *those* mental states can be correct or incorrect without being truth-apt representations. Nevertheless, I believe that these behaviors could be incorrect regardless of any mental states that our character had, just as the behaviors of our contractor, dancer, or musician could have been. To bolster this point, compare the following two characters who each receive a great gift. A fails to feel grateful *and* she behaves in ungrateful ways. B fails to feel grateful, too, but she behaves in grateful ways. It seems to me that A is making two mistakes: first, her feelings are unfitting; second, her behaviors are unfitting. B, on the other hand, seems only to be making one mistake: at least she “gets it right” behaviorally (even though, of course, there are problems with being disingenuous). Just imagine the case as one where she has to summon the willpower to act in the correct ways, despite the way she feels. And this suggests that such behavioral errors or successes are not reducible to mental errors or successes. So, I think, we have also learned that certain behaviors can be incorrect or unfitting, in and of themselves and without further ado, and that their incorrectness or unfittingness cannot be explained in terms of accurate or inaccurate representation.

Finally, we are in a position to see why even the correctness or fittingness of mental states is not necessarily a matter of representational accuracy. Instead of thinking just about our “behaviorally ungrateful” character’s behaviors, think of the last mental state or states in the chain that led to these behaviors. These would presumably be states of (or events in) what is often called the motor control system. And we can coherently imagine that such states would be imperatival in form (e.g. like intentions or desires, maybe, or like Hulse et al.’s (2004) “inner imperatives”<sup>87</sup>). For example, these might be “instructions” like “yell

---

<sup>87</sup> Here is how Hulse et al. describe these inner imperatives: “An utterance of “Get that chocolate!” plausibly has the desire-like direction of fit, even if it is an *inner* utterance (i.e., an exercise of one’s capacity for auditory imagination used to create an image of uttering that sentence). After all, the inner utterance is in the form of an

such-and-such obscenity now!”, “turn around and walk quickly out of the room!”, and so on. And I would like to say two things about such states in such cases. First, they too would be incorrect in and of themselves and without further ado, for precisely the same (non-contingent) reasons for which the behaviors in which they issue are incorrect.<sup>88</sup> Second, they are not truth-apt, accurate or inaccurate, representations of anything. So, (2) is false and the defender of the accuracy model of affective fit has to fight her own battle.<sup>89</sup>

But maybe this battle will not be too hard to win. Even if (2) is false, we have granted that at least some mental states (e.g. beliefs and perceptions) *are* correct iff accurate. So why think that affective (in)correctness is *more* like the (in)correctness of behaviors and “inner imperatives” than it is like perceptual or doxastic (in)correctness? In the rest of this section I will begin to offer an answer to this question.

What kind of failing do we think it is, really, for a person to have unfitting feelings about things or to fail to have fitting feelings about things? According to evaluative content views, it is fundamentally a *perceptual* failing (or something closely analogous to it). But that seems deeply wrong. The problem with a person’s having unfitting feelings is not that such a person sees what is not there or fails to see what is there. Rather, it is for such a person to be *oriented* incorrectly towards the world or some piece of it. As Sigrún Svavarsdóttir (2014: p. 89) puts a related point, when we criticize someone for having unfitting attitudes, “the

---

imperative, and imperatives, desires, intentions and the like all share the same direction of fit.” (2004: p. 75 – emphases added)

<sup>88</sup> My opponent suggested that behavioral incorrectness should be explained by mental incorrectness, i.e. he wanted to work “from the inside out.” But we seem to have learned that, sometimes, mental incorrectness is explained by behavioral incorrectness, i.e. sometimes things work “from the outside in.”

<sup>89</sup> Further, these states might be incorrect even in the absence of any moral or prudential failing on our character’s part. If the case of behavioral ingratitude seems too moralistic to you, even given our discussion above, then think of analogous cases involving “behavioral amusement” or “behavioral admiration.” I doubt that it needs to be a moral or prudential failing to fail to behave in ways that conventionally display amusement towards something humorous, or to fail to behave in ways that conventionally display admiration towards something admirable.

alleged mistake is not that of misrepresentation but, rather, that of misplacement of emotional and motivational energies.”

But maybe this is a bit too quick. Nomy Arpaly & Tim Schroeder (2014: pp. 120-121) draw our attention to the case of Heidi:

A drunken Heidi can find a mishap with a carafe of wine enjoyably amusing, while a more sober onlooker asks “don’t you see that you’ve just ruined your couch?” Heidi can comprehend it but not feel it, sometimes. She can intellectually grasp that something very undesired has happened, but feel as though everything is delightful. In this situation, people are often inclined to say of the drunken woman that she is subject to a sort of error: that she “doesn’t see it,” “doesn’t get it,” “can’t take it in,” and so on.<sup>90</sup>

In this case, we might think that Heidi’s feelings are incorrect, but in what ways are her emotional or motivational energies misplaced? Her amusement does not seem immoral or vicious or crass or banal or ugly, for example, and whether it fails to be in her best interest is anyone’s guess. She does seem to be making some kind of mistake, but we seem to have run out of non-accuracy-involving options for explaining what kind of mistake it is.

But this case is misleading. What is so compelling about the case of Heidi is that her amusement is so benign, and anyway, the potential problem is *not*, it seems to me, that she feels amused. There is something absurd about spilling wine on a new, expensive couch (as there is something absurd about a man’s pants falling down while he is being knighted (as Thomas Nagel (1971) points out) and absurdity can certainly merit amusement. There would be a problem, however, were her amusement entirely unambivalent, experienced without a tinge of irony, utterly without a “hard edge.” If her present experience does not also include feelings of exasperation or disappointment or annoyance, then she is getting

---

<sup>90</sup> Arpaly & Schroeder are not themselves committed to an evaluative content view of phenomenal valence. We will briefly consider a different view of phenomenal valence that is inspired by some of their work in the next chapter.

things wrong here (though in an enjoyable way). But failing to feel e.g. disappointed about ruining a cherished possession seems like a paradigmatically practical – or at any rate, non-representational – problem: she is being careless about something that she cares about. The norm she is flouting is something like: feel disappointment when one of your cherished possessions is ruined. And I have a hard time taking seriously the idea that that is an instance of the more general (epistemic?) requirement to represent things accurately. The question – what kind of mistake is Heidi making? – is a good one, but the answer – it is akin to a perceptual error – is, in any case, a bad one.

But maybe you remain unconvinced. First let us take a step back: all can agree that feelings of fear are (only) correct when directed at something that's fearsome, that feelings of admiration are (only) correct when directed at someone who's admirable, etc. Let us call these values or disvalues the “correctness conditions” for feelings. All can agree that feelings have correctness conditions, and that values or disvalues constitute these correct conditions, *whatever* one's view of affective experience and phenomenal valence. That is not in dispute. As discussed above, and in Chapter 1, however, these are not *only* the correctness conditions for feelings, but they are also the correctness conditions for valenced states of other kinds: e.g., motivations and evaluative beliefs.

On the one hand, as was hinted at above in our discussion of “behavioral ingratitude,” an accuracy model of the correctness of *motivations* is an utter non-starter: motivations are not truth- or veridicality-apt states, and so the accuracy model *cannot* apply to them. So, there is a large class of valenced states whose fittingness or correctness *cannot* be explained along the lines of the accuracy model. But one might think, on the other hand, that an accuracy model of the correctness of *evaluative beliefs* is very plausible: after all, an

evaluative belief is correct iff true. In the rest of this section I will argue, first, that that the problem with incorrect affect is more like the problem with incorrect motivation than it is like the problem with false evaluative belief. Second, I will argue that even if that were not so, the accuracy model goes wrong *even in the case of evaluative beliefs*. So, however we slice it, we should reject the accuracy model of affective fit (and really of “valenced fit” altogether), and so the argument considered at the beginning of this section gives us no good reason to embrace an evaluative content view of affective experience.

To make the first case, let us begin by focusing on what we might call *recalcitrant affect*, i.e. cases where our feelings and our evaluative judgments conflict, e.g. cases where one feels fear towards something that one believes is not dangerous (or fails to feel fear towards something that one believes is dangerous, etc.). In what sense do our feelings and our evaluative judgments *conflict* in such cases? What’s the *problem* with having recalcitrant affect? Recalcitrant affect does seem problematic, and one might think that evaluative content views have a leg up here, since they can explain the problem with recalcitrant affect as an instance of the more general problem of having truth-apt mental states with inconsistent contents, e.g. believing that p and believing that not p.<sup>91</sup> My feeling of fear represents X as dangerous, but my belief represents that X is not dangerous, the story goes, and that is problematic.

But phenomenal attitude views are much better placed to capture the relevant phenomena *in the right way*. First, for reasons considered in Section 2 above, an evaluative content view should probably follow e.g. Michael Brady (2007) and Christine Tappolet’s (2010) lead in thinking that recalcitrant affect is analogous to e.g. optical illusions, like the Müller-Lyer lines, i.e. cases in which our perceptions conflict with our beliefs. But, as

---

<sup>91</sup> This idea is inspired by Christine Tappolet’s (2010, etc.) work on recalcitrant *emotion*. Her views will be discussed briefly below.



Bennett Helm (2001) points out, this *kind* of account seems to miss something: there is nothing remotely *irrational* about such conflicts between perception and belief, but there does seem to be something rationally troubling about recalcitrant affect.<sup>92</sup> Even if we should be reluctant to say that someone is culpably irrational in virtue of having some recalcitrant feeling or other – perhaps because our feelings are not sufficiently in our control for such a judgment to be fully fair – there certainly seems to be something in the ballpark that Brady-or-Tappolet-style views miss.

So far, this is all pretty well-tilled ground. Brady and Tappolet have each attempted to respond to these kinds of objections, for example.<sup>93</sup> But what I want to suggest at this point is that we already have a perfectly adequate model for understanding recalcitrant affect, a model that does not do evaluative content views any favors. A model for understanding the problem with recalcitrant affect that is *much* better than the optical illusions model (however we flesh it out further) is the model provided by akrasia, or what we might call *recalcitrant motivation*, i.e. cases where you believe that you should  $\Phi$  but you fail to  $\Phi$  (or fail to intend to  $\Phi$ , or fail to be motivated to  $\Phi$ ), despite believing that you can  $\Phi$ . As Timothy Scanlon says (1998) says, one is most clearly irrational when one is akratic.<sup>94</sup> And on the

---

<sup>92</sup> Brady (2007: p. 275) nicely summarizes Helm's idea: "Recalcitrant emotions, and in particular phobic reactions, are taken to be paradigmatic cases of psychological *disorder*; we are *bothered* by recalcitrant emotions, we find the clash between emotional experience and evaluative judgement *disturbing*...For Helm, the fact that we are bothered and disturbed by recalcitrant emotions indicates that such emotions involve *rational* conflict with evaluative judgements."

<sup>93</sup> Tappolet (2012) attempts to meet an analogous challenge by noting that our emotion *systems* are actually fairly amenable to top-down control (over time), and so certain feelings might be unjustified or irrational if they indicate that we have not exercised the right kinds of top-down control. Brady notes that recalcitrant feelings amount to a waste of cognitive, attentional, and motivational resources by our own lights, and that they "incline the subject to accept an evaluative construal that the subject has already rejected." (2009: p. 413) Each of these kinds of ideas might be part of the correct story, but – interestingly – neither *really* depends on the truth of an evaluative content view. I hope to develop this further in future work.

<sup>94</sup> He writes: "Irrationality in the clearest sense occurs when a person's attitudes fail to conform to his or her own judgments: when, for example...a person fails to form and act on an intention to do something even though he or she judges there to be overwhelmingly good reason to do it. These are clear cases of irrationality

other hand, though there are complications here, one is at least getting something right, rationally speaking, when one's actions, intentions, and motivations accord with one's normative or evaluative beliefs.<sup>95</sup> And as we have already seen actions, intentions, and motivations are not truth-apt representations (of values or oughts or anything else). So however we explain the problem with akrasia, it *cannot* be as an instance of (any version of) the general problem of having truth-apt states with inconsistent contents.

The problem with recalcitrant affect is remarkably similar to the problem with akrasia, so insofar as we want to explain why recalcitrant affect is irrational, we should explain it in the same way we explain why akrasia is irrational (namely: not along accuracy model lines). But in *what way or ways* is the problem with recalcitrant affect remarkably similar to the problem with akrasia?<sup>96</sup> If I had a fully satisfactory answer to this question, this would no doubt be a much shorter (and much stronger) dissertation. The best I can do, however, is encourage the reader to reflect on her own experiences of recalcitrance of each kind, and the kinds of annoyance or exasperation or shame to which they each give rise, and to reflect on the kinds of criticism you are inclined to make of others when they are being recalcitrant in each way. You think you should stay away from that handsome charlatan, but there you go again, feeling incredibly attracted to seeing him again and then heading out the door to meet him for a drink. From the first-person perspective anyway, any distinctions between affect and motivation seem entirely artificial in such cases: my self-criticism of my

---

because the thought or action they involve is in an obvious sense “contrary to (the person’s own) reason”: there is a direct clash between the judgments a person makes and the judgments required by the attitudes he or she holds.” (1998: p. 25)

<sup>95</sup> The complicating cases are ones, e.g. like the case of Huckleberry Finn discussed by Jonathan Bennett (1974), Alison McIntyre (1990), and Nomy Arpaly (2000), where it looks like our judgments should actually be revised to accord with our motivations or feelings. But these complications will have to be addressed another day.

<sup>96</sup> I would like to thank Sigrún Svavarsdóttir, Justin D’Arms, and Declan Smithies for each, independently, encouraging me to try to answer this question.

recalcitrant feelings is *just exactly the same kind of thing* as my self-criticism of my recalcitrant motives. We can summarize this by saying that such self-criticisms involve attributions of irrationality in each case, but that is woefully inadequate to the *utter* similarity of these criticisms, and to the yawning chasm between them and the criticisms that seem appropriate when we encounter splits between perception and belief. This might just be table-pounding on my part, but I hope that it at least puts the reader into the frame of mind where she can see, at least a bit more clearly, the attractions of the akrasia model (as opposed to the accuracy model) of the problem with recalcitrant affect. In this case, again, it seems to me, evaluative content views explain a relevant normative phenomenon in the wrong way because they focus on the wrong things.

If this is right, then we have now seen a major respect in which feeling is more like motivation than it is like belief or perception. And, although it is difficult to articulate the next point clearly, this seems to me to suggest that affective *correctness*, too, is more like motivational correctness than it is like the correctness of evaluative belief. Return to Heidi. Suppose that Heidi's feeling of amusement about the wine spilling on her couch was actually unfitting. Now let us revise the case so that this feeling of amusement is *also* recalcitrant: she knows full well, even in her drunken state, that the spilled wine is not amusing, given that her couch is ruined. If the points above are correct, then the problem with the *recalcitrance* of her feeling cannot be explained along the lines of the accuracy model. And here is the key, but difficult to articulate point: however we understand the problem with having *recalcitrant* feelings, and however we understand the problem with having *unfitting* feelings, it seems overwhelmingly plausible that we should give the same *kind* of account of each, especially in cases like our revised Heidi case. I just cannot wrap my head around the idea that her feeling

is *unfitting* because it misrepresents the facts, but that her feeling is *recalcitrant* because it violates some practical, non-representational requirement that enjoins coherence among feelings and evaluative beliefs. Surely the problems with each are not so distinct as that! So, I claim, we should either explain both affective fit and affective recalcitrance via the accuracy model, or we should not explain either via the accuracy model. We definitely should not explain affective *recalcitrance* via the accuracy model. So, we should not explain affective *fit* via the accuracy model, either.

But suppose you are still tempted by what seems to be the success of the accuracy model in accounting for the fit or correctness of *evaluative beliefs*. Now I will argue that the accuracy model falls just as flat in the case of evaluative beliefs as it does in the case of incorrect or recalcitrant motivation. Compare the following two beliefs: (a) the belief that grass is red and (b) the belief that women are inferior to men. Both beliefs are false, and so unfitting or incorrect in that sense, but (b) is problematic in an *additional* way. It is a way of being *against* women (at least comparatively) – or at least it had better be, if evaluative content views have any hope of giving a satisfactory account of phenomenal *valence*! – and that additional error is where its unfittingness *qua valenced state* lies. But this additional problem does not seem to be an additional representational error. As with having unfitting motivations or (I have suggested) as with having unfitting feelings, it is a matter of one's being incorrectly oriented towards things, and not a matter of one's having inaccurate representations.

#### **Section 4: Enrichment**

The next argument that we will consider comes from Mark Johnston (2001). This argument begins with the following thought:

Seeing the utterly specific ways in which a situation, animal or person is appealing or repellent requires an appropriate affective engagement with the situation, animal or person. Absence of appropriate affect makes us aspect blind. The world then appears more neutral than it is, and our immediate evaluational thought and judgement becomes impoverished. (Johnston (2001: p. 181))

What I am interested in here is this idea that, without affect, our “evaluational thought and judgment becomes impoverished.” More specifically, I am interested in one of the arguments that we can extract from the following passage:

My focus will be on a certain class of values, which do not have good names in the languages I know. I mean the utterly determinate versions of such determinables as the beautiful, the charming, the erotic (in the narrower sense), the banal, the sublime, the horrific and the plain old appealing and the repellent... Within each determinable range, the determinate values in question would be inaccessible to beings without an appropriate sensibility. So these values might be called the inherently sensuous values. Thought and judgement directed at these determinate values could not be generated simply by the understanding. Something akin to sensing and sense-based imagination is required to make them available as topics for thought and judgement. While reason can include in its accounting judgements directed at such values, it cannot deliver the judgements themselves. Just as we need to sense cherry red to make a goodish range of judgements as to its nature, we need to encounter the determinate sensuous values in order to have them either as the topics or as the things predicated in our most basic evaluative judgements... If one has never been moved or affected by the determinate ways in which things are beautiful or charming or erotic or banal or sublime or horrific or appealing, then one is ignorant of the relevant determinate values. (2001: pp. 182-183)

There are at least two ways of understanding what Johnston's up to here. On one reading, call it the “epistemic reading,” Johnston's argument relies on the assumption that *there are* a number of “utterly specific” values out in the world and, then, the idea is that in order for us to “see” them, we need to be appropriately affectively engaged with things. On the other reading, call it the “psychological reading,” Johnston's argument does not depend on the realist assumption just made. The point here would be that in order for us to be able to *think* about, or make *judgments* about, utterly specific values, we need to be appropriately affectively engaged with things.

As a matter of fact, it is pretty clear that Johnston aims to defend both of these theses: after all, he spends the majority of the relevant paper arguing for a “detectivist” account of the relations between sensuous values and our feelings and judgments about them, and against “projectivist” accounts of these relations. But I will focus only on the psychological reading. It is less controversial, and we will address many of the issues raised by the epistemic reading in the next section anyway.

Put another way, I am only interested in the argument that we find in this passage that can be interpreted as follows:

- (1) Affective experiences directly enable enriched (non-impooverished) thoughts and judgments about a range of sensuous values.
- (2) If affective experiences directly enable enriched (non-impooverished) thoughts and judgments about sensuous values, then affective experiences are perceptions (or are some other kind of truth-apt representations) of sensuous values.
- (3) So, affective experiences are perceptions (or are some other kind of truth-apt representation) of those values.

We can call this the Argument from Enrichment.

So why should we believe (1)? Imagine someone who was born incapable of feeling anything approaching disgust. Through some biological quirks, he could never, and can never, feel disgusted. Maybe there are smells or tastes or scenes that he does not like, but they do not “gross him out” like a visibly maggot-infested animal carcass might do to us. Perhaps the closest he can get is a feeling of annoyance. Now suppose you are walking down the street with this strange fellow and you see (and smell) such an animal carcass on the side of the road. You feel disgusted by it, to the point of nausea, even, and you say, “That’s disgusting.” Johnston’s point here, which I think is correct, is that there is a real

sense in which your friend would have no idea what you were talking about. There is a sense in which he does not understand what it means for something to be disgusting.

Sure, he could consult a dictionary, or go to an evolutionary biology lecture, or spend hours on YouTube, and learn which kinds of things people typically call ‘disgusting,’ and even learn that they are often the kinds of things that can cause human beings to get sick if ingested, etc. So in an attenuated, derivative sense, he could understand what ‘disgusting’ means. But there is another, very important sense, in which he is still “aspect blind.” This is the distinction that I suspect Johnston is making in the passage above when he writes “while reason can include in its accounting judgements directed at such values, it cannot deliver the judgements themselves.” On the other hand, there is nothing missing from *your* ability to think and talk about the disgustingness of things. And since we can assume that the only significant difference between you and your friend is that you can feel disgust while he cannot, that seems to be what accounts for this cognitive difference. Maybe this is incorrect, but I am happy to grant that (1) is true, in something like the sense just described.

But why should we believe that the truth of (1) supports (3), via (2) or in any other way? Here is where Johnston’s appeal to the case of visual experience and thoughts/judgments about color comes in. It is at least fairly plausible to think that unless one is capable of having visual experiences as of color, one’s thoughts or judgments about colors will also be “impoverished” in much the same way that your friend’s thoughts or judgments about disgustingness were impoverished. But then we can ask: how is it that experiences of color enrich our color thoughts/judgments? A very natural answer is that they do so by giving us direct, conscious access to the very properties that the relevant thoughts or judgments are thoughts or judgments about, by representing (or maybe by

“presenting,” to use e.g. John Bengson’s (2015) terminology) those very properties (in a mind-to-world way). If this is how it works in the case of vision and color thoughts/judgments, then maybe it works in a similar way in the case of affect and thoughts/judgments about sensuous values: affect enriches our thoughts/judgments about sensuous values by giving us direct, conscious access to the very values that those thoughts/judgments are about, by representing those very properties. More concisely, affective experiences are perceptions of value.

I accept, at least for the sake of argument, that these claims about how (common-or-garden) perceptual experiences enrich perceptual beliefs is on the right track. I am also prepared to grant that we should use this perceptual model to explain how affective experiences enrich evaluative beliefs as long as there is no alternative story to tell that is at least as plausible. But if there are such alternative stories in the affective case, then we should, at best, withhold judgment about the application of the perceptual model and so we should, at best, withhold judgment about whether (1) supports (3), via (2) or in any other way.

The Argument from Enrichment, then, puts the burden squarely on the shoulders of those who would reject an evaluative content view of affective experience. At the very end of the day, whether this burden can be met will end up turning, I think, on how some foundational debates in metaethics turn out, particularly debates concerning the nature of normative or evaluative thought or judgment. But in lieu of engaging in those debates



here<sup>97</sup>, I will simply present an attitude view-friendly alternative to (2) that, it seems to me, is at least as plausible as (2) is.

First, as we have already seen, there are a number of important systematic connections between affective experiences and values and disvalues. Namely, values and disvalues constitute the correctness conditions for feelings. With this in mind, one plausible way to think about evaluative *concepts* – especially the ones that Johnston focuses on here, e.g. SUBLIME, DISGUSTING, HORRIFIC, BANAL, etc. – and so one plausible way to think about evaluative thoughts or judgments is offered by fitting attitude accounts. According to one way of fleshing out such accounts, for someone to sincerely and competently have the thought e.g. that is conventionally expressed by the English sentence ‘that is disgusting,’ one must somehow understand (at least implicitly) the points made above, e.g. that what it would be for something to be disgusting is for that thing to *merit* disgust, or for disgust towards that thing to be *fitting* or *correct* or *appropriate*. The idea is that you do not count as a fully competent user of the concept DISGUSTINGNESS, for example, unless you understand, at least implicitly, that for something to be disgusting is for it to be a fitting (correct, merited, appropriate) object of disgust. Maybe you could still make disgustingness judgments without grasping these facts, but at best, such judgments would be impoverished.

So, to make appropriately enriched evaluative judgments, you have to have appropriately enriched conceptions of two things: fit (merit, correctness, appropriateness) and the relevant response (disgust, fear, boredom, anger, delight, etc.). Since we are

---

<sup>97</sup> According to a view of normative and evaluative judgment that I find appealing, you *cannot* make sincere and competent normative or evaluative judgments without being disposed to have relevant affective experiences. Part of the *nature* of such judgments is, then, affective. For example, perhaps part of *what it is* to think that X is disgusting is to be disposed to feel disgust towards X. If this kind of account is correct, then *of course* your friend could not make non-impoverished disgustingness judgments. Unfortunately, I cannot argue for this view here, and so I simply mention it as a possibility. But it will come up again, briefly, in Chapter 6.

assuming that (1) is true, we now must ask: does an absence of affect impoverish our conception of fit, or does an absence of affect impoverish our conception of the relevant responses, or both? I think it is clear that an absence of affect *at least* impoverishes our conception of the relevant responses. We cannot, we might try saying, have an appropriately rich (i.e. non-impoverished) conception of disgust, fear, etc. without actually feeling these ways sometimes, or at least having felt these ways in the past. If you tell your friend from above that you are disgusted by this rotting animal carcass, here too, it seems, he would not really understand what you were talking about. He could at best have an attenuated, impoverished conception of what it is to feel disgusted by something. So much should be common ground among all of us who find anything like premise (1) plausible.

But if this is right, then we are now in a position to offer an attitude view-friendly alternative to (2). Affective experience enables enriched evaluative thought and judgment by giving us an appropriately rich understanding of the responses that the relevant objects make fitting (correct, etc.). Conversely, an absence of affective experience impoverishes evaluative thought and judgment by limiting our ability to understand those responses. And – here is the key point – we can accept this without believing that affective experiences are perceptions (or other kinds of truth-apt representations) of values (or of anything else). The idea here is not that feelings of disgust would enrich your friend’s disgustingness judgments by giving him direct, conscious access to the properties that such judgments are about, i.e. in the way (we are assuming) that color experiences enrich one’s color thoughts and judgments. Rather, feelings of disgust would enrich your friend’s disgustingness judgments by giving him an enriched understanding of what disgust is. Further, this could be so even if feelings of disgust do not *represent* anything about disgust. Sometimes we learn about something by

experiencing it, but not by representing it – e.g. we acquire an enriched understanding of what pleasure is by experiencing episodes of pleasure, but not because a particular episode of pleasure represents anything *about pleasure*! That, the idea goes, is how it works in this case, too.

