Norm-Expressivism: Requirements & Possibilities for Moral Emotions

Kristin Klamm-Doneen
Norm-Expressivsm: Requirements & Possibilities for Moral Emotions
Narrow Moral Emotions and Broad Capacities

Kristin Klamm-Doneen
©Kristin Klamm-Doneen, Stockholm University 2012


Distributor: Department of Philosophy, Stockholm University
To Randall, and our children
Dean and Cilla.

To my mother, Ruth Klamm,
and the memory of my father,
Phil Klamm.

Till svenska släktingar.
# Table of Contents

1. Introduction ......................................................................................................................................... 1

2. Norm-Expressivism .......................................................................................................................... 12
   2.1 Introduction .................................................................................................................................... 12
   2.2 Meaning of a Moral Claim: What Does a Moral Claim Express? .................................................. 14
   2.3 Moral Emotions: Why Guilt and Resentment (Anger)? ................................................................. 15
      2.3.1 Other Moral Emotions Shape, Yet Are Insufficient Terms For, a Moral Judgment ............. 16
      2.3.2 Moral Judgments Identify Blameworthiness in Terms of Guilt & Anger ............................ 19
      2.3.3 Moral Judgments Promote Social Coordination Through Meshing Guilt and Anger ......... 21
      2.3.4 The Nature of Moral Emotions: The Circularity Problem ...................................................... 22
      2.3.5 Why This Matters ................................................................................................................ 23
   2.4 Accepting a Norm ........................................................................................................................ 25
      2.4.1 Avowal ..................................................................................................................................... 26
         2.4.1.1 Normative Discussion ....................................................................................................... 27
      2.4.2 Normative Governance ......................................................................................................... 28
      2.4.3 Why This Matters ................................................................................................................ 30
   2.5 Types of Moral Norms .................................................................................................................. 31
      2.5.1 Why This Matters ................................................................................................................ 34
   2.6 What it Means to Think Something “Rational” ........................................................................... 34
      2.6.1 Rationality as Endorsement .................................................................................................... 35
      2.6.2 Criteria for Rationality: Coherence, Plausibility, and Grounding Reasons .......................... 35
         2.6.2.1 Coherence ........................................................................................................................ 36
         2.6.2.2 Plausibility ........................................................................................................................ 39
         2.6.2.3 Grounding Reasons .......................................................................................................... 40
      2.6.3 Why This Matters ................................................................................................................ 41

3. Behaviorist Theories of Emotion ...................................................................................................... 45
   3.1 Introduction .................................................................................................................................... 45
3.2 Philosophical Behaviorism: Gilbert Ryle

3.2.1 Ryle’s Behaviorist Theory of Mind Leads to a Theory of Emotions

3.2.2 Ryle and Feelings

3.2.2.1 Would Rylean Feelings Suffice for Norm-Expressivism?

3.2.3 Ryle on Moods and Agitations

3.2.3.1 Would Rylean Moods Suffice for Norm-Expressivism?

3.2.4 Ryle on Motives

3.2.4.1 Would Rylean Motives Suffice for Norm-Expressivism?

3.3 Psychological Behaviorism: Radical Behaviorism

3.3.1 Would Radical Behaviorism Suffice for Norm-Expressivism?

3.4 Psychological Behaviorism: Paradigmatic Behaviorism

3.4.1 Would Paradigmatic Behaviorism Suffice for Norm-Expressivism?

3.4.1.1 The Limitation of Paradigmatic Behaviorism

3.5 Conclusion

4. Cognitivist Theories of Emotion

4.1 Introduction

4.2 Gibbard’s Circularity Concern

4.3 Attributional Theories of Emotion in Psychology (Emotion as Attribution of Emotion Concepts to Inner Experience)

4.3.1 Would Attributional Theories of Emotion Suffice for Norm-Expressivism?

4.3.1.1 Sinnott-Armstrong and Gibbard Debate (C3 and C1)

4.3.1.1.1 The Nature of the ( Alleged) Problem

4.3.1.1.2 An Attributional Account of Guilt, Independent of Moral Concepts

4.3.1.2 Sinnott-Armstrong and Gibbard Debate (Moral Rebels)

4.3.1.2. Attributional Theories & the Other Conditions

4.4 Nussbaum’s (Pure) Cognitivism: Emotion as Cognitive Judgment Only

4.5 Dimensional Appraisal Theories of Emotion in Psychology (Emotion as Bodily Response to Cognition)
4.5.1 Dimensional Appraisal Theory: Ortony, Clore and Collins.