Given these two options – Johnston’s proposal and the proposal just made – the proposal just made seems preferable. Why does your friend not really understand what it is to be disgusting? Precisely because he does not really understand what it is to be disgusted. It seems uncontroversial that feelings of e.g. disgust, etc. “enrich” our understanding of e.g. disgust. And if the points made above are correct, then the fact that e.g. feelings of disgust enrich our understanding of disgust is *all* that we need to appeal to in order to show how affective experiences could enrich our thoughts about disgustingness. To go beyond that, as Johnston does, requires more argument, argument that – in any case – Johnston does not provide. This alternative proposal seems to explain all that Johnston’s account does, but without requiring any particular view of affect to be true.<sup>98</sup>

## Section 5: Epistemic Justification

The final argument that we will consider for the claim that affective experiences are truth-apt representations of value is the best of the lot. The basic argument, one I will call the Argument from Epistemic Justification, goes as follows:

- (1) Affective experiences (defeasibly) epistemically justify evaluative beliefs.
- (2) If X epistemically justifies the belief that p, then X is a truth-apt representation that p (or X is a truth-apt representation of something that implies p or makes it probable).

---

<sup>98</sup> I would like to thank Sigrún Svavarsdóttir (p.c.) for encouraging me to get clearer about why I take this proposal to be preferable to (or at least as good as) Johnston’s.

(3) So, affective experiences are truth-apt representations of values (or are truth-apt representations of things that imply the truth of evaluative beliefs or make them probable).

Premise (1) seems quite plausible. To adapt a case from Sabine Döring (2007: p. 377), imagine that you witness a toddler accidentally drop his ice cream cone and be punished harshly for that reason, and you come to feel indignant. As Döring notes, your feeling of indignation seems to justify you in believing that his treatment was wrong, or unfair, or cruel. More simply, imagine that, upon hearing a certain song for the first time, the melody pleases you and you feel attracted to listening to it again. Here, too, it seems as though your positive feelings towards the song defeasibly epistemically justify you in believing that it is good in one way or another. Less simply, consider the role that feelings seem to play in moral philosophy, particularly in our “intuitions” about thought experiments.<sup>99</sup> Some of us feel horrified by the prospect of letting the child drown in Peter Singer’s (1972) famous case, or we feel outraged by the idea that the protagonist could not “detach” herself from the violinist (Judith Jarvis Thomson (1971).) And so on. In all of these cases, and many more, it seems as though we have positive or negative feelings towards these characters and their actions or options, and it seems as though these feelings, to some extent, justify the correlative positive or negative judgments (e.g. that such-and-such is right, wrong, or permissible).

But we only get from (1) to (3) via (2), so why should we believe that (2) is true?

First, think about the least controversial kinds of cases of one mental state (defeasibly

---

<sup>99</sup> A number of philosophers have embraced theses in this general area. Ralph Wedgwood (2007), for example, suggests that moral intuitions are “offline simulations” of emotions. Svavarsdóttir (2015, pp. 233-5) suggests that it is on the basis of emotional responses to thought experiments, like Nozick’s experience machine, that we reach verdicts about value. In that way, at least, “emotional responses are [our] ultimate guides to what is of value.” (Thanks are due to Sigrún Svavarsdóttir for helpful discussion of these aspects of her view.)

epistemically) justifying another.<sup>100</sup> For example, if you believe that grass is green, then you are defeasibly justified in believing that grass is green or that grass is red.<sup>101</sup> If you believe that every bit of grass that anyone has ever seen has been green, then you are defeasibly justified in believing that the next bit of grass that someone sees will also be green. And so on. In these cases, you have defeasible justification to believe the logical implications of your beliefs, or the things that your beliefs make probable or inductively support.

But it also seems like things besides beliefs can justify beliefs: in particular, perceptions, or perceptual experiences, seem to defeasibly epistemically justify beliefs, too. For example, my visual experience as of a coffee cup on the table seems to justify my belief that there is a coffee cup on the table. Or, if vision cannot itself represent properties like being a coffee cup or being a table, perhaps the relevant visual experience at least justifies my belief e.g. that colors, shapes, and textures are arrayed in way thus-and-so. And maybe the relevant visual experience “combines with” background knowledge about the relations between such color, shape, and texture distributions and the presence of coffee cups and tables to jointly justify my belief that there is a coffee cup on the table.<sup>102</sup> However these debates shake out, it seems that perceptual experiences (perhaps in tandem with background knowledge) can also defeasibly epistemically justify beliefs.

But what do beliefs and perceptions like these have in common such that they are good candidates to be able to justify, say, the belief that *p*? Well, they are all truth-apt

---

<sup>100</sup> I will assume throughout this section that the epistemic justifiers of beliefs are all mental states, and not e.g. the worldly things that the relevant mental states are about. This assumption is innocent here since (1) claims that affective experiences – i.e. a kind of mental state – justify other mental states. But see e.g. Earl Conee & Richard Feldman (2001) for a defense of this kind of approach.

<sup>101</sup> These cases raise bootstrapping worries, of course, but I will just sweep such worries under the “defeasibility” rug here, since these cases are only used here for illustrative purposes. See e.g. Gilbert Harman (1986) and John Broome (1999) and the literatures that have grown up around each for further discussion.

<sup>102</sup> For a helpful recent discussion of what visual experiences do or do not represent, see Ned Block (2014).

representations of p, or of something that implies p or makes it probable. And at least some *non-truth-apt* representations of p (or of things that would imply p or make p probable) clearly do *not* justify a belief that p, e.g. when we *imagine* that p or we *wonder whether* p or we *desire* that p. Further, the plausible candidate justifiers all seem to be the kinds of things from which we could reason well to a belief that p.<sup>103</sup> And perhaps, as e.g. Jim Pryor's (2005: p. 189) "Premise Principle" (PP-1<sup>104</sup>) has it, these are features that any (defeasible epistemic) justifier needs to have. According to PP-1:

The only things that can justify a belief that P are other states that assertively represent propositions [i.e. represent those propositions in a truth-apt way], and those propositions have to be ones that could be used as premises in an argument for P. They have to stand in some kind of inferential relation to P: they have to imply it or inductively support it or something like that. (ibid)<sup>105</sup>

One key virtue of PP-1 is that it ensures that there are appropriate relations between any potential justifier of a belief that p and the *truth* of the belief that p. For that reason, at least, something in the ballpark of PP-1 is a very reasonable constraint.

And PP-1 seems to be a disaster for attitude views of affective experiences (given premise (1) of the Argument from Epistemic Justification). For example, if I feel pleased that I am eating a sandwich, I, according to an attitude view, have a non-truth-apt pro-attitude towards the content *that I am eating a sandwich*. But *that* content – that I am eating a sandwich – does not on its own imply or make probable that my eating a sandwich is good,

---

<sup>103</sup> This seems very important, e.g. for reasons emphasized by Bill Brewer (1999: p. 154): "A reason for S's believing that P is a fact about that person which makes her believing that thing intelligible from the point of view of rationality. If this is to happen then the selected fact about S must be somehow related to (her) believing that P. And since this relation is to make her believing that P intelligible *from the point of view of rationality*, it is necessarily a relation which obtains in virtue of the correctness of some kind of reasoning. That is to say, successfully giving such a reason makes essential reference to the premise of an inference of some kind, whose conclusion is appropriately related, most likely by identity, to the content of the belief for which the reason is being given." We will return to such issues momentarily.

<sup>104</sup> The reasons for including the number after PP will become apparent shortly.

<sup>105</sup> Pryor does not himself accept PP-1. He rejects it for a reason we will consider momentarily.

and nor can I reason well from that content to that belief. Conversely, this is all grist for an evaluative content view's mill. If my feeling pleased that I am eating a sandwich is a truth-apart representation akin to or just a kind of perception of the content that my eating a sandwich is good, then my experience *does* imply that my eating a sandwich is good and – at least insofar as we can “reason well” from perceptions to correlative perceptual beliefs – I can reason well from that feeling to that belief.

Luckily for (phenomenal) attitude views, however, we have sufficient independent reason to reject PP-1. PP-1, as stated, has a highly implausible implication which has nothing to do with affect. If an adherent of an evaluative content view sticks with PP-1, then her view inherits that problem. But, as I will argue presently, once we make the needed revisions to PP-1, attitude views can offer an attractive account of the epistemic role of affect. Either way, then, the Argument from Epistemic Justification turns out not to be nearly as powerful as it seemed.

The highly implausible implication of PP-1 is that, if it is true, then the fact that we are having a conscious experience gives us no reason at all to believe that we are having that experience. (Pryor (2005: p. 189)) Suppose that you are consciously judging that *p*, or consciously imagining that *p*, or consciously wondering whether *p*, or, say, you are feeling pleased that *p*. If PP-1 is true, then being in these states cannot not even defeasibly epistemically justify you in believing that you are undergoing them, because the *content* of my judging that *p* or imagining that *p*, etc. is just *p*: it – the content that *p* – does not imply or make probable any claims about what I am judging, imagining, or etc. If an evaluative content view needs to accept this implication to get the Argument from Epistemic Justification to work, that is a bug and not a feature of the view.

But we can slightly revise PP-1 in a way that avoids this unsavory implication, but stays true to the original intent of the principle (more or less):

PP-2: The only things that can justify a belief that P are

(i) other states that assertively represent propositions, and those propositions have to be ones that could be used as premises in an argument for P. They have to stand in some kind of inferential relation to P: they have to imply it or inductively support it or something like that; or

(ii) other states *the subject's possession of which* stands in some kind of inferential relation to P: the subject's possession of the state has to imply that P or inductively support that P or something like that.

For example, suppose that you are consciously imagining (or judging or desiring) that q. And let us say that a belief with the following content – that I am imagining (or judging or desiring) that q – has the content that p. Your consciously imagining (or judging or desiring) that q *does* stand in “inferential relations” to the (content of the) belief that p: namely, it implies that p. So, PP-2 allows us to avoid this unsavory implication while maintaining the primary virtue of PP-1: namely, we hold onto the appropriate relations between potential epistemic justifiers of a belief that p and the truth of the belief that p. So, PP-2 is preferable to PP-1.

But PP-2 actually gives attitude views a bit of room to maneuver vis-à-vis the Argument from Epistemic Justification. First, if PP-2 is true, then *it is not true that only truth-apt mental states can serve as epistemic justifiers*. Next, we have already seen that the *contents* of affective experiences, if attitude views are right about such experiences, do not justify evaluative beliefs (at least in general). But can the fact that you are *having* a particular affective experience justify you in believing that such-and-such is good or bad? The second



point is that it might, in a few different ways, and those ways are what I would like to discuss now.

Here is a radical position that one could take about value: if one has a positively valenced affective experience towards X, then X is good. If this claim about value is true, then it would turn out that the fact that you are having e.g. a positive affective experience towards X would in fact imply that X is good, and so, for all PP-2 tells us, your affective experience could justify the relevant evaluative belief! Though this proposal about value is probably false, it displays nicely the logical space that PP-2 opens up for attitude views of phenomenal valence.<sup>106</sup>

Once we reject PP-1, what constraints guide us in providing amendment or replacement principles? We want to ensure that our principles (a) allow a perception that p to justify the belief that p, (b) allow our experiences to justify the belief that we are having them (as seen above), and (c) to rule out the possibility that e.g. imagining that p, wondering whether p, or desiring that p can justify a belief that p. But there are more general desiderata as well. We might also, as suggested above, want to ensure that our amendment or replacement principles (d) guarantee an appropriate connection between putative justifiers and the truth of the belief that p.<sup>107</sup> Most generally, as Pryor (2005: pp. 192-193) points out, the most fundamental motivation for the Premise Principle is the desire to (e) avoid

---

<sup>106</sup> Also, of course, it might turn out that our affective experiences reliably occur in the presence of the relevant values – e.g. maybe our feelings of fear reliably occur in response to things that really are fearsome – even if an attitude view is true of affective experiences. This, too, might amount to the right kind of connection between affective experiences and the relevant values for such experiences to be able to “pass” the PP-2 “test.”

<sup>107</sup> As Declan Smithies (p.c.), Tristram McPherson (p.c.), and Sigrún Svavarsdóttir (p.c.) have all pointed out to me, this has an awfully reliabilist flavor to it, and so will not be especially appealing to the many non-reliabilists about epistemic justification. I take the point, but if (d) is not a constraint on adequate principles, that makes my job easier, not harder, so I see no problem with my allowing (d) as such a constraint here.

arbitrariness.<sup>108</sup> He “agree[s] that epistemologists should give principled, non-arbitrary rationales for the justifying relations they postulate.” (2005: p. 193) But, he says, “I see no reason to think that they will *have to* appeal to propositional contents to do it.” (ibid)

I would like to take this idea of Pryor’s and run with it. I will propose here a justificatory framework that allows us to satisfy (a) – (e) and shows how affective experiences can justify evaluative beliefs even if a phenomenal attitude view is the correct account of such experiences. The framework is provided by PP-3:

PP-3: The only things that can justify a belief that P are

(i) other states that assertively represent propositions, and those propositions have to be ones that could be used as premises in an argument for P. They have to stand in some kind of inferential relation to P: they have to imply it or inductively support it or something like that, or they would imply it or inductively support it in *optimal conditions*; or

(ii) other states the subject’s possession of which stands in some kind of inferential relation to P: the subject’s possession of the state has to imply that P or inductively support that P or something like that, or it would imply it or inductively support it in *optimal conditions*.

On this approach, mental state M can justify a belief that p only if M (its content or the state itself) either implies or makes probable that p, *or would in optimal conditions*.<sup>109</sup> PP-3 allows us to satisfy criteria (a) and (b) straightforwardly, and (c) falls out of it, too: wonderings whether p and desires that p and imaginings that p, for example, do not justify the belief that p because there is no reason to think that it would be the case that p, were we to have *those*

---

<sup>108</sup> Pryor accepts that this criterion rules out non-intentional qualia as justifiers, for example. See his (2005: p. 192) for further discussion.

<sup>109</sup> Alvin Goldman (1986) attempted to use a method like this (although he talked about “normal worlds” instead of optimal conditions) at least partly in order to make reliabilism compatible with what seems like the right thing to say about brains in vats. One neat lesson of our recent discussions, however, is that this basic kind of maneuver has significant appeal independently of reliabilism. But, of course, it will be difficult to specify “optimal conditions.”

states in optimal conditions.<sup>110</sup> The connection between the truth of the putatively justified belief and our justifiers is also preserved, as (d) requires: my visual experience as of a's being F, say, justifies my belief that a is F because, were I to have that experience in optimal conditions, a would be F, and so the belief that a is F would be true. And as we can now see, there is nothing arbitrary about the connection that PP-3 requires between justifiers and justified. So the move to PP-3 is well motivated independently of questions about affect.

Though the "optimal conditions" will be different for different kinds of mental states or contents, our affective experiences would co-vary with the values of things were we in optimal conditions (e.g. were we fully rational, fully informed, highly sensitive to the features of our surroundings, etc.<sup>111</sup>). This, in fact, is just another way of making a claim that we have made several times already: positively (negatively) valenced states (e.g. affective experiences, positive or negative motivations, evaluative beliefs) are the ones that are correct (etc.) to have in response to things of value (disvalue). "Fittingness," "optimal conditions," etc. should all be heard here as pointing to the same basic idea.

The important point is that for all PP-3 tells us, our having affective experiences can justify our evaluative judgments, *whatever* view of affect we adopt. Your feeling disgusted by X, say, is a state such that, were you to have it in optimal conditions, X would actually be disgusting, and so your belief that X is disgusting would be true. This is true on every theory of affect that accepts that values constitute the correctness conditions for feelings. The

---

<sup>110</sup> There is no reason, as far as I can tell, for thinking that one's imagining that p would co-vary with the truth of p in optimal conditions. The world need not be as we imagine it to be, even if everything is going well with the world and with our imagination. Likewise, there is no good reason to think that, in optimal conditions, we would only desire things that are true or are likely to be true. We can know all about what the world is like and yet wish it were otherwise. There will be tricky cases, though, e.g. hunches and suspicions. But here it is important to remember that PP-3 only proposes a *necessary* condition on something's ability to be an epistemic justifier.

<sup>111</sup> What *exactly* the optimal conditions for feelings are is a huge question that I cannot further address here.

connections between our affective experiences and the (truth of the) relevant evaluative beliefs, then, are not nearly accidental on this picture, and I take that to be the basic, compelling motivation for the initial idea that only truth-apt representations can serve as epistemic justifiers.<sup>112</sup> So if one wants to replace PP-3 with something that requires that justifiers be truth-apt representations, further argument will be needed. In the absence of such argument, we should conclude that the Argument from Epistemic Justification does not favor an evaluative content view of phenomenal valence.

Maybe, once we add even more necessary conditions that something has to meet to be capable of being an epistemic justifier, so that we can move all the way to a set of individually necessary and jointly sufficient such conditions – e.g., perhaps, by requiring the relevant states to be *conscious* – then we will see new trouble for attitude views of phenomenal valence. Maybe that will happen, but here is why I doubt it. A state that feels exactly like your feeling of outrage towards the treatment of the toddler (see the beginning of this section) will defeasibly justify (or fail to justify) the same beliefs as your feeling of outrage towards the treatment of the toddler, whether it (re)presents value (or disvalue) or not. Why does it give you reason to think that the treatment of the toddler is bad instead of good? Because of the negative valence of its attendant attitude. Why does it give you reason to think that the treatment of the toddler is bad, instead of giving you reasons to think that something else is bad, or something else altogether? Because it has the content e.g. that the toddler got slapped and yanked by the arm for accidentally dropping his ice cream. If this is mysterious, that is only because we have been overly focused heretofore on the epistemic import of beliefs and perceptions. There is nothing arbitrary about the epistemic connection

---

<sup>112</sup> I would like to thank Declan Smithies (p.c.) for discussion of these points.

between affective experiences and evaluative beliefs, even if a phenomenal attitude view is the correct account of such experiences.<sup>113</sup>

In the previous section, we appealed to the systematic connections between affective experiences and values – systematic connections that obtain *regardless* of one’s view of affective experience – to explain an important connection between affective experience and evaluative thought and judgment. In this section we have explained another connection between affective experience and evaluative thought and judgment by drawing on analogous resources. At this point, then, we should be at least fairly confident that a great many of these basic kinds of arguments for evaluative content views – i.e. arguments that appeal to relations between affect and evaluative thought or judgment – will be addressable in analogous ways. In this way, we now have reasons to think that our responses to these arguments have not just been desperate, ad hoc measures, but rather point to a more general problem with such arguments. The phenomena they appeal to do not point us towards an

---

<sup>113</sup> Some of Karl Schafer’s (2013) points about *desire* might also be worth taking on board here. He claims, plausibly, that desires present their intentional objects with “imperative force.” But he also claims that “one of the ways to present A as something that I ought to do is just to present A with imperative force.” (2013: p. 277) In this sense, then, my desire to  $\Phi$ , say, presents  $\Phi$ -ing as something that I ought to do. This is a good result, according to Schafer, precisely because it helps to explain how it is that desires can *justify* e.g. my belief that I ought to  $\Phi$ . So far, this just sounds like a standard “guise of the good” theory of desire according to which desires represent values (or oughts). But Schafer’s interesting contribution here comes at the next stage. According to Schafer, the points made above do not imply that desires have values as parts of their representational contents. Schafer’s most illuminating discussion of how this works comes in a footnote (2013: p. 278, n. 44) where he writes that “just as it is very plausible that any state that presents P with assertoric force presents P as true, it is also plausible that any state that presents P with imperative force presents P as something that ought to be done.” He also claims, very plausibly, that even though perceptual experiences present things with assertoric force, and so “present P as true,” they do not include, *in their representational content*, that such-and-such is true. Applying this point to desire, the idea is that desires that p “present p as something that ought to be done,” but do not include, *in their representational content*, that p ought to be done. Applied to affective experiences, the phenomenal attitude view-friendly idea would be that a positive feeling that p “presents” p with “imperative force,” and so “presents P as something that is good” even though that feeling does not include, *in its representational content*, that p is good. I suspect that this way of speaking might lead to avoidable confusion, but if Schafer’s view of desire can be made to work, and does not just turn out to be a standard guise of the good theory, then I see no reason why a phenomenal attitude view might not adopt an analogous proposal in the affective realm. (Thanks are due to Declan Smithies (p.c.) for discussion of these issues.)

evaluative content view of affective experience and phenomenal valence. Rather, they bring to our attention features of affective experiences that, though important, are compatible with any one of a number of views of affective experience. As such, these phenomena cannot tell in favor of one view of affective experience over another.<sup>114</sup>

## Section 6: Conclusion

As we have now begun to see, evaluative content views and phenomenal attitude views paint very different pictures of our affective lives. Nevertheless, it is important not to overstate the differences between these views. There *are* a number of very important connections between affect and value – and even between affect and *evaluation* – and it is to the credit of evaluative content views that they lead us to see and focus on these connections. Furthermore, there might even be a number of *intentional* relations between affect and value, even according to phenomenal attitude views, at least if A's being the correctness condition for B implies that there is an intentional relation between A and B.<sup>115</sup> But since a phenomenal attitude view can embrace these points as fully as evaluative content views can, there is nothing here to favor an evaluative content view.

What we have learned from this chapter is not exactly that a *phenomenal* attitude view is preferable to an *evaluative* content view, but rather that *attitude* views are preferable to *content*

---

<sup>114</sup> I would like to thank John Hurst (p.c.) for bringing the Premise Principle to my attention in the first place, and for a great many helpful conversations about direction-of-fit and epistemology (and a great many other topics broached in this dissertation as well).

<sup>115</sup> First, maybe the very fact that values constitute the correctness conditions for feelings amounts to there being an intentional relation between feelings and values, even if feelings do not represent values. Julien Deonna & Fabrice Teroni (2012) might occupy a position like this with respect to the *emotions*. They write: "The proposal holds that emotions indeed stand in intentional relations to evaluative properties, but that these intentional relations do not assume the form of emotions that have evaluative properties as their objects...Emotions are not experiences of these evaluative properties themselves but rather specific attitudes we adopt towards actual or potential bearers of evaluative properties, attitudes that are correct or incorrect depending on whether the object actually exemplifies the relevant evaluative property." (2012: p. 85) Or, as they put it more simply, "emotions stand in intentional relations to values without being about values." (ibid.) A phenomenal attitude view can, at least in principle, say the same thing about affective experiences. Second, see also the discussion of Karl Schafer's (2013) in Section 5 above.

views more generally. Attitude views have the virtues highlighted e.g. in Section 3 of this chapter and in Section 4.4 of the previous chapter, but without having the correlative vices of content views (ibid). And as we have seen throughout this chapter, attitude views seem at least as well placed as content views to capture important relationships between affect, value, and evaluation. As such, I conclude that we should embrace an attitude view of phenomenal valence. But phenomenal attitude views are not the only game in town. They have at least one major competitor – motivational attitude views – and it is to a discussion of such views that we now turn.

## Chapter 4: Affect and Motivation

### Section 1: Introduction

At a number of crucial points in earlier chapters (see especially Section 4.4 of Chapter 2 and Section 3 of Chapter 3) I have drawn on apparent similarities between affective experiences, on the one hand, and mental states like desires, motives, “inner imperatives,” and intentions on the other. Let us call these latter kinds of mental states ‘conative states,’ leaving it open exactly what (if anything) all conative states have in common, and leaving it open whether affective experiences are a species of this genus. I have suggested that affective experiences are similar to conative states in at least the following – deep and important – respects: neither affective experiences nor conative states are truth-apt representations of what they are about, and both affective experiences and conative states are valenced states in virtue of partly consisting of special (i.e. valenced) attitudes, and not in virtue of having special (i.e. valenced) contents. Let us call these attitudes that conative states partly consist in ‘conative attitudes.’

When we think about conative states – and conative attitudes, in particular – it is also very natural to think that conative attitudes are valenced in virtue of their *motivational profile*. A desire that p, for example, one might think, is a positively valenced orientation towards p



in virtue of the fact that that desire disposes one<sup>116</sup> to “act positively” with respect to p, i.e. to *do* what one believes (suspects, perceives, etc.) will bring it about that p, or will protect p, preserve p, promote p, etc. That is, one might reasonably think that a desire that p is a positive orientation towards p in virtue of a desire that p having what we earlier called *positive motivational valence* with respect to p. Likewise one might think that an aversion to p is a negative orientation towards p in virtue of the fact that that aversion disposes one to “act negatively” with respect to p, i.e. to do what one believes (etc.) will bring it about that not p, or to do what one thinks will destroy p, prevent p, help one to avoid p, reduce p, etc. That is, one might reasonably think that an aversion to p is a negative orientation towards p in virtue of its having *negative motivational valence* with respect to p.

Given what seem to be the deep and important similarities between affective experiences and conative states mentioned in the first paragraph, a very natural move at this point is to suggest that affective experiences and conative states are similar in these latter respects, too. That is, it is very natural to think that affective experiences, too, are valenced in virtue of *their* having a certain motivational profile. The proposal would have it that a feeling is a *positive* feeling towards p in virtue of the fact that that feeling disposes one<sup>117</sup> to

---

<sup>116</sup> Perhaps desires are not best thought of as being identical to any particular behavioral dispositions, but are instead better thought of as something like the “grounds” of such dispositions. When I say that a desire that p disposes A to Phi, I mean that to be neutral between these options. Or maybe desires need not be identical to *or* the grounds of any *behavioral* dispositions, but can instead be or ground dispositions to have certain feelings, etc. These issues having to do with the metaphysics of desire – or analogous issues to do with the metaphysics of intentions, etc. – will not be addressed further in this chapter. But some related issues will be discussed in the concluding chapter of the dissertation.

<sup>117</sup> We should remain neutral among several possibilities here. Maybe motivational attitude views will or should claim that feelings *are themselves* behavioral dispositions, or maybe that they are the *grounds* of such dispositions, or maybe that they are as such parts of “larger” states that *in turn* are (or ground, etc.) behavioral dispositions, or maybe that feelings necessarily come along with some *distinct* state that is (grounds, etc.) such dispositions. And so on. For present purposes, we should allow motivational attitude views to tie feelings to motivation in whichever of these ways is most plausible. As we will see, the resolution of *these* disputes will not matter vis-à-vis the debate between motivational attitude views and phenomenal attitude views. At any rate, when I say that a feeling disposes one to Phi I mean to be neutral among the various possibilities just discussed.

act positively with respect to p, and a feeling is a *negative* feeling towards p in virtue of the fact that that feeling disposes one to act negatively with respect to p. This proposal – a motivational attitude view of phenomenal valence – would have us assimilate phenomenal valence to motivational valence.

In the next section I will describe motivational attitude views in a bit more detail and bring out some of their most significant attractions. After that, the rest of the chapter will, more or less, be organized around the relations between (a) motivational attitude views and (b) the criteria of adequacy for views of phenomenal valence mentioned in the previous chapter. As a reminder, those criteria were the following: first, the properties, said by the relevant account to be the properties that unite the positive (negative) affective experiences, have to be properties that all positive (negative) affective experiences *actually share*; second, the properties appealed to should enable us to *distinguish* the positive from the negative affective experiences; and third, the properties, said by the relevant account to be the properties that unite the positive (negative) affective experiences, have to be properties the possession of which is suitably related to (more strongly: explanatory of) the *valence* of the experiences (that is, we have to be able to see, upon reflection, perhaps with some coaching, how it is that someone's instantiating *that* property amounts to that person's being *for* or *against* something.)

In these terms, what I will argue in Section 3 below is that even though straightforward kinds of motivational attitude views can pass the *third* of these tests with flying colors, they also fail to satisfy the *first* of these criteria. So, straightforward kinds of motivational attitude views are inadequate accounts of phenomenal valence. Then, in Section 4, I will argue that even though less straightforward kinds of motivational attitude

views can probably satisfy the *first* (and second) criteria above, they also fail to satisfy the *third* of these criteria. So, less straightforward kinds of motivational attitude views are inadequate accounts of phenomenal valence, too. Though I do not rule out every conceivable version of motivational attitude views in this chapter, I believe that, on the basis of the considerations adduced here, we should – at least for the time being – reject motivational attitude views.

Assuming that the arguments of this and the previous chapter are successful, i.e. that we should not embrace evaluative content views or motivational attitude views of phenomenal valence, does that mean that phenomenal attitude views are our only reasonable option? This is, more or less, the topic of the concluding section of this chapter. Spoiler alert: it's complicated. We will need another chapter to decide between the live possibilities. But that is for later. Now to motivational attitude views and their attractions.

## **Section 2: Motivational Attitude Views**

### **Section 2.1: Positive and Negative Motivational Profiles**

Motivational attitude views claim, roughly, at a first pass, that what unites the positively phenomenally valenced states is the fact that they each come along with positive motivation vis-à-vis their intentional objects, and what unites the negatively phenomenally valenced states is the fact that they each come along with negative motivation vis-à-vis their intentional objects. The proposal has it, for example, that insofar as all of the feelings in Group 1 (see earlier chapters) are positively phenomenally valenced orientations towards what they are about, they are such in virtue of the relations they stand in to positive motivation. In this section I will present a sympathetic picture of how a plausible

motivational attitude view should understand *what it is* to be positively or negatively motivated with respect to something.

First, what is it for someone to be motivated to do something? Motivation certainly has something to do with *action*, but being motivated at  $t$  to  $\Phi$  cannot, for example, require you actually *to be  $\Phi$ -ing at  $t$* : I am *motivated* right now to turn off my computer and call a friend, and yet here I sit, typing away, not even taking any steps towards bringing it about that I call my friend. A simple account of motivation that accommodates this point might have it, instead, that to be motivated to  $\Phi$  is just to be *disposed* to  $\Phi$ . But, unfortunately, one's having a disposition to  $\Phi$  is neither necessary nor sufficient for one's being motivated to  $\Phi$ .

It is not *necessary* e.g. because we seem to be capable of being *motivated* to do impossible things (at least if we do not know that they are impossible), but nobody is disposed to *do* impossible things (they are impossible!). For example, suppose that an all-powerful God *cannot* exist, e.g. because of the paradoxes to which omnipotence would give rise. Nevertheless, it seems like one might be motivated to please an all-powerful God. I am quite confident, in fact, that many people throughout history have been motivated to do precisely that. Whether an all-powerful God is possible seems not to make any difference to whether someone can be *motivated* to please "him," although it makes a great deal of difference to whether someone can be *disposed to* please "him." So, a disposition to  $\Phi$  is probably not necessary for motivation to  $\Phi$ .

A disposition to  $\Phi$  is not *sufficient* for motivation to  $\Phi$ , either, as we see when we consider things like reflexes (kicking out your leg when your knee is hit with a rubber hammer, etc.), involuntary bodily processes (digestion, etc.), clumsiness (I am disposed to hit

my shin on that damned coffee table, but I am certainly not motivated to do so!), and so on. In each of these cases we are disposed to  $\Phi$  even though we are not motivated to  $\Phi$ , so, again, the simple theory of motivation considered above will not do.

The objections to necessity above suggest one simple fix to our simple theory of motivation. Instead of trying to explain the motivation to  $\Phi$  in terms of a disposition to  $\Phi$ , we might try explaining it in terms of a disposition to do what *you believe* (suspect, perceive, etc.) will bring it about that you  $\Phi$ . We can imagine that those who are motivated to please an all-powerful God *are* disposed to do what *they believe* (etc.) will please an all-powerful God, even if such a God does not or cannot exist.

The objections to sufficiency, however, are more difficult to accommodate, and it is not worth our while to try to do so satisfactorily here.<sup>118</sup> Instead, I suggest that we settle, for now, with the view just floated – i.e. that to be motivated to  $\Phi$  is to be disposed to do what you believe (suspect, perceive, etc.) will bring it about that you  $\Phi$  – and note a few problems that any plausible motivational attitude view of phenomenal valence will want its account of motivation to avoid.

The first such problem that we (or, at least, a motivational attitude view) will want to avoid is analogous to a problem we considered in the previous chapter. Surely non-human animals and very young human children can have affective experiences. So our theory of motivation had better not imply that only incredibly sophisticated creatures can be motivated to act. For example, if one's being motivated to  $\Phi$  requires one's intending to  $\Phi$ , and one's

---

<sup>118</sup> To provide sufficient conditions for motivation, we will have to address extremely vexed cases, e.g. “Freudian slips” (see Velleman (2000) for discussion) and “deviant causal chains” (see e.g. Davidson (2001: pp. 79ff) for discussion), i.e. cases where the relevant behaviors seem to be caused by relevant attitudes and beliefs of the agent, but where it is not at all clear that the agent was *motivated* to perform the relevant behaviors. We cannot solve such problems here.

intending to  $\Phi$  is a matter of having a suite of very complicated beliefs, desires, etc., then a motivational attitude view of phenomenal valence would be straightforwardly implausible e.g. because it would imply, falsely, that non-human animals and very young human children could not have affective experiences. For these reasons, we should not require a great deal of sophistication from motivation in what follows.<sup>119</sup>

The second problem that a motivational attitude view will want to avoid is vacuousness. For example, if we say that positive feelings are positive in virtue of their coming along with positive motivation, then we had better not go on to say that positive motivations are positive in virtue of their coming along with positive feelings. In a similar spirit, we had also better avoid the following. Suppose we claim that a positive feeling towards  $\Phi$ -ing is positive in virtue of coming along with motivation to  $\Phi$ , and that one's motivation to  $\Phi$  is (at least in part) a matter of one's having a "pro-attitude" towards  $\Phi$ -ing. This is all kosher so far, but suppose that we said nothing else about what makes such pro-attitudes *pro*-attitudes. That would be desperately unsatisfying. It might not make the ensuing view vacuous, exactly, but it certainly has the feel of a good question left unanswered, a bump in the rug moved but not flattened. So, for example, if we would eventually like to understand motivation in terms of pro-attitudes, then we better have something to say about those pro-attitudes, and we had better not say that they are *pro*-attitudes in virtue of their ties to *affect*.<sup>120</sup>

---

<sup>119</sup> Maybe e.g. J. David Velleman (1992b: p. 462) is right that "human action *par excellence*" requires quite sophisticated motivational machinery. But even if so, the point in the main text can be read as claiming that one can e.g. feel pleased about something without being motivated to engage in human action *par excellence* with respect to it.

<sup>120</sup> A number of theories of desires (or intentions or "pro attitudes") have been proposed down the years – I am thinking in particular of the views of David Armstrong (1993), Harry Frankfurt (1971), Robert Stalnaker (1984), Michael Smith (1994, etc.), and maybe Donald Davidson (1963) – that might provide a model for how to meet these challenges. According to Harry Frankfurt (1971) and Michael Smith (1994), at any rate, desires

Finally, in giving an account of *positive* motivation and *negative* motivation, a motivational attitude view will have to complicate the above view of motivation somewhat, but she needs to be careful in how she does so. Maybe what it is to be positively motivated with respect to  $\Phi$ -ing, i.e. with respect to performing some particular action or behavior, is, precisely, to be motivated *to*  $\Phi$ . But, first, it seems as though we can feel positively about things other than actions or behaviors, e.g. propositions or even people. This might be easily solvable in the case of propositions – we are disposed *to bring it about* that p, perhaps – but what about people? Second, the account just offered does not seem to give us a perspicuous model for *negative* motivation. If I am negatively motivated with respect to it

---

are motivationally relevant pro-attitudes that (a) we can say illuminating things about and (b) do not require affective experiences. On these views, what is essential to desire is desire's tie to behavior. According to Frankfurt, an "effective desire" is one that "moves (or will or would move) an agent all the way to action" (1971: p. 8) and the notion of a desire (sans phrase) is "the notion of something that merely inclines an agent in some degree to act in a certain way." (ibid) (Frankfurt is not too interested in desires sans phrase. As he notes (1971: p. 7), it is possible for someone to have such a desire for X even though he "doesn't 'really' want to X" and he "would rather die than X.") And Frankfurt's clear throughout his work that desire need not have much to do with feeling. According to Frankfurt, it is possible for A to desire to do X even if "the prospect of doing X elicits no sensation or introspectible emotional response in A" and "A is unaware of any feelings concerning X." (1971: pp. 7-8) One of the more telling passages here comes in his discussion of the physician who wants *to want* to do heroin, so he can better understand what it is like for his addicted patients, but who does not want to do heroin and does not want the desire for heroin to be an effective desire. (1971: p. 9) Nevertheless, says Frankfurt: "If it is a genuine desire that he wants, then what he wants is not merely to feel the sensations that addicts characteristically feel when they are gripped by their desires for the drug. What the physician wants, insofar as he wants to have a desire, is to be inclined or moved to some extent to take the drug." (ibid) This passage strongly suggests (although it does not strictly imply, as Sigrún Svavarsdóttir (p.c.) has rightly pointed out to me) that connections to motivation and action, and *not* any connections to feeling (even "characteristic" such connections), are the essence of desire, according to Frankfurt.

Smith (1994: p. 105 – emphasis added) allows that "there is, after all, such a thing as the phenomenology of desire, as, for instance, to use one of Hume's...examples, 'when I am angry I am possessed with the passion' (1888: 416). That is, we may agree with Hume that, *on occasion*, when I have a desire, I have certain psychological feelings, analogues of bodily sensations." Smith even allows (1994: p. 114 – emphasis added) that some desires might be such as to have a certain phenomenology "essentially" ("*some* desires may be dispositions to have certain feelings under all conditions: these have phenomenological content essentially"). However, it is clear that, according to Smith, feelings are neither necessary nor sufficient for desire as such. They are not necessary, according to Smith (1994: pp. 105-107), because, if they were necessary, for example, then we would be infallible about what we desire, or too close to it for comfort. They are not sufficient, according to Smith, because feelings do not have propositional contents, whereas desires do (1994: pp. 107-108). The essence of desire is, instead, behavioral. As he helpfully puts it in recent work: "What's crucial is that it is the fact that desire either is, or is *inter alia*, a behavioral disposition that is absolutely crucial for us to understand how desire can play the role it plays in the explanation of action." (2011: p. 82)

being the case that  $p$ , is that just a matter of my being disposed to try to bring it about that *not*  $p$ ? That is not obvious. Third, as we will see later, it just plain seems too restrictive in any case, even if we could only feel *positively* about *actions*.

I suggest that we proceed from the outside in and give some room to intuition here. What kinds of *behaviors*, intuitively, are ways of helping something, say, or are ways of protecting it? Roughly, to be motivated to engage in what you believe (etc.) are these kinds of behaviors towards  $X$  is for you to be motivated positively towards  $X$ . Which kinds of *behaviors*, intuitively, are ways of harming something, say, or avoiding it? Roughly, to be motivated to engage in what you believe (etc.) are these kinds of behaviors towards  $X$  is for you to be motivated negatively towards  $X$ . It is the motivation to engage in these kinds of behaviors to which we would have to appeal, it seems to me, if we want to claim that positive or negative phenomenal valence is to be explained in terms of motivation. The less obvious it is that  $A$  sees  $\Phi$ -ing as a way to help or protect  $X$ , for example, the less obvious it is that  $A$ 's being motivated to  $\Phi$  amounts to  $A$ 's being positively oriented towards  $X$ . In what follows, we will be quite permissive about what counts as positive or negative motivation. But the farther we get from the paradigm cases of e.g. helping behavior or harming behavior, the less plausible it will seem that we have cases of positive or negative orientations at all. This is the third problem that motivational attitude views will want to watch out for when giving their account of motivation.

## **Section 2.2: The Attractions**

Now that we have a better understanding of what motivational attitude views are, it is time to turn to their attractions. I take it that there are at least two major reasons for embracing such a view.



First, there is something directly intuitive about such views. To see just how plausible such a view is, we can begin by considering a case from Michael Smith (1994: p. 60) having to do with *moral judgment* instead of affect. Suppose you and I are debating about whether each of us is morally required to give money to Oxfam. After some back and forth, I tell you, seemingly sincerely, that I am convinced: we are morally required to give money to Oxfam. As it happens, a minute later, an Oxfam representative knocks on the door, fundraising. But I have not the least tendency to give him any money for the cause. You would be puzzled. You would either suspect that I was being insincere a minute earlier when I told you that I was convinced that we should give money to Oxfam, or you would suspect that though I was motivated to give money, that motivation was over-ridden in this case by some other desire or set of desires. Thoughts like these have seemed to many philosophers to show that there are very close relations between moral judgment and motivation.<sup>121</sup>

With respect to affective experience, an analogous case is *much* more compelling (and I say that as someone who accepts that moral judgments are very closely tied to motivation indeed). Amend the case above so that now, instead of my telling you that we are morally required to give to Oxfam, I tell you that I *feel strongly attracted to* giving money to Oxfam and that nothing could please me more. Once again, the Oxfam representative then comes to the door, and I do not lift a finger to give him so much as a penny (and not because there are no spare pennies about or because I am busy or incapacitated). Again, but with even more confidence now, you would suspect me of pulling your leg, or you would suspect that my benevolent motives are being overridden by some stronger, perhaps less charitable,

---

<sup>121</sup> See e.g. A.J. Ayer (1952), C.L. Stevenson (1937), R.M. Hare (1952), Simon Blackburn (1984), and Allan Gibbard (1990) for some seminal discussions of the (putative) motivational import of moral judgments.

motivations. We can explain your suspicions quite well if motivational attitude views are true. And when we think of all manner of affective experiences, it just seems unfathomable that someone could feel these ways without even being *disposed* to do what one thinks could bring about (or prevent) their objects.

Putting a related point a bit differently, we often seem to act in certain ways *because* we feel attracted to certain things. To explain how feelings can issue in action, we only seem to need to “add” instrumental or constitutive beliefs, and we do not seem to need to posit an “additional” desire (see e.g. Simon Blackburn (1984) for discussion of a similar point about moral judgments.) I feel attracted to  $\Phi$ -ing, I believe that by  $\Psi$ -ing I will  $\Phi$ , and so (all else equal) I  $\Psi$ . The feeling (itself, without further ado) seems capable of playing the “desire role” vis-à-vis action, and motivational attitude views are well positioned to explain that.

The second major reason to embrace motivational attitude views is that such views promise to make phenomenal valence unmysterious. Motivation, as we have understood it here, is essentially just a causal notion. To be motivated to  $\Phi$  is, more or less, just to have some internal states that dispose you to behave in certain ways when you are in certain other internal states, or when you are confronted with certain external stimuli. Granted: the notion of motivation above does make reference to other mental states (i.e. beliefs, suspicions, or perceptions), so it is not straightaway reductive in the way that, say, a direct identification of motivation with certain brain states would be. But understanding phenomenal valence as motivational valence sits well with e.g. *behaviorist* or *functionalist* theories of the mind<sup>122</sup> in ways that, say, phenomenal attitude views might not.

---

<sup>122</sup> Behaviorists and functionalists share the general idea that internal states of a creature are mental states of certain kinds in virtue of those internal states (typically or necessarily) playing certain causal roles. Ultimately, the idea goes, mental states are, as such, states that lead – perhaps in combination with other states – from

Many of the great functionalists of the past several decades – e.g. Armstrong (1993), Lewis (1966, 1972), and Stalnaker (1984) – are explicitly motivated by the desire to show that minds are not entirely *sui generis* pieces of reality, only contingently related to the physical (e.g. neural) objects and properties with which the natural sciences have shown them to be so intimately bound up. Functionalists attempt to “fit” the mental into the physical order in roughly two steps: (1) show that what it is for X to be a certain kind of mental state is for X to have a certain syndrome of causes and effects; (2) show that something physical can have that syndrome of causes and effects. In these ways, the mental need not “threaten” the physical.

Sometimes, however, this procedure will have to be complicated. Consider, for example, Robert Stalnaker’s functionalist understanding of desire, an understanding of desire that is almost *identical* to the account of motivation in play here: “To desire that P is to be disposed to act in ways that would tend to bring it about that P in a world in which one’s beliefs, whatever they are, were true.” (1984: p. 15) If we are trying, as Stalnaker is, to fit the mind into the physical realm – or, as he puts it, to “explain intentional relations [e.g. beliefs and desires] in naturalistic terms” (1984: p. 6) – then, clearly, this account of desire only gets us halfway there. Surely we need to show how to fit *belief* into the physical order, too. And when one encounters Stalnaker’s account of belief – “to believe that p is to be disposed to act in ways that would tend to satisfy one’s desires, whatever they are, in a world in which P (together with one’s other beliefs) were true” (1984: p. 15) – one begins to worry that we

---

environmental input to behavioral output. For seminal philosophical discussions of such views, see e.g. Gilbert Ryle (2000), David Lewis (1966, 1972), David Armstrong (1993), and Ned Block (1980a, 1980b).

cannot fit both belief *and* desire into the physical order. At least one of them will have to remain “dangling,” to paraphrase J.J.C. Smart (1959).<sup>123</sup>

But functionalists are well aware of this issue and have attempted to “break into the circle” (Stalnaker (1984: p. 19)) in a number of ways. Stalnaker himself tries to break into the circle (1984: p. 19) by endorsing a “causal account of belief.” (1984: p. 18)<sup>124</sup> Other functionalists (see e.g. David Lewis (1966, 1972)) respond in other ways, e.g. by giving “network analyses” (Michael Smith (1994: p. 45)) that promise to enable us to fit mental properties into the physical order wholesale, but not piecemeal. When supplemented in these ways, functionalist accounts of the mind thus look extremely attractive, at least to those of us who hope, like Stalnaker, to “explain” the mind “in naturalistic terms.”

To summarize, motivational attitude views attempt to explain phenomenal valence in terms of motivation. The account of motivation that such views should adopt is in all relevant respects a functionalist account of motivation (compare Stalnaker’s account of desire). Functionalist accounts of mental states and properties, at least when supplemented in the ways just mentioned, promise to help us fit the mind into the physical order, and that

---

<sup>123</sup> “That everything should be explicable in terms of physics (together of course with descriptions of the ways in which the parts are put together – roughly, biology is to physics as radio-engineering is to electro-magnetism) except the occurrence of sensations seems to me to be frankly unbelievable. Such sensations would be “nomological danglers.”” (J.J.C. Smart (1959: p. 142)

<sup>124</sup> “Belief is a version of the propositional relation I called *indication*. We believe that P just because we are in a state that, under optimal conditions, we are in only if P, and under optimal conditions, we are in that state because P, or because of something that entails P.” (1984: p. 18) But as he notes, many mental states will satisfy these kinds of conditions, e.g. “the retinal image that forms on the eye” (ibid). To explain how belief differs from such states, he appeals to the relations noted above between belief and desire (and, ultimately, to the relations between desire and behavior).

is quite an appealing prospect. In this way, then, a motivational attitude view is quite an appealing account of phenomenal valence.<sup>125</sup>

Further, and the importance of this point cannot be stressed enough, motivational attitude views promise to achieve the foregoing without losing *valence* altogether. There is a straightforwardly comprehensible sense in which you are *positively* oriented towards  $\Phi$ -ing if you are disposed to do what you believe will bring about your  $\Phi$ -ing, for example. So if you tell me that my positive feelings are positive in virtue of their coming along with such tendencies, I know exactly what you mean and can see right away that I need to take your proposal seriously. Suppose, on the other hand, that you claimed that my positive feelings are positive *in virtue of* the fact that they tend to occur in a certain part of my brain. This proposal might well be physicalistically respectable, but I have no idea why I should accept it. What does any part of a brain have to do with valence? Motivational attitude views promise to avoid such puzzlement altogether, and so are marked out as very serious contenders for the truth in this area.

Those are the two major reasons to embrace a motivational attitude view of phenomenal valence, it seems to me.<sup>126</sup> At any rate, I hope that it is now clear enough, at least in outline, what a motivational attitude view of phenomenal valence is and why a reasonable person might believe it. A guiding thought behind such views seems to be that if

---

<sup>125</sup> Thanks are due to Sigrún Svavarsdóttir (p.c.), Justin D'Arms (p.c.), and Declan Smithies (p.c.) for all, independently, encouraging me to get clearer about what the (putative) advantage here for motivational attitude views is supposed to be.

<sup>126</sup> There are a number of more in-the-weeds reasons to embrace such a view, too. One such reason, for example, has to do with the heterogeneity objection discussed in Chapter 2. Chris Heathwood (2007) defends a "motivational theory of pleasure," for example, almost solely on the grounds that it provides a tidy solution to the heterogeneity objection. And so on. I have already made my case that a phenomenal attitude view can respond to the heterogeneity objection, and, at any rate, I take it that the two reasons to favor a motivational attitude view just mentioned in the main text are already substantial enough to justify us in taking such views seriously. Other reasons will have to be addressed elsewhere.

I would not take *any* steps to bring it about that p, even if I knew perfectly well that such steps are essentially cost-free and that they will guarantee that p, well, that just seems to imply that I am entirely *indifferent* as to whether p. Whatever I might “feel” towards p in such a situation, I certainly must not feel *positively* (or *negatively*) about whether p. Although there is something admirably frank and intolerant of bluff in this guiding thought, I will now argue that we should reject motivational attitude views of phenomenal valence in favor of a phenomenal attitude view.

### **Section 3: Failing the First Criterion**

I will not argue here that phenomenal attitude views do as well as motivational attitude views in terms of physicalistic respectability (although we will discuss related issues briefly in Section 3.4 below). Nor will I present an alternative account of the relations between affect and behavior that shows that phenomenal attitude views are preferable (or just as good) on that score. Instead, I will argue that motivational attitude views offer us an inadequate of phenomenal valence. The putative advantages of such views just discussed *would* be virtues of such views *if* such views at least satisfied the minimal criteria of adequacy that apply to accounts of phenomenal valence. But they do not.

As a reminder, those criteria were the following: first, the properties, said by the relevant account to be the properties that unite the positive (negative) affective experiences, have to be properties that all positive (negative) affective experiences *actually share*; second, the properties appealed to should enable us to *distinguish* the positive from the negative affective experiences; and third, the properties, said by the relevant account to be the properties that unite the positive (negative) affective experiences, have to be properties the possession of which is suitably related to (more strongly: explanatory of) the *valence* of the

experiences (that is, we have to be able to see, upon reflection, perhaps with some coaching, how it is that someone's instantiating *that* property amounts to that person's being *for* or *against* something.)

As I mentioned near the end of the last section, if A is motivated to  $\Phi$ , or A is positively motivated with respect to  $\Phi$ -ing in some looser sense, then A is positively oriented towards  $\Phi$ -ing. So, *if it is true* that, whenever someone, A, has e.g. a positively phenomenally valenced experience towards  $\Phi$ , A is positively motivated towards  $\Phi$ -ing, then motivational attitude views of phenomenal valence would satisfy the *third* criterion just listed. In this section, I will assume that that is the approach that motivational attitude views would take. That is, I will assume that, according to motivational attitude views, necessarily, if A has e.g. a positively phenomenally valenced experience towards  $\Phi$ , then A is positively motivated towards  $\Phi$ -ing.

But I will argue that this claim is false: there are (or could be) positively phenomenally valenced affective experiences that are unattended by any such motivations.<sup>127</sup> And so motivational attitude views, understood in this way, fail to satisfy the *first* of our criteria, and so are inadequate accounts of phenomenal valence. In Sections 3.1 – 3.4 I will present a series of cases to motivate the relevant worry. Then, in Section 3.5 I will present a more general consideration that supports the same conclusion.

### **Section 3.1: Wishes and Hopes**

The first cases worth thinking about when thinking about possible splits between affect and motivation are the most commonplace: the affective experiences characteristic of wishes and hopes. Right now, I feel strongly attracted to my (not my team's, but *my*) scoring the

---

<sup>127</sup> We will weaken – or at least complicate – this claim somewhat at the end of Section 3.4 below.

winning goal in a World Cup final. I wish I could do that, and I hope I do. But such wishes and hopes seem idle. They seem not to influence my behavior at all.

Tim Schroeder (2004: Ch. 1) presents such cases as potential counterexamples to theories according to which *desires* essentially involve dispositions to act. But even if one denies that wishes and hopes like these pose a serious problem for motivational theories of *desire*, one might think that they do pose serious problems for motivational theories of *affect*.<sup>128</sup> Despite the apparent motivational idleness of such wishes and hopes, they are fully psychologically real and, very often, are quite affectively rich. So – and here is the important point – we seem to understand, via typical, everyday experiences, *what it would be* to feel positively about something without being disposed to do anything with respect to it.

One could deny this, of course. After all, it is probably (ok, definitely) true that *were* I to come to believe that there *was* something I could realistically do to make it likely that I will score the winning goal in a World Cup final, I would be disposed to do that thing, at least if doing it were not too costly in terms of other things that I care about. And this would not be a matter of my acquiring a wholly *new* motivational state, but rather would be the natural and immediate consequence of my pre-existing wish or hope combining with my new belief.

This is fair enough, but we might wonder how effective this response really is. It seems to me that I know for sure that I feel positively towards scoring the winning goal, now, just by reflecting on my feeling. But according to a motivational attitude view of phenomenal valence, it seems, I could not know for sure whether I feel positively, now, towards scoring the winning goal unless I could figure out what kinds of things I would attempt to do in such far-off counterfactual circumstances (i.e. in the bizarre circumstance

---

<sup>128</sup> I would like to thank Sigrún Svavarsdóttir (p.c.) for encouraging me to bring attention to this point, here and below.



where I believed I was in a position to score the winning goal in a World Cup final). Even if Michael Smith (1994: pp. 106-107, 113-114) is right that the epistemology of *desire* is the epistemology of such counterfactuals, the epistemology of *phenomenal valence* does not seem to be.<sup>129</sup>

### Section 3.2: The Past

Setting ordinary kinds of wishes and hopes aside, now consider feelings directed towards things that happened in the past, e.g. feelings of regret or pride. If I feel regret about having been mean to the new kid in elementary school, what behavioral dispositions does that come along with? I will never be in elementary school again and I know it, even more surely than I know that I will never score a goal in the World Cup. To deal with such cases effectively, I think that a motivational attitude view needs to make certain revisions to her account of positive and negative motivation.

Motivation, of whatever kind, is essentially forward-looking. To be motivated to do something is to be set up to behave certain ways in the future. In any case, we should *not* say that my feeling of regret about having been mean to the new kid is a negative orientation to my past behavior in virtue of the fact that I am disposed to try to do what I believe will make it so that I *was* never mean to the new kid. That is perverse. But I do not think that a motivational attitude view is committed to such perversity. We should be a bit looser in our understanding of what it takes to be positively or negatively motivated with respect to something. Perhaps the motivational import of my feeling of regret includes, say, an

---

<sup>129</sup> Smith writes that “desires are states that have a certain functional role...According to this conception, we should think of desiring to  $\Phi$  as having a certain set of dispositions, the disposition to  $\Psi$  in conditions C, the disposition to X in conditions C’, and so on, where in order for conditions C and C’ to obtain, the subject must have, *inter alia*, certain other desires, and also certain means-ends beliefs, beliefs concerning  $\Phi$ -ing by  $\Psi$ -ing,  $\Phi$ -ing by X-ing and so on.” (1994: p. 113) Further, he claims that “the epistemology of desire is simply the epistemology of dispositional states – that is, the epistemology of such counterfactuals.” (ibid)

inclination to prevent or avoid those *kinds* of mean treatments in the future, or an inclination to make *amends* for such past mean treatments.

I am happy to allow these kinds of wiggle room – these indirect, vicarious, or restorative, kinds of tendencies – to be included in our account of positive and negative motivation. Any account of motivational valence will, I suspect, need to appeal to some such maneuvers. These kinds of maneuvers seem preferable in these kinds of cases, anyway, to the alternative, harder-line view that has it that your feeling of regret is motivating in virtue of the fact that it disposes to you to try to do what you believe will bring it about that you were never mean to the new kid in the past.

Nevertheless, it seems to me that phenomenal attitude views give a *better* account of the valence of backward-looking feelings like feelings of regret, better even than the more expansive, softer-line account proposed above. Inclinations to prevent or avoid *kinds* of mean treatments in the future or inclinations to make *amends* for such past mean treatments do not seem like essential components of feelings of regret, but rather like typical effects, the kinds of effects that feelings of regret might have on particularly thoughtful people, say, but not on everyone, and not necessarily. Even if our view of positive and negative motivation is quite expansive, we still see that there can be significant space between affect and motivation.

### **Section 3.3: Action-Thwarting Contents**

The next reason for worry about motivational attitude views is also brought out by perfectly mundane cases. Oftentimes, what we feel positively about is not only something that we know we cannot effect, but is something that, as we might put it, is explicitly our-action-thwarting. Here is Galen Strawson (1994) on one such humdrum case:

When I want Wimple rather than Ivanov to win the World Chess Championship, I do not wish – let alone necessarily wish – that I could affect the outcome. I want something to happen. I do not wish that I could do anything about it. Desire does not necessarily involve the will.<sup>130</sup> (1994: 287)

If we replace ‘want’ and ‘desire’ with ‘feel positively towards’ in this passage, the point becomes even more plausible. I might feel positively toward the prospect of a committee making up their mind in my favor without my interference (Schroeder (2004)<sup>131</sup>), or toward the prospect of the ball on a roulette wheel landing on red fairly, or toward the finicky kitchen sink finally just breaking of its own accord so the landlord will have to fix it. In all such cases, and a million besides, the *content* of our feelings seems to *preclude* – sometimes as a logical matter – action on our parts. Once again we seem in all such cases to understand what it is to feel positively or negatively about such-and-such without motivation entering into our thoughts about the matter.

But as noted earlier, we should often allow that certain positive or negative motivations with respect to p can take the form of indirect or vicarious or restorative behavioral dispositions. But what such dispositions could we appeal to in these cases? Maybe Strawson is disposed to tell people how great he thinks Wimple is. But this response will not apply perfectly generally: a moose might feel positively about the prospect of her young calf figuring out on his own, i.e. without her interference, how to traverse a rocky pass. But she is not disposed to tell anyone anything. Maybe, in Schroeder’s case, you are disposed to pace back and forth in the hallway while awaiting the committee’s decision. But how is this behavior positively valenced *at all*?

---

<sup>130</sup> Strawson’s own view of desire has it that desires are very closely tied to dispositions to experience pleasure, and so to this extent at least he disagrees quite profoundly with the views of desire discussed earlier.

<sup>131</sup> Again, Schroeder is concerned here to argue that you can *desire* such-and-such without being relevantly motivated, whereas I am arguing that you can *feel positively* about such-and-such without being relevantly motivated.

Strawson (ibid), in a notably tongue-in-cheek way, provides a very helpful warning against going too far down the road that we are now on:

When we release a dart or a bowling ball, we know we can do nothing about what happens next. Nonetheless, we may lean to the left as we will our dart or bowl to go to the left, seeking to generate and project leftward-movement-producing energy by a strange sort of psychophysical clenching.

One of the strengths of a motivational attitude view, recall, was that it promised to explain phenomenal valence in a physicalistically respectable way while not losing sight of *valence*, i.e. while satisfying our third criterion. But if we allow behaviors like leaning to the left in the case Strawson discusses, and maybe like pacing back and forth while awaiting the committee's decision, to count as instances of *positive* or *negative* motivation, then we risk undermining motivational attitude views by threatening one of their major selling points.

### **Section 3.4: The Weather Watchers**

Now we will try to make further progress by going farther out into logical space. Strawson also tries to have us conceive of creatures that are not capable of *any* kinds of behavior – or even *attempts* at behavior – *at all*, but (for our purposes) nevertheless seem to have affective experiences.<sup>132</sup> He calls these creatures the “Weather Watchers” and here is how he describes them:

The Weather Watchers are members of a race whose natural course of development leads from an active, mobile youth to a state of immobility, rooted to the ground, in which they retain basic sensory and intellectual capacities. Perhaps they have fixed compound eyes, indistinguishable from the rest of their surface, and look to us like lichen-covered standing stones. Being of a happily adaptive disposition, they progressively lose all desire for what they cannot have – action – and...they retain many desires about how the weather should go and about the naturally shifting scenery... Their memory span is limited, and as time goes on, all their memories of their former capacity for action fade away. Their thoughts fall into different patterns. (1994: p. 254)

---

<sup>132</sup> Strawson actually uses these creatures to show that *desires* need not come along with behavioral dispositions, but I will translate the relevant points into claims explicitly about affect here.

A Weather Watcher, we are to imagine:

sees the sky and hopes the clouds are bringing rain. It watches a seed lodge in a gap between two rocks by the edge of the river. It forms the belief that a tree may grow there before long, and hopes that it will. It sees a second seed settle near by and hopes that it will not germinate, since it dislikes the species of tree in question. (1994: p. 255)

It seems to be possible that such creatures have affective experiences but are not disposed to engage in *any* behaviors that they believe are related to the intentional objects of those experiences, even in the fairly indirect, vicarious, restorative ways described above.

But there are at least three kinds of responses one might make to the Weather Watchers. First, one might worry that, if Weather Watchers do not have the relevant behavioral dispositions, then quadriplegic people or brains in vats could not have those dispositions either, and that is a bad result since those people do (or could) have those dispositions. Second, one could appeal to even more indirect kinds of behaviors, e.g. betting behaviors, to show that the Weather Watchers are still compatible with motivational attitude views (Smith (1998)). Third, one could object that if the Weather Watchers are possible, then we would have to reject any kind of functionalism about affect, and that would be too hard to stomach. We will now discuss each of these responses in turn.

First, there are significant enough differences between the Weather Watchers and, say, quadriplegic people that our claims about the former need not have any interesting implications about the latter. The most interesting feature of Strawson's discussion of the Weather Watchers, for my money, is the "rooting story." As I conceive of it, when they become rooted to the ground, they undergo massive *internal* changes, in some ways just as pronounced as the changes that occur in fetal development. Or perhaps a better analogy is

with the changes that occur e.g. in severe strokes. It seems possible that, beginning at the time they become rooted to the ground, the parts of their brains that are responsible for behavior and even for attempts at behavior stop receiving blood, e.g. as happens in certain kinds of stroke, and so eventually lose all original function. Or, it seems possible that part of the rooting process involves the repurposing of those parts of the brain, along the lines of how – during evolutionary history or during fetal development – features that were once used for one thing (e.g. gills, tails) come to do something entirely different. And it *seems* to be possible that, despite these changes, affect might persist. But, we can suppose, quadriplegic people and brains in vats need not undergo *any* such internal changes. Their brains are (typically) identical to ours and so it seems that they *are* disposed to try to perform behaviors relevant to the intentional objects of their affective experiences – precisely as we are when we have analogous such experiences. Mentally, we might say, they are just like us.

But the interesting thing about Strawson’s Weather Watchers is that they are *not* just like us mentally. All tendencies even to *try* to behave have gone and only passive perceptions and feelings remain.<sup>133</sup> For Weather Watchers to *act*, you would need to add not only “cooperative limbs,” say, as you might with quadriplegic people or brains in vats, but the parts of a *brain* necessary for any kind of behavior or attempts at behavior. But adding these parts would be cheating (Strawson (1994: pp. 272-273)). In “adding” these parts of the brain, we would be adding things “that actually turn the beings under consideration into agents.” And even if it is true that a Weather Watcher *would* be motivated to  $\Phi$  *if we were to turn him into an agent*, that does not imply that any given Weather Watcher *is* motivated to  $\Phi$

---

<sup>133</sup> Again, Strawson claims that the Weather Watchers can have *beliefs*, *desires*, etc. I am not committed to any such claim. Beliefs and desires might be more “active” kinds of states than the Weather Watchers can muster. Perception and feeling, however, seem to be able to be had by very – even entirely – “passive” creatures.

(now, as he is). But quadriplegic people and brains in vats are agents now, they are just unlucky. In any case, there is a sufficiently large gap between the Weather Watchers and quadriplegic people and even brains in vats to see that accepting Strawson's verdict about the former need not force us to accept any repugnant conclusions about the latter.

If this point works, then it also seems to undercut e.g. Michael Smith's (1998) response to the Weather Watchers. He attempts to show that even the Weather Watchers must be disposed to engage in certain kinds of *betting* behavior, at least. Going back to the case of Wimple and Ivanov, Smith writes the following:

A subject desires that Wimple beats Ivanov fairly and squarely only if, in the closest possible worlds in which she has a desire to gamble, and is offered a choice between a gamble in which the pay-off is that Wimple beats Ivanov fairly and squarely and a gamble in which the pay-off is something that she wants less but assigns only a somewhat higher probability, she chooses the gamble in which the pay-off is that Wimple beats Ivanov fairly and squarely...What could we make of a gambler who professes to desire that Wimple beats Ivanov fairly and squarely, but who then refuses such a gamble? (1998: pp. 450-451)

This is an important possibility to consider. Applied to our version of the Weather Watchers, it would say that a Weather Watcher is, insofar as he feels positively about X, disposed to place certain bets if offered. In general, it seems legitimate (all else equal) to expand our account of positive motivation still further to include such (hypothetical) betting behavior. And maybe this can help motivational attitude views respond to the Wimple and Ivanov case.

But it will not help such views to respond to the Weather Watchers. Placing a bet is still a behavior, and to turn the Weather Watchers into beings capable of behaving, we would have to employ the problematic cheat discussed earlier. The appeal to (hypothetical) betting

behavior does not solve this problem (or, if it does, the problem could have been solved more easily).<sup>134</sup>

Finally, we might worry that if we allow that Weather Watchers can have affective experiences – i.e. if we allow that the mental relation *phenomenal valence* can be instantiated in those conditions – then we will be unable to provide a behaviorist or functionalist account of phenomenal valence. Strawson might encourage this response, given that he explicitly conceives of his *Mental Reality* (the book in which the Weather Watchers appear) as a sustained argument against what he calls “neo-behaviorist” accounts of the mind. And there certainly seems to be a tension between some of the things that I, at any rate, would like to say about the Weather Watchers and some of the bolder claims made by certain behaviorists or functionalists. David Armstrong (1993), for example, wrote the following:

The concept of a mental state is primarily the concept of a *state of the person apt for bringing about a certain sort of behaviour*...In the case of some mental states only they are also *states of the person apt for being brought about by a certain sort of stimulus*. But this latter formula is a secondary one. (1968/1993: p. 82 – emphasizes original)

The case of the Weather Watchers – and all of the other cases discussed in this section – suggests that this is incorrect. The *concept* of an affective experience, or the *concept* of phenomenal valence, is not, it seems to me, the concept of a state of a person apt for bringing about a certain sort of behavior.

But, first, it is not obvious that we actually *need* to deny anything Armstrong says here. As just discussed, the concept of e.g. a positive affective experience towards p does

---

<sup>134</sup> Somebody applying Smith’s idea to our topic might have something a bit different in mind here, an idea according to which e.g. the positive valence of a Weather Watchers’ affective experience is not explained by any disposition that *that* Weather Watcher has, but by the dispositions that some *other* creature has – his counterpart, say. This *kind* of idea will be explored further in Section 4 below.



not seem to be the concept of a state that, as such, involves a *disposition* to engage in relevant behavior. But need Armstrong deny this? Perhaps we should understand his appeal to “aptness” differently. Maybe what it is for a mental state to be “apt” for bringing about a certain behavior is, rather, a matter of what behaviors it would bring about *in optimal conditions*.<sup>135</sup> As suggested in the previous chapter, a phenomenal attitude view can – and *should* – accept that feelings and motivation (and evaluation) will align in optimal conditions. If all we need for a physicalistically respectable functionalism about phenomenal valence are *these* kinds of relations between affect and motivation, then our verdict on the Weather Watchers need not pose a problem here.

But if our hand is forced and we must choose between a functionalism about phenomenal valence that does not embrace these subtleties, on the one hand, and what seems like the right thing to say about the Weather Watchers and our other cases, on the other, then we should opt for the latter, even if that raises the stakes of our discussion. As Hume notes:

Love is always follow'd by a desire of the happiness of the person below'd, and an aversion to his misery: As hatred produces a desire of the misery and an aversion to the happiness of the person hated. (1978: II, 2, vi: p. 368)

If we understand his references to love and hatred here as references to affective experiences, and his references to desires and aversions as references to motivational states<sup>136</sup>, then maybe Hume is right: maybe any token instances of love or hatred that we are likely to encounter will be attended by such motivations. Maybe that is even so as a matter

---

<sup>135</sup> Cf. Stalnaker's account of belief presented on p. 120 above.

<sup>136</sup> This is not the place for Hume exegesis, but I suspect that this way of understanding Hume is fairly accurate.

of natural law, say, or of brute metaphysical necessity (what do I know about such things?).

Nevertheless, Hume also plausibly points out that

this order of things, abstractly consider'd, is not necessary. Love and hatred might have been unattended with any such desires, or their particular connexion might have been entirely revers'd. If nature had so pleas'd, love might have had the same effect as hatred, and hatred as love. I see no contradiction in supposing a desire of producing misery annex'd to love, and of happiness to hatred. (1978: II, 2, vi: p. 368)

So even if our response to the Weather Watchers does pose a problem for certain *kinds* of functionalism about the mind, that *is* a problem for those kinds of functionalism about the mind. Even those of us – like myself – who want a unified account of mind and nature need to face up to those problems even if that means tinkering with, or even rejecting, otherwise attractive attempts to find a place for minds in nature.

Another way to respond to our discussion of the cases above – the Weather Watchers in particular – is to note that they show, at most, that it is *conceivable* that affect comes radically apart from motivation. But *this*, as we have known for a long time, does not straightforwardly imply that it is *possible* for affect to come so radically apart from motivation. And, to be fair, I have provided no argument here for thinking that conceivability teaches us anything about possibility in these particular cases. So, for all I have said here, it might be that there are *metaphysically* necessary connections between affect and motivation even if these connections are not *conceptually* necessary. And so any accounts of the mind that only attempt to establish such metaphysical connections, and not such conceptual connections, will – the idea continues – be unaffected by e.g. the Weather Watchers. In particular, if an adherent of a motivational attitude view only wanted to argue that phenomenal valence can be explained in terms of motivational valence in the sense that the former is *identical* to or is

metaphysically *grounded* in or *supervenient* upon the latter, then she need not lose much sleep over the Weather Watchers.<sup>137</sup>

These points are all fair, and well taken. Unfortunately, to say anything of much interest by way of response would require us to wade very deeply into the mind-body problem. That is impossible here (but see the end of Chapter 6 for brief additional discussion). But here is one point worth making.<sup>138</sup> At the end of the day, I am defending phenomenal attitude views of phenomenal valence. Such views appeal to phenomenal properties P and N to unify the positive and negative affective experiences. Now suppose that someone *accepts* that P and N unify the positive and negative affective experiences, but then goes on to claim, *in addition*, that P and N are identical to or grounded in or supervenient upon certain relevant motivational tendencies. It does not seem to me that an adherent of a phenomenal attitude view *needs* to have any quarrel with such a person. My real opponent here is the person who argues that P and N do *not* unify the positive and negative affective experiences and who claims that motivation does so *instead*.

And *this* opponent, I think, is now on the back foot. In *all* of the cases above – not just the Weather Watchers – appeals to P or N seem to provide a more compelling explanation of the phenomenal valence of the relevant subjects' mental states than appeals to the motivational tendencies of their states. This is not nearly the final word on these issues, of course, but it does, I think, give us good reason to think that *insofar as phenomenal attitude views and motivational attitude views disagree*, phenomenal attitude views get the better of the exchange.

---

<sup>137</sup> Thanks are due to Sigrún Svavarsdóttir (p.c.) and Declan Smithies (p.c.) for encouraging me to consider these issues explicitly here.

<sup>138</sup> The following points were inspired by a number of questions raised by Declan Smithies during my dissertation defense, and I thank him for raising those questions.

### Section 3.5: Feeling Motivated vs. Being Motivated

All of that said, the Weather Watchers are *very* bizarre. Maybe, as just noted, they are even *impossible* (even if they are conceivable). And we should be wary about drawing significant conclusions – e.g. about the truth or falsity of motivational attitude views of phenomenal valence – on the basis of intuitions about such strange cases. So, in this section, I will present a more general reason for thinking (modulo the points made at the end of the previous section) that affect need not involve motivation, a reason that does not depend directly on our intuitions about cases as bizarre as the Weather Watchers.

The idea that affect implies relevant motivation seems most plausible when we consider a certain subset of affective experiences, i.e. the affective experiences that, as we might put it, *feel motivating*, i.e. the ones that make you feel energized, up for it, pushed, pulled, drawn, repelled, twitchy, ready, prepared for action, etc. How, we might ask ourselves, could *cravings* and *yearnings* fail to come along with dispositions to act? But here it is worth making two points.

First, not all affective experiences feel motivating in these ways. Many feelings just wash over us and seem to call for nothing at all, e.g. the feeling of lazy satisfaction about having eaten a large meal. So even if all craving and yearning experiences were motivating, we have no particular reason in advance to expect that these other kinds of affective experience are motivating. And, second, even if it were true (though it is not) that all affective experiences *felt* motivating, there is in general a distinction between *feeling* X and *being* X, e.g. feeling hungry or thirsty vs. being hungry or thirsty<sup>139</sup>, feeling like you are hurt, or spinning in circles, or like you are not making sense vs. being hurt, spinning in circles, or

---

<sup>139</sup> The idea is that we can understand being hungry or thirsty e.g. as your body needing nourishment or water so as to be able to continue to perform its functions. Thanks are due to Sigrún Svavarsdóttir (p.c.) for urging me to get clear about this.

not making sense. If such distinctions apply in general, then why not think they apply here, too? Why couldn't you feel motivated without being motivated? Why in the world should we believe that *feeling as though* you are disposed to do what you believe will result in your  $\Phi$ -ing<sup>140</sup> means that you *really are* disposed to do what you believe will result in your  $\Phi$ -ing, given, as we have just seen, that this kind of inference is in general a bad one? Given that this kind of inference is in general a bad one, we should expect in advance that it will also be problematic in this case.<sup>141</sup>

These two points – (a) that not all affective experiences “feel motivating,” and (b) that even those that do are subject to the general distinction between feeling that you are some way and your being that way – do not refute motivational attitude views. But I think they go a long way towards undercutting whatever introspective, first-personal appeal they might have had. And they do so without relying on anything nearly as unusual as the Weather Watchers.

## Section 4: Failing the Third Criterion

### Section 4.1: A Brave New World

---

<sup>140</sup> This might not be an apt description of any of *your* affective phenomenology, but Armstrong (1993: pp. 134–135) describes the experience of *intention* in the terms quite like this: “Suppose I form the intention to strike somebody. My mind is in a certain state, a state that I can only describe by introspection in terms of the effect it is apt for bringing about: my striking that person...My direct awareness of this mental cause is simply an awareness of the sort of effect it is apt for bringing about...We are directly, that is to say, non-inferentially, aware of something ‘apt for bringing about the striking of X’.”

<sup>141</sup> There is also some empirical work that might suggest that affect and motivation can come apart fairly radically. Nomy Arpaly & Tim Schroeder (2014: p. 118) point out that some “drugs...tend to flatten impulses to act, but not in lock-step with the flattening of feelings. Someone who, on a high dose chlorpromazine, is no longer disposed to feel much of anything in response to eating, walking, or talking can nonetheless remain disposed to eat, walk, and talk.” And Lisa Barrett and her colleagues (2007: p. 379) have found, as they put it, that “mental representations of emotion often, but not always, include some arousal-based content (i.e., feeling as if the mind or body is active, as in aroused, attentive, or wound-up, versus feeling that the mind or body is still, as in quiet, still, or sleepy). Felt activation is typically related to, *but does not have a one-to-one correspondence with*, actual physiologic activity.” (emphasis added) Such findings are merely suggestive for our purposes, but they are suggestive.

Suppose that we have now seen that it is *not* a necessary truth that all positive (negative) affective experiences have a positive (negative) motivational profile with respect to their intentional objects. If that is not a necessary truth, then it seems to me that motivational attitude views are inadequate accounts of phenomenal valence because they fail to satisfy our first criterion (i.e. they are extensionally inadequate). In this section, however, I would like us to consider views that bear a strong family resemblance to motivational attitude views of phenomenal valence, at least, but that could probably be developed in such a way that they *are* extensionally adequate. The views to be considered in this section attempt to show that phenomenal valence *still* might be explicable in terms of motivational valence, *even if any token affective experience might occur without relevant motivation*.

Even if such views will not end up counting as motivational attitude views, strictly speaking, they would still, if correct, undercut phenomenal attitude views substantially. That is why they are worth addressing here.<sup>142</sup> What we will learn, however, is that they buy extensional adequacy at the cost of failing to satisfy our third criterion. That is to say, the properties that are said to unite the positive (negative) affective experiences, are *not* properties the possession of which are suitably related to (more strongly: explanatory of) the *valence* of the experiences (that is, these views do *not* enable us to see, upon reflection, even with coaching, how it is that someone's instantiating the relevant properties amounts to that person's being *for* or *against* anything.)

## **Section 4.2: Representing Motivation**

The first such attempt that I would like to explore is inspired primarily by the relations between *pleasure* and *desire* hypothesized by Tim Schroeder (2004) and Nomy Arpaly & Tim

---

<sup>142</sup> I would like to thank Sigrún Svavarsdóttir (p.c.) for encouraging me to be explicit about how (and where) these views fit into the broader dialectic of this dissertation.

Schroeder (2014).<sup>143</sup> Arpaly & Schroeder’s idea, very roughly, is that feelings of pleasure *represent* certain features of how things stand with our desires:

Pleasure and displeasure tell us about changes in intrinsic desire satisfaction. They form a kind of sense modality: an interoceptive modality, allowing us to sense facts about what is going on inside us regarding our intrinsic desires (a sense more like our sense of limb orientation—proprioception—than like vision, in this respect). And because of this, it is probably the case that any creature capable of pleasure and displeasure is one that has intrinsic desires. But intrinsic desires are not *made up* of these feelings as a result, because the feelings *represent* facts about intrinsic desires. (2014: p. 119 – emphases original)

I will not discuss further Arpaly & Schroeder’s motivations for such a view, nor their fascinating further elaborations of it. What I am interested in, rather, is how an *analogous* view might be applied in such a way as, potentially, to allow us to explain phenomenal valence in terms of motivational valence.

The analogous idea is this: phenomenal valence is to be explained by the fact that affective experiences *represent* our positive or negative motivations. Suppose that I feel positively about going to the concert in two days time. The first part of the idea to be explored here has it that this feeling represents (in a truth-apt, mind-to-world way) that I am positively motivated with respect to going to the concert in two days time.<sup>144</sup> Let us call this the “Representation Claim.” The second part of the idea is that my feeling is *positively valenced*

---

<sup>143</sup> We also discussed similar kinds of views towards the end of Section 3.3 of Chapter 2.

<sup>144</sup> Perhaps there are two basic kinds of positive (negative) affective experiences: feeling attracted to (averse to) X, and feeling pleased that (pained by) X. The view now being considered might try to explain the relevant similarities and differences between these sets of experiences in the following way. Perhaps when I feel attracted (averse) to p, I represent that I am positively (negatively) motivated with respect to p, but when I feel pleased (pained) that p, I represent that I am positively (negatively) motivated with respect to p, *and p*. So feeling pleased (pained) that p involves the additional representation of one’s motivation being “satisfied.” I see this as a very promising area for further research. See Section 5 below for brief discussion of potentially related issues.

in virtue of the fact that it represents *positive* motivation (and the same would apply, *mutatis mutandis*, to *negative* feelings and *negative* motivation). Let us call this the “Valence Claim.”<sup>145</sup>

The Representation Claim might well be true. Our arguments (in Section 3.3 of Chapter 2) against the idea that e.g. feelings of fear represent only internal changes in the fearful person were not exactly arguments for thinking that feelings of fear *do not* represent such changes. Rather, they were only, in the first instance, arguments against thinking that they *only* represent such changes. If affective experiences are in the business of “giving us information,” say, then it is not so implausible to think that the information that our feelings give us is precisely information about what we are now inclined to do or forego, as the Representation Claim has it. And a further, major benefit of this approach is that it is consistent, at least in principle, with what we wanted to say e.g. about the cases discussed in Section 3 above. Just as one can represent that a unicorn is nibbling the grass on one’s lawn despite the fact that there is *not* a unicorn nibbling the grass on one’s lawn, so one can represent (via an affective experience or in any other way) that one is disposed to do what one believes will bring it about that one  $\Phi$ s without it *being the case* that one *is* disposed to do what one believes will bring it about that one  $\Phi$ s. Misrepresentation is *always* possible. So, even if affective experiences represent motivation – even if it is a *necessary* truth (even a *conceptually* necessary truth) that affective experiences represent motivation, and so the view can establish a perfectly general connection between phenomenal valence and motivation, and so the view can, in *that* way, satisfy our first criterion – that is perfectly consistent with it

---

<sup>145</sup> If one embraces this conjunction of claims, then there is a clear sense in which one embraces a *content* view instead of an *attitude* view of phenomenal valence: phenomenal valence would be explained by its representing a certain content, and would not be explained by the attitude the experience has us take towards that content. But leaving it there would be misleading, since the content represented would only *itself* explain phenomenal valence in virtue of the fact that that content refers to a *pro-attitudes* of the agent. We will return to such issues momentarily.



not being a necessary truth that if one feels positively about  $\Phi$ -ing, then one is positively motivated with respect to  $\Phi$ -ing.

However, even if the Representation Claim is true, the Valence Claim is certainly false.<sup>146</sup> Even if affective experiences represent relevant motivation in the ways suggested, that does not allow us to explain phenomenal valence in terms of motivation. Such a view fails to satisfy our third criterion. The argument for this conclusion is quick but, I think, decisive:

- (1) If the Representation Claim entails the Valence Claim, then our *beliefs* about our motivations are valenced states.
- (2) Our beliefs about our motivations are not valenced states.
- (3) So, the Representation Claim does not entail the Valence Claim: that is, even if affective experiences as such represent relevant motivation, that does not allow us to explain phenomenal valence in terms of motivational valence.

First, if the Representation Claim allows us to explain phenomenal valence in terms of motivational valence, then it looks like we must also embrace the following general principle, on pain of being desperately ad hoc: if a state,  $S$ , represents that one is motivated positively with respect to  $\Phi$ , then that state,  $S$ , is *itself* a positively valenced state with respect to  $\Phi$ . How else *could* the Representation Claim allow us to explain phenomenal valence? But if that general claim is true, then it looks like our *beliefs* about our own motivations – e.g. my belief that I am disposed to do what I believe will bring it about that I finish this chapter tonight – are also ways of being positively or negatively oriented towards things – in this case, my *belief that I am motivated to finish this chapter* is itself a way for me to be for or in favor of *finishing this chapter*.

---

<sup>146</sup> I suspect that even if the Representation Claim is true, it is at most contingently true (why should we believe that it is a necessary truth?).

But this is false. This belief is not a way for me to be for or pro or in favor of anything at all. This comes out most clearly when such beliefs are false (although that does not really make a difference). Suppose that my belief above is false. That is, suppose that I am not at all motivated to finish this chapter tonight, even though I believe that I am so motivated. In this case, given all that we know about me, I seem to be *perfectly indifferent* to finishing this chapter. Unless we add information – e.g. that I *feel* positively about finishing the chapter, or that I think that it would be *good* if I finished the chapter – it seems like there is no reason whatsoever to think that I care one jot about finishing this chapter. I might regard my (putative) motivation as an alien quirk, for example. Or suppose that not only am I not motivated to finish this chapter, but that I am strongly motivated to put it off, I feel nothing but contempt for it, and I believe that working on it is an utter waste of my time. If you were to try to figure out where I stood with respect to my finishing this chapter, you should not conclude “well, he’s antagonistic to it in some pretty serious respects, but at least he’s in favor of it in this way: he believes that he’s motivated to finish it.” The second premise of the argument above, then, is in good shape, and so this way of explaining phenomenal valence in terms of motivational valence will not work.

### **Section 4.3: Mad Affect**

Now I would like to consider a second view that also attempts to explain phenomenal valence in terms of motivational valence in a roundabout way. This view is inspired by David Lewis’s (1980) account of “mad pain.” Lewis has us consider “a strange man who sometimes feels pain, just as we do, but whose pain differs greatly from ours in its causes and effects.” (1980: p. 216) Focusing on the strange effects of this character’s pain, Lewis writes that whereas

our pain is generally distracting...his turns his mind to mathematics, facilitating concentration on that but distracting him from anything else. Intense pain has no tendency whatever to cause him to groan or writhe, but does cause him to cross his legs and snap his fingers. He is not in the least motivated to prevent pain or to get rid of it. In short, he feels pain but his pain does not at all occupy the typical causal role of pain. (ibid)

According to Lewis, mad pain is possible (ibid). And we can already begin to see how this discussion might be relevant for our purposes. In some ways, Lewis's mad pain character might resemble the Weather Watchers.

But Lewis is one of the 20<sup>th</sup> century's great functionalists about the mind. He writes (and compare the Armstrong passage quoted earlier) that

the concept of pain, or indeed of any other experience or mental state, is the concept of a state that occupies a certain causal role, a state with certain typical causes and effects. It is the concept of state apt for being caused by certain stimuli and apt for causing certain behavior. Or, better, of a state apt for being caused in certain ways by certain stimuli plus other mental states and apt for combining with certain other mental states to jointly cause behavior. (1980: p. 218)

How can he square these claims with the possibility of the madman? The madman seems precisely to be in pain, despite the fact that his pains do *not* occupy the typical causal role of pain, are not apt for being caused by, or causing, the usual things, etc.

In an attempt to square this circle, Lewis writes that

[The madman] is in pain, but he is not in a state that occupies the causal role of pain for him. He is in a state that occupies the causal role of pain for most of us, but he is an exception. The causal role of a pattern of firing of neurons depends on one's circuit diagram, and he is hooked up wrong...The thing to say about mad pain is that the madman is in pain because he is in a state that occupies the causal role of pain for the population comprising all mankind. He is an exceptional member of that population. The state that occupies the role for the population does not occupy it for him. (1980: p. 219)

The madman is in pain, then, but only because his experience is a token of a type *other* tokens of which (the "normal" ones?) play a certain causal role.

This is an easy enough view to state, but one that is very difficult to understand. The question that generates my confusion is the following: *in virtue of what* is the madman's state a pain? On the one hand, as we noted above, according to Lewis,

the concept of pain, or indeed of any other experience or mental state, is the concept of a state that occupies a certain causal role, a state with certain typical causes and effects. It is the concept of state apt for being caused by certain stimuli and apt for causing certain behavior. Or, better, of a state apt for being caused in certain ways by certain stimuli plus other mental states and apt for combining with certain other mental states to jointly cause behavior. (1980: p. 218)

Now, when we try to apply this to mad pain, it looks like our answer to the question above is something like this: the madman's state is a pain in virtue of the fact that it is a state of a *type* that is apt for being caused in certain ways by certain stimuli plus other mental states and apt for combining with certain other mental states to jointly cause behavior *in the madman's population at large*. Call this the "communal-functional answer" to our question.

But Lewis also says the following very plausible things:

Pain is a feeling. Surely that is uncontroversial. To have pain and to feel pain are one and the same. For a state to be pain and for it to feel painful are likewise one and the same. A theory of what it is for a state to be pain is inescapably a theory of what it is like to be in that state, of how that state feels, of the phenomenal character of that state. Far from ignoring questions of how states feel in the odd cases we have been considering, I have been discussing nothing else! (1980: p. 222)

A natural understanding of this passage suggests a (conceptually) distinct answer to our earlier question: the madman's state is a pain in virtue of the fact that it *feels* a certain way (the pain way). Call this the "phenomenological answer."

Lewis would not (or at least should not) rest content with these two, seemingly distinct answers to our question. Instead, he would (or should) try to show how these two answers amount to the same thing, in some important sense, at the end of the day. I see two

basic ways for him to do so, one of which is much more implausible than the other. The less implausible one is the one that I would like to use as a model for our purposes below.

The very implausible reconciliation goes as follows: the *phenomenology* of the madman's state (how it feels) is determined by the fact that it is a token of a type of state that plays a certain causal role in the madman's population generally (though not for the madman). Another way to put this proposal is that the phenomenology of the madman's state constitutively depends on facts outside of the madman's brain. But Terence Horgan & John Tienson (2002: pp. 526-527), it seems to me, provide a decisive objection to any such view. Here I quote them at length:

Phenomenology does not depend constitutively on factors outside the brain. Now, it is obvious enough that in normal humans, phenomenology does depend *causally* on some such factors; but one need only consider how this causal dependence works in order to appreciate the lack of constitutive dependence. First, phenomenology depends causally on factors in the ambient environment that figure as *distal* causes of one's ongoing sensory experience. But second, these distal environmental causes generate experiential effects only by generating more immediate links in the causal chains between themselves and experience, viz., physical stimulations in the body's sensory receptors—in eyes, ears, tongue, surface of the body, and so forth. And third, these states and processes causally generate experiential effects only by generating still more immediate links in the causal chains between themselves and experience—viz., afferent neural impulses, resulting from transduction at the sites of the sensory receptors on the body. Your mental intercourse with the world is mediated by sensory and motor transducers at the periphery of your central nervous system. Your conscious experience would be phenomenally just the same even if the transducer-external causes and effects of your brain's afferent and efferent neural activity were radically different from what they actually are—for instance, even if you were a Brain in a Vat with no body at all, and hence no bodily sense organs whose physical stimulations get transduced into afferent neural inputs. Among your logically possible phenomenal duplicates, then, are beings whose sensory experience is radically illusory, in the manner of the famous Evil Deceiver scenario in Descartes' First Meditation—or its contemporary version, the Brain in a Vat.

For these reasons, we should look for some other way of reconciling the communal-functional answer to our question with the phenomenological answer.<sup>147</sup>

A better answer has us amend the phenomenological answer. On the view I recommend (for Lewis and for anyone else who is attracted by the communal-functional answer to our question), the *phenomenology* of the madman's state does *not* depend on the relevant facts about his community *at all*. The phenomenology of his state, we should suppose (in keeping with the lesson from Horgan & Tienson), is fully determined by states or activities of his central nervous system. Let us say that his state has phenomenal character C1. C1, let us suppose, is identical to the phenomenal character of (some of) our pain experiences. Now here is the move: it is *not* in virtue of the fact that the madman's state has C1 that it is a pain. What *makes* his state *a pain* are the facts about the causal roles that other tokens of that state-type play in his population. Another way to put it is like this: whether or not a mental state *is a pain* does constitutively depend on factors outside of one's brain (e.g. on facts about one's community). But since phenomenology cannot constitutively depend on factors outside of one's brain, whether or not a state is a pain does not constitutively depend on its phenomenology.

What this view ends up allowing is the possibility that (a) a state could be phenomenally identical to a pain state, and yet not *be* a pain state. This might strike you as implausible. You might think that if something feels exactly like a pain, then it is a pain. This is not a crazy idea. But to maintain it, in the light of what we have seen, you either have

---

<sup>147</sup> Some philosophers accept that phenomenology can constitutively depend on factors outside the brain. These are the "phenomenal externalists" (see e.g. William Lycan (2001) for a notable defense of a view of this kind). I take the passage quoted above from Horgan & Tienson decisively to refute any such view. Nevertheless, I will simply note that if an adherent of a motivational attitude view needs to embrace phenomenal externalism in order to get her view to work, that saddles her with substantial theoretical burdens.

to (b) deny that mad pain is possible (i.e. because no state that was apt to play the “mad” causal role could *feel* like a pain), or you have to (c) allow that phenomenology can constitutively depend on factors outside the brain (e.g. on facts about one’s community), or you have to (d) allow that whether or not something is a pain does not constitutively depend on its causal role (not even on the causal role typical of the type of state it is). Given all we have seen in this chapter, (b) seems implausible, and Horgan & Tienson make a compelling case against (c). Even if (d) is the best way to go at the end of the day, I take it that (a) is at least worth taking seriously at this point, given our options.<sup>148</sup>

Finally, then, we have a model for the type of view of phenomenal valence that I want to consider. According to the view I have in mind, whether or not a creature’s experience is a positive or negative orientation towards what it is about – i.e. whether an experience has phenomenal *valence* – depends on whether it is a token of a type that, in some relevant population, comes along with, or is apt to come along with, positive or negative motivation. More simply (but less accurately), the idea is that a state has phenomenal valence only if it is a token of a type whose tokens are motivating *for us*, say.

On the (a)-type interpretation of Lewis presented above, whether or not something is a pain does not depend on that thing’s phenomenology. The idea here about affect is similar: whether or not something is a positive or negative feeling does not depend on that

---

<sup>148</sup> Declan Smithies (p.c.) has suggested to me an objection to Lewis’s account, as understood by way of (a), based e.g. on David Chalmers’ (1996) idea that mental terms “live” a “double life.” The basic idea is that we have a phenomenal concept of pain that is distinct from any functionally defined concept of pain. Now, we can certainly define a functional concept of pain on which playing functional role R is necessary and sufficient for being pain, and so by that definition, a state can feel just like pain without being pain. But, Smithies reminds us, we also need to recognize a phenomenal concept of pain on which feeling just like pain is necessary and sufficient for being pain. And since Lewis cannot plausibly deny that we have such a concept, if he takes option (a), then – according to Smithies – he’s just changing the subject. This might well be a decisive objection to any view like (a), but I am inclined to give Lewis more leeway here. Even if he is not changing the subject in quite this way, I will argue below that this kind of view is inadequate.

thing's phenomenology. Instead, it depends on the motivational import of other tokens of states of that kind. What makes it the case that the Weather Watchers' experiences have phenomenal valence, say, is not the fact that they have experiences that are phenomenally identical to our affective experiences (though they do, we can imagine). Rather, it is the fact that the relevant experiences are the kinds of experiences such that e.g. when *we* undergo them, *we* are relevantly motivated. A bit more specifically, the idea is that such experiences are tokens of *attitude*-types whose tokens are or tend to be motivating in the relevant population (whatever that population is<sup>149</sup>).

The most interesting thing about this proposal is that it is probably extensionally adequate in its way. What I mean by that is that it will very likely be true that any feeling that is intuitively a positive or negative feeling about such-and-such will be the kind of feeling that is such that, when *we* have it, say, it comes along with relevant motivation. It is difficult to imagine an experience that is clearly a positive (negative) feeling about something that is not like this, at least if we think that e.g. feelings directed towards impossible or necessary intentional objects, or feelings directed towards things happening without our interference, can be relevantly motivating (at least in the ways discussed above). But as I will now argue, even if this proposal is extensionally adequate, and so meets our first criterion, it still cannot enable us adequately to explain phenomenal valence in terms of motivational valence.

---

<sup>149</sup> Declan Smithies (p.c.) suggested to me the following serious objection to the proposal just floated. What makes it the case that the Weather Watchers' experiences are of the same kind as ours? Not the fact that they play the same motivational role, since they do not play any motivational role at all. Rather, Smithies plausibly suggests, it is the fact that they have the same phenomenal character as ours. But now it is not clear that their experiences inherit their valence from ours rather than simply having it in virtue of their phenomenal character. And so, in this way, the idea that this kind of view promises to explain phenomenal valence in terms of motivational valence in *any* way is rendered extremely dubious. A first blush response to this objection is to note that maybe the Weather Watchers have brains like ours, and it is in virtue of that that they count as members of our community. But this carries little weight, especially when we remind of ourselves of the point made earlier that the Weather Watchers are most plausibly possible when we imagine that their "brains" are *very* different from our own!



To see why, we should first look at an analogous view in the metaethical literature. This is the view that e.g. Jon Tresan (2009) calls “communal internalism.” This view claims that

moral beliefs in a community entail [pro- and con-] attitudes in that community, but not necessarily in every member of the community who has moral beliefs. Communal internalists say that moral beliefs require the characteristic moral practices of socialization, norm-enforcement, and self-guidance, but once such practices are up and running, moral beliefs may be acquired by individuals who do not themselves participate in such practices. (2009: p. 180)

The important feature of this view for our purposes is that it allows that particular individuals – “mad” individuals, we might call them – can have moral beliefs without being motivated in accordance with them, but what *makes it the case* that these are bona fide moral beliefs is that they are tokens of a type of mental state, other important tokens of which in the community *are* attended by relevant motivations in those who token them.

In these ways, at least, this view is very much like Lewis’s account of mad pain (as interpreted via (a) above) and the view of affect currently under consideration. And there is something incredibly unsatisfying about all of these views. Whether or not someone is in pain, whether or not someone feels positively or negatively about something, whether or not someone thinks that such-and-such is right or wrong – these seem like very important questions to ask and answer about a person. The answers to *each* of these questions, for example, seem to affect what it is rational for the relevant person to do, what is in their interest, whether and to what extent the person is virtuous, and how it is morally permissible to treat the person. But the distant connections to motivation posited by the kinds of views now under consideration seem not to make *any* difference whatsoever in *any* of these ways.

Focusing only on the case of affect for a moment, presumably it is *prima facie* seriously wrong, as such, for you to make me do things that I feel strongly negatively about doing, and it is *prima facie* seriously wrong *precisely because* I feel strongly negatively about doing them. In this way at least, phenomenal valence seems to be very morally important. But it is not even *prima facie* seriously wrong, *as such*, for you to make me do things that *other people* in my community are motivated to avoid, reduce, prevent, etc. These facts about communal motivation seem to make no difference at all (at least no difference of this kind) to how you should treat me. But the valence of my affective experiences does make such a difference. This suggests that phenomenal valence is something over and above what we might now call “communal motivational valence.”

And lest we think that this is special pleading on behalf of a phenomenal attitude view, *very* similar points apply to the case of pain, and slightly less similar points apply to the case of moral beliefs. For example, I display a certain lack of virtue – call it a virtue of being principled, or being morally strong – when I act against my all things considered moral beliefs, *precisely because* I act against my all things considered moral beliefs. But I do not display *any* lack of virtue, *as such*, in acting against a belief that has the property of being such that, when some other people in my community have tokens of that type of belief, they are motivated in accordance with it. This suggests that whatever is special (if anything) about moral beliefs is not nearly captured by the thesis of communal internalism.

The point of this discussion is to suggest that even if these communal internalist views are extensionally adequate, they do not – or at least the affective version does not – explain what it was being enlisted to explain. Communal motivational internalism about phenomenal valence loses the primary potential virtue of motivational attitude views:

namely, that they seem capable of satisfying our third criterion from above and thus enabling us to *understand* phenomenal valence. And so I conclude that we still have not been given an adequate recipe for explaining phenomenal valence in terms of motivational valence.<sup>150</sup>

### Section 5: Conclusion

Luckily, to my mind, the arguments of this chapter do not imply that there are no necessary or a priori knowable connections between affective experiences and motivation. As we have pointed out many times throughout the dissertation, positive feelings and positive motivations are both, as we might say, accountable to the good. When something is good, it would be fitting (correct, apt, etc.), at least all things being equal, for you to feel positively about it and to be positively motivated with respect to it. These claims seem true, and there is nothing contingent or a posteriori about the connections between affect and motivation that they posit. Further, these claims are fully available to a phenomenal attitude view of phenomenal valence.

We saw in the previous chapter that evaluative content views are the most promising kind of content view, but also that they are inadequate. So, it looks like we are forced to embrace an attitude view of phenomenal valence. It seems to me as though there are two plausible species of this genus, two attitude views that have a hope of satisfying our third criterion: phenomenal attitude views and motivational attitude views. If the latter will not

---

<sup>150</sup> I would also like to note that if the points just made suffice to cause problems for the views just considered, then they should also apply, with equal force, to analogous claims one might make about the (putative) evaluative contents of affective experiences. Suppose one argued, for instance, that even though particular *token* affective experiences need not be or come along with representations of relevant values, nonetheless, any given experience is an *affective* experience (i.e. has affective valence) only in virtue of the fact that it is a token of a *type* of state that normally (in the relevant community, etc.) is or comes along with representations of the relevant values. First, this communal internalist thesis is even less plausible than the others considered in the main text – I am not even sure that I understand it. Second, even if it makes sense, it is subject to analogues of the objections made above. I leave the following as an exercise for the reader: how in the world could *this* kind of internalist thesis help us to make *any* headway in understanding e.g. how affective experiences justify evaluative beliefs (see previous chapter)?

do, and it now seems like they will not, then, I tentatively conclude that we should conclude with the former.

Briefly, however, I would like to consider one final response that a skeptic about phenomenal attitude views might propose. One might still think that, even though there are cases of non-motivating affect and there are cases of non-evaluating affect, there are *no* cases of *neither*. For example<sup>151</sup>, perhaps one accepts that (a) feeling pleased (pained) that p and (b) feeling attracted (averse) to p are *both* kinds of affective experiences, i.e. they both have phenomenal valence. One might also think, for example, that all (a)-type experiences are representations of value, but that they need not be motivating, and that all (b)-type experiences are motivating, but that they need not represent values. If one embraced this (not absurd!) series of claims, then one could allow that there are many cases of affect without motivation, and many cases of affect without evaluation, without allowing that there are *any* cases of affect without motivation *or* evaluation. And if this series of claims is true, then it is far from obvious that we need to countenance phenomenal valence. We could then make do with what seem to be the better-understood notions of evaluation and motivation. We can call this the “divide and conquer” strategy.

My response to this strategy is, in a way, the topic of the next chapter. There I will argue that – supposing that one can represent that A is valuable *and* be positively motivated with respect to A without one feeling positively towards A – positive evaluation and positive motivation, even in tandem, are not sufficient for love. What we should conclude from this, I think, is that the apparent promise of the divide and conquer strategy is merely apparent. Evaluation and motivation just cannot “get us” all the things that affect can get us.

---

<sup>151</sup> What follows is just one possible elaboration of the view I would like to consider

Explanations of phenomenal valence that only avail themselves of motivation and evaluation fall short, and so, at last, we can reasonably conclude, a phenomenal attitude view is our best bet.

## Chapter 5: Affect, Love, and Importance

### Section 1: Introduction

In this chapter I will argue that feeling is necessary for *love*. This claim, on its own, does not favor one view of phenomenal valence over another. But one thing that we will see in this chapter is that evaluation and motivation, individually or in combination – unless they come along with feeling – are not sufficient for love. This claim, I think, *does* point to problems with evaluative content views and motivational attitude views of affective valence. It does not refute them outright. But it suggests, as we might put it, that such views look in the wrong place, that such views are misguided.

Feeling is not necessary for love, say, because of any extra motivational oomph that feeling provides. Feeling is not necessary for love because of any extra values it makes manifest to us, or because it makes those values manifest to us in a special way, either. Instead, as a phenomenal attitude view would have it, feeling is necessary for love because of the phenomenally valenced attitudes it includes. These thoughts – in combination with the points made in previous chapters – give us sufficient reason to conclude that there is room for, and a need for, phenomenal attitude views of phenomenal valence, contrary to what the divide and conquer strategy from the last chapter would have us believe. On these bases, I conclude that phenomenal attitude views of phenomenal valence are correct.

### Section 1.1: Overview

We can love any number of things, e.g. works of art, gods, games, nations, causes, ideals, landscapes, pets, plants, and, of course, people. And our love for people, at least, can come in several different varieties, e.g. romantic, familial, friendship-style, etc. and can be requited or unrequited. Is there any unity behind this diversity? All can agree that there are some very general features shared among all such cases. First, for A to love B is for A to be, at least in part, positively oriented towards B. We can hate, resent, and wish ill upon what we love, but we cannot only hate, resent, or wish ill upon what we love. However, not all positive orientations towards things amount to love. We like, want, enjoy, pursue, protect, respect, admire, appreciate, value, and care about many things in this world. Of those myriad things that we favor, in one way or another, there are many that we do not love. What marks the difference? We might try saying, second, that for A to love B is, at least in part, for B to be particularly important to A. We favor many things, in one way or another, that are not especially important to us.

But things get more difficult once we move beyond these truisms. The more difficult question that I would like to ask here is whether *feeling* is essential to love: can A love B without A tending to feel positively towards B (or at least tending to feel negatively towards B's faring poorly, etc.)? Here I will argue that feeling is essential to love, that A cannot love B without A tending to feel positively (etc.) towards B.

I will regard it as a serious point in favor of a particular view of love if what it says that love is seems, on reflection, to be the kind of thing that would answer to our desires to love and to be loved, inchoate as these desires typically are. And on the other hand, I will regard it as a serious point against a particular view of love if what it says that love is does not seem to be, on reflection, the kind of thing that would answer to our desires to love and

to be loved, inchoate as these desires typically are. As an example of how this works, suppose I told you that for A to love B is for A to have a mild, fleeting inclination to help B flourish. You might reasonably point out that *whatever* it is that you want when you want to love and to be loved, it is something more than that. And that suggests that mild, fleeting inclinations are not enough for love. Or suppose I told you that for A to love B is for A to crave B's company like an addict craves his next fix, a craving that entirely swamps A's reason and makes him utterly a slave of his passions. Here too, you might point out that *whatever* it is that you want when you want to love and to be loved, it is not *that*. Whether or not an account of love fits in these ways with our desires to love and to be loved might not be the only thing to consider, of course. But it is one thing to consider and, I think, a very important thing to consider. And when we consider it carefully, we learn that feeling is necessary for love.

In the rest of Section 1, I will explain what I mean when I say that love requires feeling. Then I will briefly situate the claim that love requires feeling in the contemporary philosophical landscape. As we will see, some recent philosophers have denied that love requires feeling, and even those who have accepted it have not done enough to defend it. Then, having finished with preliminaries, we will make the case in Sections 2 and 3 that feeling is necessary for love. In these sections I present and discuss a series of cases. What these cases seem to show is that, roughly, love without feeling is very disappointing. When we want to be loved, we would be crushed to discover that our putative lover feels nothing for us, or that he feels about us the same way he does about strangers. When you want to love something or someone, it is very disappointing to discover that you feel nothing for it or him. And it will turn out that these points suggest that love requires feeling. In Section 4,



the concluding section, I point to some important lessons that we will have learned here, particularly – but not only – those having to do with phenomenal valence.

### **Section 1.2: Love Requires Affective Dispositions**

I will argue here that love requires feelings, i.e. affective experiences. But the claim that love requires feelings should be understood fairly weakly for present purposes. First, it should not be understood as the implausible claim that, for A to love B at t A has to feel positively about B at t. We continue to love things and people when we are asleep, distracted, and depressed, i.e. when we feel nothing or when our affective systems are otherwise engaged. Rather, it should be understood along dispositional lines: the claim that I will call “love requires affective dispositions” (L-RAD) says that for A to love B, A must be disposed to feel positively towards B. Second, we should be permissive as to what counts as feeling positively towards B. After all, someone might be disposed to feel positively towards positive changes in B’s *well-being*, or be disposed to feel negatively towards negative changes in B’s well-being (cf. Robert Nozick (1993)). Or, someone might be disposed to feel positively towards being *guided* by B (cf. Michael Smith (2015)). For present purposes, we should allow people with any such dispositions to count as satisfying the necessary condition asserted by L-RAD.<sup>152</sup> Our interest here is in the general idea that there is a necessary, non-normative connection between A’s loving B and A’s having relevant B-related affective experiences. The finer details of more specific proposals need to be addressed elsewhere.<sup>153</sup>

---

<sup>152</sup> The kinds of wiggle room I am asking for here with respect to what counts as positive (negative) feelings towards B are exactly analogous to the kinds of wiggle room I allowed to my opponents with respect to what counts as positive evaluation (see note 71 above) and, especially, positive motivation (Chapter 4). So I assume that none of my opponents will begrudge me these kinds of wiggle room here. In any case, none of this will significantly affect the arguments to follow.

<sup>153</sup> I will also not discuss the *specific* phenomenology of love here, e.g. whether love is warm or cool, more like pleasure or attraction or awe, etc.

### Section 1.3: Friends and Foes

It might seem like I am pushing on an open door here in suggesting that love requires feeling, given that this claim is understood in this weak, general way. The claim that love requires feeling might strike one as more or less obvious or, at any rate, as not worth arguing for. But several important theories of love have been proposed that *deny* that affect is necessary for love. If love is an evaluative belief, as certain Stoic or neo-Stoic views might have it, and one can have the relevant belief without feeling, then love does not require feeling. And a number of philosophers have defended views according to which motivations or volitions – of a certain kind, or arrayed in a certain structure – are sufficient for love, even though such motivations or volitions do not (according to such views) entail affective experiences. In a particularly striking passage along these lines, Harry Frankfurt writes the following:

Love may involve strong feelings of attraction, which the lover supports and rationalizes with flattering descriptions of the beloved. Moreover, lovers often enjoy the company of their beloveds, cherish various types of intimate connection with them, and yearn for reciprocity. These enthusiasms are not essential. Nor is it essential that a person like what he loves. He may even find it distasteful. As in other modes of caring, the heart of the matter is neither affective nor cognitive. It is volitional. Loving something has less to do with what a person believes, or with how he feels, than with a configuration of the will that consists in a practical concern for what is good for the beloved. (2004: pp. 42-43)

Relatedly, Nomy Arpaly & Tim Schroeder (2014) argue that *desire* is the essence of love, and that desire is only contingently (though non-accidentally) related to affect.<sup>154</sup>

---

<sup>154</sup> Arpaly & Schroeder (2014) argue that “for A to love B is for A to intrinsically desire B’s wellbeing a great deal.” (2014: p. 94) But they also believe that A can have such a desire without being disposed to *feel* any particular way about B. On their view, for example (as discussed briefly in Chapter 4), pleasure *represents* (roughly) the satisfaction of one’s desires (relative to expectation). (See Schroeder (2004) for further discussion.) But the relevant desire – and so love – can exist without its being represented by a pleasure, or by anything else, and even without one’s being *capable* of feeling pleasure. In this way, feelings are contingently (though non-accidentally) related to desire. Eleanor Stump (2006) and Michael Smith (2015) also put desire at the heart of love and are at best agnostic about the claim that love requires feeling.

I will not consider such views or arguments in detail here. I will simply accept that they impose a burden on the defender of L-RAD: L-RAD needs argument. And while a number of philosophers *seem* to accept that feeling is essential to love, e.g. Nozick (1993), J. David Velleman (1999), and Niko Kolodny (2003), there are problems with relying on any of them to shoulder the present burden.

Nozick offers us an outline of an argument for thinking that love requires feeling in the following passage:

What is common to all love is this: Your own well-being is tied up with that of someone (or something) you love. When a bad thing happens to a friend, it happens to her and you feel sad for her; when something good happens, you feel happy for her. When something bad happens to one you love, though, something bad also happens *to you*...If a loved one is hurt or disgraced, you are hurt; if something wonderful happens to her, you feel better off...love in some amount is present when your well-being is affected to whatever extent (but in the same direction) by another's. (1993: p. 417)

Simplified, the thought is that the lover is disposed to undergo changes in well-being depending on how his beloved is faring, and these changes in one's well-being involve affective changes, so love involves affect. The problem with this account, for present purposes, is that one could accept that you are made worse off by your beloved's struggles even if one denies that your reduction in well-being is even in part a matter of your coming to *feel* negatively about anything. Desire-satisfaction and objective list theories of well-being both open up such space, as do some of Nozick's own contributions to discussions of well-being.<sup>155</sup> To see whether Nozick's points about the relationships between love and well-being support a tight connection between love and affect, we would have to wade deeply

---

<sup>155</sup> For a seminal discussion of these different kinds of theories of well-being, see Derek Parfit (1984: Appendix I). For Nozick's discussion of the "experience machine," see his (1974).

into these debates. Though this may be a promising way to proceed, I am inclined to see if we can find a more direct route to L-RAD.

Velleman (1999) and Kolodny (2003) both accept that for A to love B is, at least in part, for A to be “emotionally vulnerable” to B. But as was discussed above, it is not at all clear what the relations are between *emotions* and affective experiences. And neither Velleman nor Kolodny give us reasons to think that for A to love B is for A to be *affectively* vulnerable, in particular, to B. Velleman says that “love...feels...like a state of attentive suspension, similar to wonder or amazement or awe.” (1999: p. 360) But he nowhere argues that this, or any other feeling, is *essential* to love.<sup>156</sup> Likewise, Kolodny says that when A is emotionally vulnerable to B, “A *may* feel content when B is well, elated when B meets with unexpected good luck, anxious when it seems that B may come to harm, grief-stricken when B does” and that “A *may* feel indignant when B’s standing or merit is questioned” (2003: p. 152 – emphases added) But even Frankfurt could accept these claims. And even when Kolodny tells us, in a more committal way, that “loving parents...*are* inclined to take particular pleasure in the delightful features of their own children, greater pleasure than they would take in the same features of other people’s children” (2003: p. 152 – emphasis added), this claim is advanced without argument.

So it is unclear whether anyone has so much as proposed L-RAD yet, let alone defended it. In the next two sections I will attempt to take steps towards filling this gap by

---

<sup>156</sup> At this point in his discussion, Velleman is primarily concerned to argue against views according to which *desires* constitute the essence of love. The passage quoted in the main text is his attempt to cast doubt on such theories on phenomenological grounds. The entire passage is: “Love does not feel (to me, at least) like an urge or impulse or inclination toward anything; it feels rather like a state of attentive suspension, similar to wonder or amazement or awe.” (1999: p. 360) Stephen Darwall (2010: p. 2) makes what seem to be related points: “If someone about whom I care is miserable and suffering, I will be disposed to emotional responses, for example, to sadness on his behalf, that cannot be explained by the mere fact that an intrinsic desire for his welfare is not realized. Taken by itself, all that would explain would be dissatisfaction, disappointment, or frustration.” I am neutral here about these claims made by Velleman and Darwall.

focusing on the relationships between feeling and what we want from love. I will argue that we could not get what we want from love from motivation and evaluation (separately or conjointly).<sup>157</sup> Feeling is necessary, too.

## **Section 2: From the Beloved's Perspective**

### **Section 2.1: Fred and Wilma**

Being *loved* is often great. It is not always great, of course, but when it is, what is so great about it? Why do we seek it out so fervently? Why do our poets and song-writers and novelists obsess about it? In this section we will learn that, when we think carefully about why we want to be loved, or at least about what we want from those who love us, we will see that affect is necessary for love.

One possibility is that we want to be loved because people who love us tend reliably to help us out and are willing to make sacrifices to do so, e.g. they might give us rides to the airport, help us move, accompany us at a concert or in the kitchen, stand shoulder-to-shoulder (or face-to-face<sup>158</sup>) with us when outrageous fortune shoots its arrows, enable us to send our genes into the next generation, etc. Call this “the help-mate model” of love.

Although this is almost certainly part of it – let us not fog up our already rose-colored glasses– it cannot be the whole of it. To see this, suppose that someone reliably acted towards you, in every instance, just as a loving person would act towards you, but who did not actually love you. This great actor secretly despised you, but would *always* be your loyal help-mate, so as to get into heaven, say. Only the most austere behaviorist would

---

<sup>157</sup> If the relevant evaluative beliefs or motivations or volitions entail affective experience, then, of course, there need be no dispute here. But if such beliefs or motivations or volitions do *not* entail affective experience, then, I will argue, such beliefs or motivations or volitions are not sufficient for love.

<sup>158</sup> See Neera Badhwar (2003) for the metaphor: friends stand shoulder-to-shoulder, (romantic) lovers stand face-to-face.

disallow this possibility. According to the help-mate model, we should be just as happy with this person as we are with a genuine lover, we would get all that we could legitimately want from a lover from this super-maid or super-butler. But that seems downright sociopathic.

What more do we want? I would like to suggest that we want people who love us to have certain psychological features and not just to behave in (putatively) loving ways. What psychological features do we want when we want to be loved and not just helped? There are many *prima facie* reasonable possibilities here. We plausibly want the lover to be intrinsically motivated to help us, for the lover not to be alienated from those motivations, for those motivations to be accompanied by various kinds of appreciation of our value, for those motivations and evaluations to be based on an especially accurate view of us (see Troy Jollimore (2011)), etc.

All of these psychological features might be a part of the story, i.e. a part of why we want to be loved and why we regard it as such a good thing to be loved. But now I will try to suggest that none of them, nor even all of them together, are the whole of it. Part of the value of being loved, part of what we want from a lover, consists in the lover's *feeling* certain ways about us.

Consider Wilma in her old age. She has been happily married to Fred for decades. They have raised children together, traveled together, eaten nearly every meal together, and helped one another in the myriad ways that one's most important partners do. Wilma has always considered herself very lucky to have met, fallen in love, and lived the greatest part of her life with Fred. He has acted in every way the supportive, generous, and thoughtful husband.

But today (imagine that this is some years in the future), after Fred is done with his yearly physical, his doctor asks Wilma to come in to speak with him, alone, for a moment. Fred's doctor says "I have what might be some troubling news for you. We finally got some new brain-scanning equipment this month – really state of the art, but nevermind that – and so we've been able to run some tests on Fred this year that we've never been able to run. And what we've found is quite surprising. It turns out – probably as a result of a very rare pre-natal condition – that Fred has a series of brain lesions that prevent him, and always have prevented him – how should I say this? – from feeling positively or negatively about anything at all. He can't *feel* pleasure, sadness, regret, excitement, none of it, and he never has! But he's *fine* and perfectly normal in all other respects..."

After tens of minutes of shock coming and going, worries expressed and answered, Wilma eventually comes to believe the doctor (and is especially relieved to hear that Fred's health is not in any jeopardy). She gets Fred from the waiting room and they return home, arm in arm. But Wilma has trouble sleeping that night. She tosses and turns, but unsure why – and then it becomes apparent: "Jesus," she thinks, "*Fred's never loved me at all*. Not at all! Not for one moment! All those loving looks, those sweet words, those thoughtful gestures – all *empty*! I don't blame Fred for anything – it's his brain, it's not his fault...god, poor Fred! But-"

Despite the sympathetic entreaties of her daughter and her friends, Wilma remains inconsolable. Has Fred cared *for* her? He certainly has, but has he cared *about* her? Has he acted and (coolly) thought in ways that someone *would* act who did love her? Beyond doubt. But given Fred's total lack of affect – he was *never* pleased to see her, was *never* pained by her pain, *never* felt pride in her achievements, *never* felt angry when someone treated her unfairly,

*never* felt anxious when her work took her to dangerous areas and her calls were late, *never* felt the faintest longing to stroke her hair in all those years of apparent intimacy, Wilma's inconsolability seems reasonable. She wanted to be loved, and for all the great things she has gotten over the years, she has not gotten that.

Fred is an extreme case, but we are considering an extreme view, i.e. that you can love someone without having even the faintest tendency to feel in any way positively towards them! At any rate, I think we can learn the same lesson by looking at more humdrum cases. Just imagine that any one of the (seemingly) loving things your partner does for you is utterly unaccompanied by you-related affect. And not just because your partner is distracted, or depressed, or undergoing any of the typical kinds of conditions that prevent affective dispositions from manifesting.<sup>159</sup> There you are on stage, receiving your Ph.D., and your partner is in the audience, utterly affectively indifferent to the proceedings, despite their great importance to you. Insofar as you want your partner to love you, this would be crushing, or at least seriously disappointing.

Why are one's realizations about one's putative lover's lack of affect so terrible? Precisely because affect is a crucial feature of what we want when we want to be loved. We want to make a vivid, vibrant phenomenological difference to those who love us. We want to be important to them, and this cannot happen without affect. Relevantly similar points apply even when we do not want to be loved. When we believe that someone loves us – whether we want them to or not – we reasonably expect them to feel certain ways about us. Wilma's realization about Fred is not only crushing, but it is also, more weakly, incredibly surprising. And insofar as we expect feeling wherever we find love, and we are shocked

---

<sup>159</sup> For great lists of and discussions of such conditions, see e.g. Michael Stocker (1979) and Michael Smith (1994).



when we do not find it, that is a good reason to think that love is tied very closely to affect indeed.

## Section 2.2: Partiality and Strangers

Next I will show that, when we want to be loved, we want to be *preferred* in certain ways to certain things. The thought is not that, in wanting to be loved, we want to be most preferred (in the relevant ways).<sup>160</sup> The idea, instead, is that – at a minimum – we want those who love us to prefer us (in the relevant ways) to strangers.<sup>161</sup> When we want to be loved we want someone to be oriented towards us in ways that they are not oriented towards John or Jane Doe. We want, to that extent, to be seen as special.

In particular, I think that part of the partiality or preference we want from those who love us is an affective asymmetry: we want those who love us to feel differently about us than they do about strangers. I will try to show this via a discussion of Bernard Williams' (1981) "one thought too many" objection to impartialist moral theories.<sup>162</sup>

Suppose that two people are drowning near Barney, each of the drowning people is equal in terms of their intrinsic value, and Barney can only save one of them. Further, one of them is a stranger and the other one is Barney's wife, Betty. According to some impartialist moral theories, the virtuous agent in such a case should, before acting, think

---

<sup>160</sup> When I want my close *friends* to love me, I do not need them to prefer me to all of their other close friends. When I want my *parents* to love me, I do not even *want* them to prefer me to my brother. And I have friends who are "polyamorous," and they seem, at least, not to need the people by whom they are *romantically* loved to prefer them to the other people they romantically love.

<sup>161</sup> Matt McCall (p.c.) has brought it to my attention that proposing this partiality condition as a necessary condition on love will likely conflict with many religious accounts of love, e.g. the accounts offered by Christian traditions according to which God loves everyone equally, or the accounts offered in Buddhist traditions according to which the Buddha instructs us each to love everyone as a mother loves her newborn (and only) child. I have two things to say. First, it might be that such traditions rely on a problematic theory of love: perhaps an entirely impartial love is no love at all. Second, and setting that possibility aside, we can alter the point in the text slightly to read: when we want to be loved, we want (thereby) to be preferred (in the relevant ways) to the *unloved*.

<sup>162</sup> This objection has featured prominently in much recent work on love. See especially Philip Pettit (1997) and Kieran Setiya (2014).

about what that theory would permit or require in the circumstance. But according to Williams, this requires “one thought too many”:

It might have been hoped by some (for instance, by his wife) that his motivating thought, fully spelled out, would be the thought that it was his wife, not that it was his wife and that in situations of this kind it is permissible to save one’s wife. (1981: p. 18)

It is this last thought, according to Williams that expresses a kind of alienation of the man from his wife. His action should manifest a more direct kind of concern for his wife that should not, according to Williams, be – or need to be – justified by any more general, impartial principle.

I happen not to find Williams’ particular criticism here entirely compelling<sup>163</sup>, but there is something important in the ballpark. What would be reasonably upsetting for Betty, I think, is if Barney were *entirely* indifferent between the stranger and Betty. If there were no more psychological pull towards saving Betty than there was towards saving the stranger, Betty could reasonably complain, or at least be reasonably disappointed.

Casting ourselves in the role of Betty, how would we have wanted Barney’s psychology to unfold as he transitioned from grim awareness of the situation confronting him to his eventual action of saving us or the stranger? What psychological states or transitions of Barney’s could have answered to our initial desires to be loved in the first place? What we want and expect here, I propose, is a gripping affective asymmetry in Barney’s psychology (of a kind that we would not expect from a stranger). As he considers the various possibilities – (a) save us, (b) save the stranger, or (c) do something else – we want Barney to be significantly more affectively attracted to (a) than to the alternatives, and

---

<sup>163</sup> There need not be anything objectionably alienating about trying to think, *in part*, in impartial terms about what kinds of preferential treatment it is ok to accord one’s loved ones, it seems to me. In fact, this seems like one of the best uses to which one *can* put morality!

to be significantly more affectively averse to the alternatives than to (a). Given the high stakes involved, I do not think it is over-stating things to suggest that what we want is that Barney feels horrified – at least for spells – about doing anything other than (a), and not as horrified about doing (a).

We want and expect Barney to be gripped by our plight in ways that he is not gripped by the plight of the stranger, even if he is (and should be) gripped by the plight of the stranger. To see this, alter the case a bit, so that now he can either save us or *five* equally valuable strangers, but not both. In such a case, we need not find it unloving of Barney to opt for saving the five strangers, not even if he does so, ultimately, on the basis of impartial moral considerations. What would be unloving is if he was not (affectively) crushed by this decision in a way that he would not be were he to save five strangers instead of one *stranger*.

Even if love does not justify behavioral partiality of a kind that Betty might hope for here – I am neutral on that score – it certainly seems to justify affective partiality.<sup>164</sup> But even if it does not really justify affective partiality, we at least seek love out in part precisely because it does involve such affective partiality. Love would not mean nearly as much to us if it did not. In our revised case where Barney could either save Betty or five equally valuable strangers, we might even be inclined partially to excuse Barney's saving Betty, if he does so, or at least – if we are not at all inclined to let him off the hook – we can appreciate the tragic nature of his choice and the intelligibility of his action. But it is far from obvious that we would go even this far if Barney felt nothing for Betty, or felt no more towards her than he did towards a stranger (or five strangers).

---

<sup>164</sup> The idea that there are different norms governing action and feeling is not new. See e.g. Bernard Williams (1965) and Allan Gibbard (1990) for seminal discussions.

Each of the cases above – Fred and Wilma, and Barney and Betty – suggest that a crucial feature of what we want when we want to be loved is affective. Our desire for love would not – could not – be satisfied by someone not disposed to feel anything for us. Further, they suggest that when we believe that someone loves us, we expect them, *as such*, to feel certain ways about us. L-RAD, I claim, best explains why Wilma and Betty (or you or I) would reasonably be crushed, or at least seriously disappointed, or at least surprised to the point of shock, were they to find out that their putative lovers felt nothing for them, or felt no more towards them than they do towards strangers.

The case of Fred, in addition, should lead us to be suspicious of both evaluative content views and motivational attitude views of phenomenal valence. We can imagine, it seems, that Fred has all of a lover's motives vis-à-vis Wilma, and maybe we can even imagine that Fred can represent all of the same evaluative properties of Wilma that a lover can. But if Fred can be positively oriented towards Wilma in these ways, and yet he can fail to love her because he fails to feel positively towards her, then it looks like evaluative content views and motivational attitude views tell the wrong story about phenomenal valence. According to such views, Fred is positively oriented towards Wilma in all of the ways that someone would be who felt positively towards her. So there should not be anything missing here vis-à-vis Fred's orientation toward Wilma. And yet there clearly is something missing, something very important: namely, positively phenomenally valenced attitudes.

### **Section 2.3: Objections**

I will return to the points just made about evaluative content views and motivational attitude views later. Here I would like to consider two kinds of objections to the idea that L-RAD

provides the best explanation of the phenomena just discussed.<sup>165</sup> Each kind of objection has it that L-RAD is more general than anything that can be supported by consideration of Wilma and Betty. The first kind of objection points out that each of these cases involves romantic love as it occurs in loving relationships. Then, we are asked, why not only conclude, at most, that affect is necessary for romantic love as it occurs in loving relationships? What justification is there for embracing the much more general L-RAD?

It is true that the Wilma and Betty cases each involve romantic love as it occurs in loving relationships, and it is true that L-RAD applies much more broadly. And I chose to focus on such cases because these are the cases that evoke most vividly the thought that affect is a crucial part of what we want when we want to be loved. Nevertheless, we can learn similar lessons (though perhaps less vividly) from cases of non-romantic love, requited or not. It can matter to us whether our friends or family love us, too. And when that matters to us, it seems to me, we would be crushed to discover that they felt nothing for us, or no more than they did towards strangers.

Imagine how horrible it would be to learn that the person who you have long thought of as your best friend was utterly affectively indifferent to your successes or failures, and was even fully affectively *opposed* to spending time with you, say. Likewise, imagine the profound disappointment of the teenager who wants and expects her mother's love, only to discover that her mother could not care less about her in an affective way, and that aiding her child always felt like a chore. Much like Wilma and Betty, our shocked friend and our disappointed daughter would reasonably think that their desires to be loved were not satisfied, precisely because affect is entirely missing (or entirely inverted). So insofar as

---

<sup>165</sup> There is a deeper methodological objection to consider as well, and I will discuss it briefly in the concluding section of this chapter.

Wilma and Betty support the idea that affect is required for romantic love, these cases support the idea that affect is required for friendship love or familial love. And L-RAD seems best suited to offer a deep and unifying explanation of these results.

Next, the point that all of these cases involve (putatively) loving relationships is well taken. Unless we shared a relationship with someone, we almost certainly would not *expect* them to love us or feel certain ways about us, and so we would not be shocked or surprised to learn that they felt nothing for us. But this shock or surprise is not the only thing that L-RAD is being enlisted to explain. It is also the crushing disappointment experienced when someone realizes that someone whose love they want feels nothing for them.

If you desperately want that wonderful-but-barely-known co-worker down the hall to love you, you would reasonably be crushed to learn (through a mutual acquaintance, say) that she felt nothing for you. And this is despite lacking a real relationship with her or any real expectation that she does love you (“why would she?” you ask yourself later). I think we can see, then, that if Wilma and Betty (and our shocked friend and disappointed daughter) suggest that affect is necessary for the love that we want from someone with whom we have a relationship, affect is necessary also for the love that we want in the absence of any such relationship. At any rate, in none of these romantic or non-romantic, relationship-involving or non-relationship-involving, cases are we likely to say or think anything remotely like “what I want is that she loves me, but I couldn’t care less how she feels about me.”

The second kind of objection to be considered is much more challenging. It can take at least two forms: (i) the love that we want or expect in these cases is *typical* love, and typical love involves affect; (ii) the love that we want or expect in these cases is *ideal* love, and ideal love involves affect. The real force of these objections consists in the following. If

we are to make our cases comprehensible in ways that are required for us to justify taking our intuitions or sentiments about them seriously, then it looks like we *must* imagine that (i) or (ii) is correct. What Wilma wants, for example, is that Fred loves her in *ideal* ways. What Betty expects, for example, is that Barney loves her in *typical* ways. These are the things that we really seem to be imagining in these cases, one might reasonably think. But if these are the things that we are really imagining, then the most that we can reasonably conclude on the basis of these cases is that typical love requires affect or that ideal love requires affect.

These are very strong objections. Not all loves are typical and not all loves are ideal. And it *is* very hard to imagine e.g. that Wilma does not want ideal love and that Betty does not expect typical love. My substantive response will not come until the end of the next section, but here is a point worth making in the interim. We should accept either of these alternative hypotheses only if we are given reasons for thinking that some stripped-down, bare-bones, affect-less phenomenon deserves to be called love at all. In the absence of such reasons, we are justified, I think, in accepting the more general L-RAD on the basis of our earlier discussions.

### **Section 3: To Love and to Feel**

#### **Section 3.1: Affect and the Lover**

In this section I would like to make the case, now from the lover's perspective, that affect is necessary for love. If anything, the case for thinking that affect is an essential part of what we want when we want *to love* is much easier to make. It is not entirely clear why we should ever care about someone else's unobservable (in fact if not in principle) internal mental life. But caring about our own mental lives seems unavoidable, especially where affect is concerned. A partial and somewhat crass answer to the question of why affect is an essential

part of what we want when we want to love goes as follows: part of what we want when we want to love, and part of why we want to love things in the first place, is that loving feels good, and feeling good is affective through and through. Though this answer is partial and somewhat crass, we could certainly do worse. If love held out no promise of enjoyment, delight, contentment, or satisfaction, then I suspect that love would be much less popular than it is.

But this account misses some things that are worth catching. A somewhat better account starts with the idea that part of why we want to love things is that we want what we do (and who we are and who and what we surround ourselves with) to *seem* and *feel important* (to us). The idea is not – or at least is not exactly – that love is important to us because it is pleasant. Rather, the idea is that at least part of the story about why loving is important to us is that we want projects and activities and ideals with which we can feel fully affectively engaged and in which we can feel deeply affectively invested and about which we can feel securely content and satisfied. Aside from contentment and satisfaction, none of these feelings need be particularly pleasant, and even contentment and satisfaction are not pleasant in quite the same way that a full belly and sexual gratification are pleasant. Nor is the idea that loving and caring about things reduces the pain of life. As Velleman (1999) points out and as every lover knows, loving can often increase the amount of pain in our lives since it makes us so vulnerable to other things and people. The exquisite anguish of heartbreak, for example, is impossible without love. But pain experienced in the course of pursuing what is important to you or what you love is thereby given instrumental value, or at least that is how it seems and feels from the inside. It does not hurt less, but now it is for a good cause. The suffering is, quite literally, no longer pointless (at least from the inside).



Such feelings of engagement and investment and satisfaction are part and parcel of love and care and there is no mystery about why we would seek them out and cherish them. These experiences are the kinds of things that can stave off even the most extreme forms of boredom, listlessness, anxiety, and depression, at least when we are in their grip. When they have their hold on us, the things with which we are engaged, in which we are invested, and about which we are satisfied feel important and feel meaningful. Skeptical worries e.g. about the meaninglessness of it all *sub specie aeternitatis* (Thomas Nagel (1971)) do not get nearly the same grip – even if we know we cannot answer them, that is not so bothersome as it is when such feelings abandon us. And such feelings can do this work in a way that does not seem from the inside like some kind of dodge, or bad faith, or trick, as e.g. avoiding such concerns and depressions via constant thrill-seeking or constant sleep might. When things are really and truly important to us, we have far less need for such subterfuge. There is a security that love as such offers and it does so, at least in part, in virtue of its ties to affect.

Nevertheless, there is a worry that this account makes love look too self-focused. If a crucial part of what we want when we want to love is that we have something in our lives that makes *us* feel certain ways, that seems to make the beloved merely a means to securing our own positive sensations. And there seems to be something self-defeating about this, at least if we accept (as we should) that love is a way of intrinsically caring about things.

But this objection is misguided. Feelings, like desires, can be intrinsic or instrumental. I can feel attracted to eating broccoli for its own sake or I can feel attracted to it only as a means to securing health (or I can feel attracted to it in both ways).<sup>166</sup> Our

---

<sup>166</sup> You can test whether feelings are intrinsic or instrumental in much the same way you would test desires for these features. If my feeling of attraction to eating broccoli would tend to survive my learning that broccoli

beloved is, as the objector suggests, not merely instrumentally important to us. On the present proposal, that means that at least some of the positive feelings we have towards our beloved are directed at them intrinsically. At least some of these feelings would survive the knowledge that our beloved or his well-being was not “helpful” to us.

I do not see what is objectionably self-regarding or self-focused about wanting to love *in part* because it involves positive feelings intrinsically directed at the beloved. I cannot imagine that Betty would complain, for example, were she to learn that one of the deepest sources of satisfaction in Barney’s life is *her*, that she herself is the cause and the intentional object of his joy and satisfaction. So even those of us who do self-consciously want to love, and who do see affect as an important part of what we want when we want to love, need not see this as any kind of defect or failing. The fact that the lover does “get something” out of love should neither be neglected nor unduly criticized. But it might be perverse for someone to want to love *only* so that they could have something in their lives that made them feel good in one or more of the ways discussed above.<sup>167</sup>

### Section 3.2: Objections Again

Finally, it is by taking up the perspective of the lover, as we have just been doing, that we can answer the objections that ended the previous section. Even if there is nothing to choose between the claim that Wilma was disappointed because (a) she was not loved and (b) she was not loved in an *ideal* way, or between the claim that Betty was disappointed because (c) she was not loved and (d) she was not loved in a *typical* way, these alternative interpretations

---

actually had no positive health effects, then the feeling is intrinsic, but if it would not tend to survive that knowledge, then it is instrumental, for example, all things being equal.

<sup>167</sup> Nozick’s (1974) discussion of the “experience machine” is relevant here. We can accept that we cannot get all that we want out of life in the experience machine without accepting that pleasure is irrelevant to well-being. Likewise, we can accept that we cannot get all that we want out of love just by feeling certain ways without accepting that affect is irrelevant to love.

of what we want and expect (i.e. (b) and (d)) do not get a grip at all in the *first-personal* case. It would be a very strange case indeed where someone wanted to love in an ideal way or in a typical way.

Consider how we actually tend to react when particular activities or things or people lose their affective grip on us. Many of us, when we are young, care a great deal about a variety of things about which we care not at all further into adulthood. These could be games or bands or sports teams or people or many other things. Typically, when our affective interest in something wanes, our motivations to pursue or protect it and our assessment of its relative value will decline, too. But these things need not (and, I think, do not) always wane together. Suppose that Elroy is a long-time and devoted fan of a sports team, the Jets. He excitedly follows the Jets' games and any and all news surrounding the team (players and coaches coming and going, predictions about next season, etc.). The Jets' losses, especially in games against rivals, break his heart and their victories fill Elroy with a deep and thorough contentment and satisfaction. His nerves are reliably shot after close contests. He talks avidly about the team to anyone who will listen. He identifies with the Jets and with his desires for their success, desires which he fully endorses having. In short, Elroy loves the Jets.

Now suppose that some years pass and although Elroy is still intrinsically *motivated* to go to the games, follow team news, talk about them, buy team merchandise, etc., and he still reliably *does* each of these things and (putatively) still thinks that the team is just as valuable as it has ever been, none of it any longer engages him affectively. His actions, etc. have not become purely a matter of habit yet, although they are very near that stage. Noticing that his affective interest has been flagging, that he is "just going through the motions" as he puts it,

and being saddened by this, he even takes steps to renew and bolster his affective interest, e.g. by going on a weekend trip to an away game with people he knows are Jets fans. He still wants to love the team, we can suppose. But it does not work. He just does not care anymore, in an affective sense, about the Jets or how they are faring.

We can imagine Elroy saying, on this basis, that the team is no longer important to him, that he no longer loves the Jets.<sup>168</sup> And we might reasonably agree with Elroy's assessment. Put in a similar situation, we might draw the same conclusion. Perhaps more importantly, recall that Elroy was *saddened* by the fact that his affective interest in the team has been flagging. The best explanation of this fact, it seems to me, given the story as told, is that his desire to love the Jets is unsatisfied. He *wants* to love the team, but unfortunately, he now realizes, he longer does.

Now return to the earlier objections to our discussions of Fred and Wilma and Barney and Betty. Notice how implausible these kinds of objections or re-interpretations are in the present case. The idea that Elroy wanted to love the Jets in *ideal* ways, or that he

---

<sup>168</sup> While reading about soccer on the internet one day recently (i.e. well after writing this bit about Elroy), I stumbled across an article (Morshead (2016)) in which the author describes his journey from loving to not loving his favorite boyhood soccer (football) team. What emerges is that our story of Elroy might not be perfectly realistic, but it is actually surprisingly true to life. Here are a few representative passages that are relevant for our purposes: "On a Tuesday night last October, I realised I was no longer in love with my football club, Swindon Town...It wasn't always like that. My dad took me to my first match in 1995 – an FA Cup clash with Marlow – and I quickly developed an insatiable curiosity, which soon turned into an infatuation. Every Christmas I wanted the new kit; I was a mascot twice; I'd rerun Saturday's goals on a Sunday morning in my parents' living room; and my cupboards were filled with programmes and mementos. That honeymoon feeling persisted through my teenage years, whatever state the club found itself in, and continued into my working life when I arrived at the Swindon Advertiser in 2009. Today, I can't motivate myself to smile or sulk as the final scores roll in. On that night in October, I struggled with the concept. I had discovered apathy for my team for the first time. Swindon lost to Oldham Athletic and, as I looked out on the County Ground's floodlights from my study window, I reflected on the result with a shrug...I wanted to look on it with childish enthusiasm and to be in the away end at Elland Road, jumping into a stranger's arms as Charlie Austin scored the third goal in a 3-0 win...I miss having to restrain my emotions in the press box when Swindon score; recently they recovered from 3-1 down to beat Crewe 4-3 and I caught myself groaning about the resultant 91st-minute rewrite...I hate that Swindon went to Wembley last May for a play-off final, lost 4-0 and three days later I was over the experience."

expected to love the team in *typical* ways, is totally unmotivated here. Those are awfully precious hopes and expectations to attribute to an ordinary enough person. We have no reason to think that he cares about ideal or typical love (does anyone?), or that we do when we find ourselves in Elroy's shoes. He just wants to love the team and he is sad that he does not.

If you want to love someone, and you realize that you feel nothing for them, you would reasonably be disappointed, and again, the best explanation of this fact is that affect is essential to love and so your desire to love is unsatisfied. But the idea that what you really wanted or expected *from yourself* was ideal or typical love has nothing to recommend it. L-RAD, then, can provide a unified explanation of what we want and expect from love, both when we want love from others and when we want to love thus-and-so ourselves. And so, I conclude, love full-stop, and not just ideal or typical love, requires affect.

And we can imagine that Elroy, at the key point of the story, still has all of the motives and represents all of the values that somebody would have/represent who felt strongly about the Jets. If that is right, then Elroy would be positively oriented towards the Jets in all the ways that an evaluative content view or motivational attitude view would say that someone who feels strongly about the Jets is. And yet he is *not* positively oriented towards the Jets in all the ways that someone who feels strongly about the Jets is. This is precisely what makes it the case that Elroy no longer loves the Jets. And so, again, we learn here not only that affect is necessary for love, but that evaluative content views and motivational attitude views cannot, even if they join forces, give us a proper account of phenomenal valence. A proper account of phenomenal valence would help us to see what Elroy (or Fred) are missing.

#### Section 4: Conclusion

Before concluding, it is worth mentioning another kind of objection that one might offer to the account presented here. Wilma, Betty, Elroy, and the rest of our characters were stipulated (a) to want and expect love of or from B, (b) to be disappointed and surprised when feeling was missing, and (c) to conclude on that basis that love was missing. Let us grant that, given (a), reaction (b) is quite natural and reasonable and that, given (b), reaction (c) is quite natural and reasonable. Even so, why think that we have learned anything about *love* here? Haven't we just learned what ordinary people like Wilma, Betty, etc. *think* about love? It is not obvious whether lessons like these, lessons at most about the *concept* of love, can teach us about love itself.<sup>169</sup>

In response, the idea is that L-RAD is intended as a metaphysical claim about love itself, but one that is undergirded by an investigation into some important features of the concept of love. Whether this way of proceeding is plausible will end up turning on whether it is ever plausible to think that the nature of a thing (or property or relation) can be illuminated by reflecting on the roles that the concept of that thing (etc.) plays in our lives. All I can do here is to place my bet that, sometimes, this happens. It certainly seems to do so in the case of love. But even if it does not, even if we understand L-RAD merely as a claim about the concept of love, that still imposes a burden on those who would offer a non-affective account of love: why should we accept that such accounts are accounts of *love* at all?

Finally, suppose that it turns out that affect, despite appearances, is not necessary for love (in any sense). Even if we have learned nothing here about love, we have learned a

---

<sup>169</sup> I would like to thank Sigrún Svavarsdóttir (p.c.) and Justin D'Arms (p.c.) for each independently encouraging me to consider issues closely related to the objection just presented.

crucial lesson about *affect*. At the very least, we now have good reasons to believe that the itches that affect scratch cannot be scratched fully by utterly non-affective motivations or evaluations. Suppose, say, that what we can call “affective love” is the only kind of love that requires affect. What we have learned above, in these terms, is that affective love is itself very important to us and that we could not get all that we want from it from anything utterly non-affective. And we have also learned here, in combination with the points made in earlier chapters, that the most promising answer to the question with which we began this dissertation – what is it to feel positively or negatively about something? – is that provided by a phenomenal attitude view – namely, it is for your conscious orientation towards it to have a certain feel and flavor, a feel and flavor that we find, for example, in love.

## Chapter 6: Conclusion

In this chapter I will briefly summarize where, I think, it would be especially worthwhile for us to go next. I would like to begin that discussion with a bit of intellectual autobiography.

Two roads on which I had previously been traveling each, more or less independently, led to my philosophical interest in affective experience. First, I was (and am) interested in the relationships between desires and reasons for action. It seemed obvious to me early on in my philosophical career that wanting something gives you a reason to do it, even if that reason is often very weak and is outweighed by countervailing considerations. If you prefer pears to bananas – to use a case from Joseph Raz (1999: p. 62) – then (*all else equal*) it would be crazy for you to choose the banana! You have more reason to choose the pear than the banana, it seems, and this is so precisely because you want the pear but not the banana. There is nothing objectively preferable about pears, for example.

But then one encounters philosophical theories of desire (see e.g. some of those discussed in Chapters 4 and 5) according to which desires are just (or are just the grounds of) behavioral dispositions. As Warren Quinn (1993) convincingly argues, if that is all that desires essentially are, then desires do not necessarily provide reasons for action. To paraphrase Quinn (1993: p. 236), the mere fact that we are set up to go in a particular direction does not make it even *prima facie* rational to go in that direction.



At this point, we can either give up the intuitive idea that desires provide reasons, or we can adopt a different account of desire. Given only these two options, I suppose I am inclined to opt for the latter. But these are not *really* our only two options. Instead of arguing for one theory of desire or another, we can go back and try to figure out why the thesis that desires provide reasons for action seemed compelling in the first place, e.g. in the pear-banana case. In my case, when I tried this approach, I discovered that what really drew me to that thesis was the connection that seemed to me to obtain between desire and *feeling* (i.e. between desire and what I would now call affective experience). If you are choosing between a pear and a banana, and – while deliberating – the prospect of eating the pear fills you with feelings of delight whereas the prospect of eating the banana fills you with feelings of nausea and annoyance – then we can see why choosing the pear is the *prima facie* rational thing for you to do. This thought was bolstered and enriched, for me, by consideration of Declan Smithies’ (2012) arguments for thinking that consciousness has an especially important *epistemic* role to play, too.<sup>170</sup>

If desires come along with such tendencies to feel, then desires might provide reasons for action after all. But notice that desire is no longer the star of the show, not really. Rather, *affect* has taken center stage. So, it seemed to me, we should try to acquire a richer understanding of affect at least in part so we could acquire a richer understanding of practical reason. I have hardly said anything at all here about practical reason, but I hope that the account of affect offered in this dissertation can serve in part as a prolegomenon to such work. This hope was one of two major considerations that led me to write this dissertation in the first place.

---

<sup>170</sup> I first came across the relevant arguments in a seminar that Smithies taught at Ohio State University in the winter of 2010.

The second such consideration was also metaethical-cum-moral psychological. It begins with my interest in a long-standing debate about the nature of normative and evaluative judgment. This debate as typically formulated has to do with the (potential) relations between normative and evaluative *judgment*, on one hand, and *motivation* on the other. For example, take (a) the judgment that  $\Phi$ -ing is the right thing to do and (b) motivation to  $\Phi$ . Some philosophers, the ‘motivational judgment internalists,’ tell us, roughly, that (a) implies (b): it is necessary, and a priori knowable, that if someone judges that  $\Phi$ -ing is the right thing to do, then they are motivated to  $\Phi$ . And this is not because we all happen to be great people – rather, the story goes, it is of the essence of such judgments that they come along with motivation in these ways. I would not count as making such a judgment at all were I not relevantly motivated.

Other philosophers, the ‘motivational judgment externalists,’ deny the claims just made. According to these philosophers, it is possible, at least for all we can tell a priori, that one makes such a judgment without being relevantly motivated. I might believe that  $\Phi$ -ing is the right thing to do, say, even if I lack *any* tendency to  $\Phi$ , even of the most feeble kind.

These debates are interesting for a number of reasons. Some have thought, for instance, that if motivational judgment internalism is true e.g. of moral judgments, then there are no moral truths and we cannot acquire moral knowledge, since our moral judgments, on this picture, would not even be beliefs!<sup>171</sup> Despite these potential implications (and not because of them, unlike A.J. Ayer (1952), maybe), I have long been attracted to motivational judgment internalism. When I reflect on cases of those who putatively think that such-and-such is right or wrong, good or bad, but who would not lift a finger to promote or prevent

---

<sup>171</sup> See Michael Smith’s (1994) for what I regard as the seminal discussion of these issues.

such-and-such, I always come away thinking that they do not *really* think that such-and-such is right or wrong, good or bad. Really, they are using the relevant words or concepts in an “inverted commas sense” (R.M. Hare (1952)), or they are being ironic or insincere in some other way, or they are conceptually confused or self-deceived. The “amoralist” cases that pop up from time to time (see e.g. Michael Stocker (1979), David O. Brink (1989), and Sigrún Svavarsdóttir (1999) for some notable contributions to this genre) have not, in my view, been convincing. C.L. Stevenson (1937: p. 16) wrote that

“goodness” must have, so to speak, a magnetism. A person who recognizes X to be “good” must *ipso facto* acquire a stronger tendency to act in its favour than [sic] he otherwise would have had.

This continues to seem to me to get at something true and important.

But I have come to worry about *motivational* judgment internalism for two reasons. First, there are the kinds of cases discussed in Section 3 of Chapter 4 above. For example, I can certainly make normative or evaluative judgments about things that I know I cannot affect, e.g. things that happened in the past, or things that preclude my interference, e.g. the outcome of Wimple vs. Ivanov. But as we have seen, motivation aimed at such things can be difficult to make sense of. And maybe – although I am not nearly as confident about this – the Weather Watchers can make normative or evaluative judgments, too.

But even if such cases pose a problem for motivational judgment internalism as characterized above, they do not seem to challenge what is, for me, anyway, the more fundamental idea that makes motivational judgment internalism attractive in the first place. The more fundamental idea is, as I would put it (and as I have put it in earlier chapters), that to believe that something is good (right, rational, etc.) is to be *for* that thing in some way. I just could not, and cannot, believe that it is possible for someone to believe, say, that torture

is wrong while being utterly, entirely *indifferent* to torture. Genuinely to believe that torture is wrong is as such, it seems to me, a way of being *against* torture. But *this* more fundamental idea is not challenged by cases of the kind discussed in Section 3 of Chapter 4. In fact, those cases are interesting (as we have seen) precisely because they seem to show that we can be positively or negatively oriented towards something without being positively or negatively motivated with respect to it. But once we have arrived at this point, we will need to provide some alternative account of what the *valence* of normative and evaluative judgments consists in, given that it cannot (always) be their motivational force. This contributed heartily to my writing of this dissertation.

Second, I arrived at worries about motivational judgment internalism analogous to Quinn's worries about behavioristic accounts of desire (see above). On the one hand, it seems like normative and evaluative judgments can make actions that accord with them *prima facie* rational. For instance, if I believe that I ought, all things considered to  $\Phi$ , then – all else equal – it would be rational for me to  $\Phi$ , or at least there would be something rationally problematic about my not  $\Phi$ -ing. But I began to wonder why that was. Why is there anything to be said, rationally speaking, for one's doing what one thinks one ought to do? What is so special about such judgments? Roughly, I came to think that the explanation of the rationalizing force of normative and evaluative judgments should go by way of their *valence*. It is in virtue of the fact that such judgments are (perhaps special) ways of favoring or disfavoring their intentional objects that they make actions that accord with them *prima facie* rational. But if we explain the valence of normative and evaluative judgments entirely in terms of their motivating force, as a motivational judgment internalist thesis might suggest we do, then Quinn's arguments will rear their ugly head here, too. Again, the mere fact that

I am set up to go in a particular direction does not make it even *prima facie* rational to go in that direction. So motivational judgment internalism will at least need supplementation, if not outright replacement.

Eventually, and unsurprisingly, I came to think that the right solution to these problems is analogous to the solution we offered in the case of desire: normative or evaluative judgments rationalize actions that accord with them in virtue, at least in part, of their ties to *affect*. This kind of view, which we can call ‘affective judgment internalism,’ made further investigation into affective experiences pressing, as a great many conversations with Sigrún Svavarsdóttir over the course of a number of years helped me to see. The idea to be explored was that normative or evaluative *judgments* rationalized certain *motivations* in virtue of the fact that such judgments come along with tendencies to have relevant *affective experiences*. For example, suppose that I believe that I ought to go to the store. The rough idea is that this judgment (all else equal) rationalizes my being motivated to go to the store because this judgment entails a disposition for me to feel positively towards going to the store. But to see whether this kind of idea has *any* hope of being interesting and true, we need to see what affective experiences really are and what relation they stand in, or do not stand in, to evaluation and motivation. And I hope that this dissertation has gone some way towards filling those gaps.

These considerations having to do with desire and with normative or evaluative judgment are primarily what led me to write this dissertation about affective experiences. But as often happens, the original questions have remained unanswered here. In other work (Smithies & Weiss (in progress)) I argue that affect is necessary for desire, and so attempt to answer our first set of questions. In future work I hope to defend affective judgment

internalism and in so doing I intend to rely heavily on many of the points developed in this dissertation, especially the points made in Chapters 3 and 4. One might also be interested in seeing how this account of affective experience bears on debates about the *emotions*, too, especially on the debates between feeling theories of the emotions and their opponents. I am especially interested in seeing what impacts our account of affect might have on accounts of *pleasure* (or vice versa). It is quite possible that this dissertation, really, at the end of the day, is precisely a sustained defense of an unorthodox view of pleasure. I am not sure what hangs on whether this is so, but I am eager to look into these issues more deeply.

What emerges from these various discussions will, I suspect, be a highly affect-centric picture of human psychology and practical reason. And if the points made in Chapter 5 above were correct, at any rate, then affect should also have a starring role to play in theories of well-being, the meaning of life, and potentially even virtue and freedom. And it now looks like affect is well-suited to play all of these important roles not because of any values it putatively clues us into or because of any behavioral dispositions it putatively comes along with, but because of the valenced phenomenology of the attitudes it includes.

But the points made in this dissertation are not only of interest to ethicists, metaethicists, and moral psychologists. We have also broached some very challenging, and very deep, issues in the philosophy of mind. As suggested in Chapter 2, much of this dissertation can be understood as a sustained argument against pure intentionalist views of the phenomenology and intentionality of affective experience.<sup>172</sup> Such views suggest that conscious experience is always in the business of telling us how things are. But feeling

---

<sup>172</sup> The relevance of Chapters 2 and 3 to pure intentionalism should be clear enough. In earlier drafts of the dissertation, I attempted to describe ways in which the points made in Chapter 4 were also problematic for pure intentionalism, but these points have not survived into the final draft. I hope to return to these issues in the future.

positively about things, if the impure intentionalist account defended above is correct, is not in the business, at least not primarily, of telling us how things are. I am sure that I have not considered all of the *prima facie* plausible responses a pure intentionalist could offer to the challenges raised earlier, and I am quite sanguine about the possibility of illuminating further debate about these issues.

A set of questions about intentionality that I have tried to sidestep entirely here are questions about wide (“external”) vs. narrow (“internal”) intentional content. At a number of points I have made claims about the intentionality or phenomenology of affect that seem to implicitly rely on the idea that there is such a thing as narrow content (see especially Section 3.3 of Chapter 2 and Section 4.3 of Chapter 4.) Maybe such claims are unsustainable. Maybe all intentional content is wide or external intentional content, the kind of intentional content that intrinsic duplicates need not share. And maybe, if so, that will cause trouble for one or more of the views defended above. These are issues that will need to be addressed in the future. But I hope that I have at least made the case that, even if my Brain in a Vat or Twin Earth twins and I do not share any intentional *contents* in common, we can at least share *attitudes* in common: namely, the phenomenally valenced attitudes embodied by affective experiences. These attitudes are intentional in that they are about or directed towards things. So even if all intentional content is wide or external, it seems like we still have good reasons to believe that there are important *intentional* similarities between ourselves and our intrinsic duplicates. As we might put it, our *stances* towards the world might be just the same, even if the world towards which we are each so oriented is entirely different. And maybe this will open up some new terrain in these fraught debates.

The final major issue I hope to address in the philosophy of mind vis-à-vis the arguments of this dissertation have to do with how these arguments and their presuppositions and implications might impact possible solutions to the mind-body problem. I do not take anything said above to *imply* that affective experiences are immaterial, non-physical, or non-natural substances or properties (thank goodness). But the points made above, and the ways in which I have often argued for them, do not do materialism, physicalism, or naturalism about the mind – and in particular, about consciousness – any favors, either. In particular, as discussed briefly in Chapter 4, the view of affect defended here seems to sit uneasily with some noteworthy functionalist accounts of the mind.<sup>173</sup> But I hope that, in any case, our discussions of consciousness in the future are less narrowly focused on perceptual experience than they have tended to be to this point. Affective experiences deserve to be, if not front and center, at least close to it, in such discussions.

The most pressing questions of all, at least of questions to do with feelings – e.g. how should I feel about things? how, without operating in bad faith, can I bring it about that I feel how I want to feel? – have not been answered here. But at least we have a somewhat better grip now on what such questions might really be asking, and that is something worth feeling positively about.

---

<sup>173</sup> The view defended here, and the way it has been defended, is probably inconsistent with what David Chalmers (1999) calls ‘type-A materialism.’ Type-A materialism has it that there is an “a priori entailment” from the physical facts to the mental (including the phenomenal) facts. Type-B materialism, on the other hand, claims that there is no such a priori entailment, but that there is an a posteriori entailment from the physical facts to the mental facts. Roughly, both kinds of materialism agree that mental (including phenomenal) facts are identical to (or reducible to, or grounded in: things get complicated here) physical facts, but they differ on the epistemic status of such identifications (etc.). Type-B materialism seems to me to be preferable to type-A materialism and, I hope, it is consistent with the view defended here and the way it has been defended. Whether that is so, however, is another important area for future research. Thanks are due to Declan Smithies (p.c.) for discussion of these issues.



## References

- Anscombe, G.E.M., 2000, *Intention*, 2<sup>nd</sup> Edition, Harvard University Press
- Armstrong, D.M., 1993, *A Materialist Theory of the Mind*, Revised Edition, Routledge
- Arpaly, Nomy, 2000, “On Acting Rationally against One’s Best Judgment”, *Ethics* 110: 488-513
- Arpaly, Nomy and Timothy Schroeder, 2014, *In Praise of Desire*, Oxford University Press
- Ayer, A.J., 1952, *Language, Truth, and Logic*, 2<sup>nd</sup> Edition, Dover Books
- Badhwar, Neera K., 2003, “Love” in *The Oxford Handbook of Practical Ethics* (LaFollette ed.), Oxford University Press: pp. 42-69
- Barrett, Lisa Feldman, 2006, “Valence is a Basic Building Block of Emotional life”, *Journal of Research in Personality* 40: 35-55
- Barrett, Lisa Feldman and Batja Mesquita, Kevin N. Ochsner, and James J. Gross, 2007, “The Experience of Emotion”, *Annual Review of Psychology* 58: 373-403
- Bayne, Tim and David Chalmers, 2003, “What is the Unity of Consciousness?”, *The Unity of Consciousness: Binding, Integration, and Dissociation* (Cleeremans ed.), Oxford University Press: 23-58
- Bayne, Tim and Michelle Montague (eds.), 2011, *Cognitive Phenomenology*, Oxford University Press
- Bengson, John, 2015, “The Intellectual Given”, *Mind* 124.495: 707-760
- Bennett, Jonathan, 1974, “The Conscience of Huckleberry Finn”, *Philosophy* 49.188: 123-134
- Berridge, Kent and Piotr Winkielman, 2003, “What is an Emotion? (The Case for Unconscious ‘Liking’)”, *Cognition and Emotion* 17.2: pp. 181-211

- Blackburn, Simon, 1984, *Spreading the Word: Groundings in the Philosophy of Language*, Oxford University Press
- Block, Ned, 1980a, "What Is Functionalism?" in *Readings in Philosophy of Psychology*, Vol. 1 (Block ed.), Harvard University Press: 171-185
- Block, Ned, 1980b, "Troubles with Functionalism" in *Readings in Philosophy of Psychology*, Vol. 1 (Block ed.), Harvard University Press: 268-307
- Block, Ned, 1990, "Inverted Earth", *Philosophical Perspectives* 4: 53-79
- Block, Ned, 1995, "On a Confusion about a Function of Consciousness", *The Behavioral and Brain Sciences* 18.2: pp. 227-47
- Block, Ned, 2003, "Mental paint" in *Reflections and replies: Essays on the philosophy of Tyler Burge* (Hahn and Ramberg eds.), MIT Press: 165-200
- Block, Ned, "Conceptual Role Semantics", from <http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/ConceptualRoleSemantics.html>
- Block, Ned, 2014, "Seeing-As in the Light of Vision Science", *Philosophy and Phenomenological Research* 89.3: 560-572.
- Brady, Michael S., 2007, "Recalcitrant Emotion and Visual Illusions", *American Philosophical Quarterly* 44.3: 273-284
- Brady, Michael S., 2009, "The Irrationality of Recalcitrant Emotions", *Philosophical Studies* 145: 413-430
- Brandt, Richard B., 1998, *A Theory of the Good and the Right*, Prometheus Books
- Brentano, Franz, 2002, "The Distinction Between Mental and Physical Phenomena", *Philosophy of Mind: Classical and Contemporary Readings* (Chalmers ed.), Oxford University Press: 479-484
- Brewer, Bill, 1999, *Perception and Reason*, Oxford University Press
- Brink, David O., 1989, *Moral Realism and the Foundations of Ethics*, Cambridge University Press
- Broome, John, 1999, "Normative Requirements", *Ratio* 12.4: 398-419
- Burge, Tyler, 1979, "Individualism and the Mental", *Midwest Studies in Philosophy* 4.1: 73-121
- Byrne, Alex, 2001, "Intentionalism Defended," *The Philosophical Review* 110.2: 199-240.

- Chalmers, David, 1996, *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press
- Chalmers, David, 1999, "Materialism and the Metaphysics of Modality", *Philosophy and Phenomenological Research* 59.2: 473-496
- Chalmers, David, 2004, "The Representational Character of Experience", *The Future for Philosophy* (Leiter ed.), Oxford: Clarendon Press: 153-182
- Chang, Ruth, 2004, "Can Desires Provide Reasons for Action?" in *Reason and Value: Themes from the Moral Philosophy of Joseph Raz* (Wallace, Pettit, Scheffler, and Smith eds.), Oxford University Press: 56-91
- Chappell, Richard Yetter, 2012, "Fittingness: The Sole Normative Primitive", *The Philosophical Quarterly* 62.249: 684-704.
- Chisholm, Roderick, 1942, "The Problem of the Speckled Hen", *Mind* 51.204: 368-373
- Clore, Gerald L., 1994, "Why Emotions Require Cognition", *The Nature of Emotion: Fundamental Questions* (Ekman and Davidson eds.), Oxford University Press: 181-192.
- Conee, Earl and Richard Feldman, 2001, "Internalism Defended", *American Philosophical Quarterly* 38.1: 1-18
- Crane, Tim, 1998, "Intentionality as the Mark of the Mental", *Royal Institute of Philosophy Supplement*: 229-252.
- Crane, Tim, 2009, "Intentionalism", *The Oxford Handbook of Philosophy of Mind Online* (McLaughlin, Beckermann, and Walter eds.), Oxford University Press: 1-25 [accessed March 21<sup>st</sup>, 2016]
- Damasio, Antonio, 1994, *Descartes' Error: Emotion, Reason, and the Human Brain*, Putnam
- D'Arms, Justin & Daniel Jacobson, 2000, "The Moralistic Fallacy: On the 'Appropriateness' of Emotions", *Philosophy and Phenomenological Research* 61.1: 65-90
- D'Arms, Justin & Daniel Jacobson, 2003, "The Significance of Recalcitrant Emotion (or, Anti-Quasijudgmentalism)", *Royal Institute of Philosophy Supplement* 52: 127-145
- Darwall, Stephen, 2010 *Welfare and Rational Care*, Princeton University Press.
- Davidson, Donald, 1963, "Actions, Reasons, and Causes", *The Journal of Philosophy* 60.23: 685-700

- Davidson, Donald, 2001, "Freedom to Act" in *Essays on Actions and Events*, 2<sup>nd</sup> Edition, Oxford: Clarendon Press: 63-83
- Deonna, Julien A & Fabrice Teroni, 2012, *The Emotions: A Philosophical Introduction*, Routledge
- Döring, Sabine, 2007, "Seeing What To Do: Affective Perception and Rational Motivation", *Dialectica*, 2007: pp. 363-394
- Dretske, Fred, 1994, *Naturalizing the Mind*, MIT Press
- Dretske, Fred, 2002, "Conscious Experience", *Philosophy of Mind: Classical and Contemporary Readings* (Chalmers ed.), Oxford University Press: 422-435
- Duncker, Karl, 1941, "On Pleasure, Emotion, and Striving", *Philosophy and Phenomenological Research* 1.4: 391-340
- Farkas, Katalin, 2008, "Phenomenal Intentionality Without Compromise", *The Monist* 91.2: 273-293
- Farkas, Katalin, 2009, "Not Every Feeling is Intentional", *European Journal of Analytic Philosophy* 5.2: 39-52.
- Faucher, Luc and Christine Tappolet (eds.), 2007, *The Modularity of Emotions*, University of Calgary Press
- Feldman, Fred, 2004, *Pleasure and the Good Life: Concerning the Nature, Varieties and Plausibility of Hedonism*, Oxford: Clarendon Press
- File:My Lai massacre.jpg, n.d., in *Wikipedia*, retrieved March 20<sup>th</sup>, 2016 from [https://en.wikipedia.org/wiki/File:My\\_Lai\\_massacre.jpg](https://en.wikipedia.org/wiki/File:My_Lai_massacre.jpg)
- Fish, William, 2010, *Philosophy of Perception: A Contemporary Introduction*, Routledge
- Fodor, Jerry, 1998, *Concepts: Where Cognitive Science Went Wrong*, Oxford: Clarendon Press
- Foot, Philippa, 1978, "Hume on Moral Judgment", *Virtues and Vices and Other Essays in Moral Philosophy*, University of California Press: pp. 74-81
- Frankfurt, Harry G., 1971, "Freedom of the Will and the Concept of a Person", *The Journal of Philosophy* 68.1: 5-20
- Frankfurt, Harry G. 2004. *The Reasons of Love*. Princeton University Press.
- Frege, Gottlob, 1948, "Sense and Reference", *The Philosophical Review* 57.3: 209-230

- Frijda, Nico H., 1994, "Varieties of Affect: Emotions and Episodes, Moods, and Sentiments", *The Nature of Emotion: Fundamental Questions* (Ekman and Davidson eds.), Oxford University Press: 59-68
- Gazzaniga, Michael, 1985, *The Social Brain*, Basic Books
- Gibbard, Allan, 1990, *Wise Choices, Apt Feelings: A Theory of Normative Judgment*, Harvard University Press
- Goldie, Peter, 2000, *The Emotions: A Philosophical Exploration*, Oxford University Press
- Goldman, Alvin, 1986, *Epistemology and Cognition*, Harvard University Press
- Gunther, York H., 2003, "Introduction" in *Essays on Nonconceptual Content* (Gunther ed.), MIT Press: 1-21
- Gunther, York H., 2003, "Emotion and Force" in *Essays on Nonconceptual Content* (Gunther ed.), MIT Press: 279-289
- Gunther, York H., 2004, "The Phenomenology and Intentionality of Emotion", *Philosophical Studies* 117: 43-55
- Hare, R.M., 1952, *The Language of Morals*, Oxford: Clarendon Press
- Harman, Gilbert, 1986, *Change in View*, MIT Press
- Harman, Gilbert, 1990, "The Intrinsic Quality of Experience", *Philosophical Perspectives* Vol. 4: 31-52
- Heathwood, Chris, 2006, "Desire Satisfactionism and Hedonism", *Philosophical Studies* 128: 539-563
- Heathwood, Chris, 2007, "The Reduction of Sensory Pleasure to Desire", *Philosophical Studies* 133: 23-44
- Helm, Bennett, 2001, *Emotional Reason: Deliberation, Motivation, and the Nature of Value*, Cambridge University Press
- Hill, Christopher S., 2009, *Consciousness*, Cambridge University Press
- Horgan, Terence and John Tienson, 2002, "The Intentionality of Phenomenology and the Phenomenology of Intentionality", *Philosophy of Mind: Classical and Contemporary Readings* (Chalmers ed.), Oxford University Press: 520-534

- Hulse, Donovan, Cynthia N. Read, Timothy Schroeder, 2004, "The Impossibility of Conscious Desire", *American Philosophical Quarterly* 41.1: 73-80
- Hume, David, 1978, *A Treatise of Human Nature*, 2<sup>nd</sup> Edition (Selby-Bigge ed.), Oxford University Press
- James, William, 1884, "What is an Emotion?", *Mind* 9.34: 188-205
- James, William, 1950, *The Principles of Psychology*, Vols. 1 & 2, Dover
- Johnston, Mark, 2001, "The Authority of Affect", *Philosophy and Phenomenological Research* 63.1: 181-214
- Jollimore, Troy, 2011 *Love's Vision*, Princeton University Press.
- Kaplan, David, 1989, "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals", *Themes from Kaplan* (Almog, Perry, and Wettstein eds.), Oxford University Press: 481-565
- Kind, Amy, 2003, "What's So Transparent about Transparency?", *Philosophical Studies* 115.3: 225-244
- Kolodny, Niko, 2003 "Love as Valuing a Relationship", *Philosophical Review* 112.2: pp. 135-189
- Kraut, Robert, 1986, "Feelings in Context", *The Journal of Philosophy* 83.11: 642-652
- Kriegel, Uriah (ed), 2013, *Phenomenal Intentionality*, Oxford University Press
- Kripke, Saul, 1982, *Wittgenstein on Rules and Private Language: An Elementary Exposition*, Harvard University Press
- Lewis, David, 1966, "An Argument for the Identity Theory", *The Journal of Philosophy* 63.1: 17-25
- Lewis, David, 1972, "Psychophysical and Theoretical Identifications", *Australasian Journal of Philosophy* 50.3: 249-258
- Lewis, David, 1980, "Mad Pain and Martian Pain" in *Readings in Philosophy of Psychology*, Vol. 1 (Block ed.), Harvard University Press: 216-223
- Lycan, William G., 1996, *Consciousness and Experience*, Bradford Books/MIT Press
- Lycan, William G., 2001, "A Defense of Phenomenal Externalism", *Philosophical Perspectives* 15: 17-35

- Lycan, William G., 2008, "Phenomenal Intentionalities", *American Philosophical Quarterly* 45.3: 233-252
- McIntyre, Alison, 1990, "Is Akratic Action Always Irrational?" in *Identity, Character, and Morality: Essays in Moral Psychology* (Flanagan and Rorty eds.), MIT Press: 379-401
- Martin, Michael G.F., 2003, "The Transparency of Experience", *Mind & Language* 17.4: 376-425
- Montague, Michelle, 2009, "The Logic, Intentionality, and Phenomenology of Emotion", *Philosophical Studies* 145: 171-192
- Moore, George Edward, 1953, *Some Main Problems of Philosophy*, Allen & Unwin
- Morshead, Sam, 03/31/2016, "I was a Swindon Town fan...until it became my job to report on the club", *The Guardian* (<http://www.theguardian.com/football/when-saturday-comes-blog/2016/mar/31/swindon-town-fan-reporter-club>)
- Nagel, Thomas, 1971, "The Absurd", *The Journal of Philosophy* 68, 20: pp. 716-727
- Nagel, Thomas, 1974, "What is it Like to be a Bat?", *The Philosophical Review* 83, 4: pp. 435-450
- Nozick, Robert, 1974, *Anarchy, State, and Utopia*, Basic Books
- Nozick, Robert, 1993, "Love's Bond." *The Philosophy of (Erotic) Love* (Solomon & Higgins eds.), University Press of Kansas: pp. 417-432
- Nussbaum, Martha, 2004, "Emotions as Judgments of Value and Importance" in *Thinking about Feeling: Contemporary Philosophers on Emotions* (Solomon ed.), Oxford University Press: pp. 183-199
- Parfit, Derek, 1984, *Reasons and Persons*, Oxford University Press
- Parfit, Derek, 2011, *On What Matters: Vol. 1*, Oxford University Press
- Pautz, Adam, 2013, "Does Phenomenology Ground Mental Content?" in *Phenomenal Intentionality* (Kriegel ed.), Oxford University Press: pp. 194-235
- Peacocke, Christopher, 1983, *Sense and Content: Experience, Thought, and Their Relations*, Oxford University Press
- Pettit, Philip, 1997, "Love and Its Place in Moral Discourse" in *Love Analyzed* (Lamb ed.), Westview Press: pp. 153-163

- Prinz, Jesse J., 2004, *Gut Reactions: A Perceptual Theory of Emotion*, Oxford University Press
- Prinz, Jesse J., 2007, *The Emotional Construction of Morals*, Oxford University Press
- Prinz, Jesse J., 2011, "The Sensory Basis of Cognitive Phenomenology", *Cognitive Phenomenology* (Bayne and Montague eds.), Oxford University Press: 174-197
- Pryor, James, 2005, "There Is Immediate Justification" in *Contemporary Debates in Epistemology* (Steup and Sosa eds.), Blackwell Publishing: 181-202
- Putnam, Hilary, 1975, "The Meaning of 'Meaning'" in *Mind, Language, and Reality*, Cambridge University Press: 215-271
- Quine, W.V.O., 1960, *Word and Object*, MIT Press
- Quinn, Warren, 1993, "Putting Rationality in its Place" in *Morality and Action*, Cambridge University Press: 228-255
- Rabinowicz, Wlodek & Toni Rønnow-Rasmussen, 2004, "The Strike of the Demon: On Fitting Pro-attitudes and Value", *Ethics* 114: 391-423
- Raz, Joseph, 1999, *Engaging Reason: On the Theory of Value and Action*, Oxford University Press
- Ryle, Gilbert, 2000, *The Concept of Mind*, University of Chicago Press
- Scanlon, Timothy, 1998, *What We Owe to Each Other*, Harvard University Press
- Scarantino, Andrea, 2010, "Insights and Blindspots of the Cognitivist Theory of Emotions", *British Journal for the Philosophy of Science* 61: 729-768
- Schafer, Karl, 2013, "Perception and the Rational Force of Desire", *The Journal of Philosophy* 110.5: 258-281
- Schnall, Simone, Jonathan Haidt, Gerald L. Clore, and Alexander H. Jordan, 2008, "Disgust as Embodied Moral Judgment", *Personality and Social Psychology Bulletin* 34.8: 1096-1109
- Schroeder, Timothy, 2004, *Three Faces of Desire*, Oxford University Press.
- Schroeder, Timothy and Ben Caplan, 2007, "On the Content of Experience", *Philosophy and Phenomenological Research* 75.3: 590-611
- Scott, A.O., 10/17/2013, "Seems This Anarchist is a Laptop Hogger", *The New York Times* (<http://mobile.nytimes.com/2013/10/18/movies/the-fifth-estate-with-benedict-cumberbatch-as-julian-assange.html>)



- Seager, William, 2002, "Emotional Introspection", *Consciousness and Cognition* 11: 666-687
- Seager, William and David Bourget, 2007, "Representationalism about Consciousness" in *The Blackwell Companion to Consciousness* (Velmans and Schneider eds.), Blackwell: 261-277
- Searle, John R., 1983, *Intentionality: An Essay in the Philosophy of Mind*, Cambridge University Press
- Setiya, Kieran, 2014, "Love and the Value of a Life", *Philosophical Review* 123.3: 251-280
- Siewert, Charles, 1998, *The Significance of Consciousness*, Princeton University Press
- Siewert, Charles, 2004, "Is Experience Transparent?", *Philosophical Studies* 117.1/2: 15-41
- Singer, Peter, 1972, "Famine, Affluence, and Morality", *Philosophy & Public Affairs* 1.3: 229-243
- Smart, J.J.C., 1959, "Sensations and Brain Processes," *The Philosophical Review* 68.2: 141-156
- Smith, Michael, 1994, *The Moral Problem*, Blackwell Publishing
- Smith, Michael, 1998, "Galen Strawson and the Weather Watchers", *Philosophy and Phenomenological Research* 58.2: 449-454
- Smith, Michael, 2011, "Scanlon on Desire and the Explanation of Action" in *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon* (Wallace, Kumar, and Freeman eds.), Oxford University Press: 79-97
- Smith, Michael, 2015, "The What and Why of Love's Reasons", Draft of July 2<sup>nd</sup>, 2015, downloaded from:  
<http://www.princeton.edu/~msmith/mypapers/TheWhatandWhyofLovesReasonsJULY2015.pdf> [cited with permission]
- Smithies, Declan, 2012, "The Mental Lives of Zombies", *Philosophical Perspectives* 26: 343-372
- Smithies, Declan, 2014, "The Phenomenal Basis of Epistemic Justification" in *New Waves in Philosophy of Mind* (Sprevak & Kallestrup eds.), Palgrave MacMillan: pp. 98-125
- Smithies, Declan and Jeremy Weiss, in progress, "Affective Phenomenology and Reasons for Action"
- Solomon, Robert C. and Lori D. Stone, 2002, "On 'Positive' and 'Negative' Emotions", *Journal for the Theory of Social Behavior* 32.4: 417-435

- Solomon, Robert, 2004, "Emotions, Thoughts, and Feelings" in *Thinking about Feeling: Contemporary Philosophers on Emotions* (Solomon ed.), Oxford University Press: pp. 76-88
- Sprigge, Timothy L. S., 1988, *The Rational Foundations of Ethics*. Routledge & Kegan Paul
- Stalnaker, Robert, 1984, *Inquiry*, Cambridge University Press
- Stevenson, C.L., 1937, "The Emotive Meaning of Ethical Terms", *Mind* 46.181: 14-31
- Stocker, Michael, 1979, "Desiring the Bad: An Essay in Moral Psychology", *The Journal of Philosophy* 76, 12: pp. 738-753
- Stoljar, Daniel, 2004, "The Argument from Diaphanousness", *Canadian Journal of Philosophy* 34.sup1: 341-390
- Strawson, Galen, 1994, *Mental Reality*, MIT Press
- Stump, Eleanor, 2006, "Love By All Accounts", *Proceedings and Addresses of the American Philosophical Association* 80, 2: pp. 25-43
- Svavarsdóttir, Sigrún, 1999, "Moral Cognitivism and Motivation", *The Philosophical Review* 108.2: 161-219
- Svavarsdóttir, Sigrún, 2014, "Having Value and Being Worth Valuing", *The Journal of Philosophy* 111.2: 84-109
- Svavarsdóttir, Sigrún, 2015, "Detecting Value with Motivational Responses" in *Motivational Internalism* (Björnsson, Strandberg, Olinder Eriksson, and Björklund eds.), Oxford University Press: 213-237
- Tappolet, Christine, 2010, "Emotion, Motivation, and Action: The Case of Fear" in *The Oxford Handbook of Philosophy of Emotion* (Goldie ed.), Oxford University Press: 325-349
- Tappolet, Christine, 2012, "Emotions, Perceptions, and Emotional Illusions" in *Perceptual Illusions: Philosophical and Psychological Essays* (Clotilde ed.), Palgrave-Macmillan: 207-224
- Thomson, Judith Jarvis, 1971, "A Defense of Abortion", *Philosophy & Public Affairs* 1.1: 47-66
- Tresan, Jon, 2009, "The Challenge of Communal Internalism", *The Journal of Value Inquiry* 43: 179-199

- Tye, Michael, 1995, *Ten Problems of Consciousness*, MIT Press
- Tye, Michael, 2000, *Consciousness, Color, and Content*, The MIT Press
- Velleman, J. David, 1992a, "The Guise of the Good", *Noûs* 26.1: 3-26
- Velleman, J. David, 1992b, "What Happens When Someone Acts?", *Mind* 101.403: 461-481
- Velleman, J. David. 1999. "Love as a Moral Emotion." *Ethics* 109, 2: pp. 338-374.
- Velleman, J. David, 2000, "Introduction" in *The Possibility of Practical Reason*, Oxford University Press: 1-32
- Watson, Gary, 1975, "Free Agency", *The Journal of Philosophy* 72.8: 205-220
- Wedgwood, Ralph, 2007, *The Nature of Normativity*, Oxford University Press
- Williams, Bernard, 1965, "Ethical Consistency", *Proceedings of the Aristotelian Society, Supplementary Volumes* 39: pp. 103-124
- Williams, Bernard, 1981, "Persons, Character, and Morality", *Moral Luck*, Cambridge University Press
- Williamson, Timothy, 2002, *Knowledge and its Limits*, Oxford University Press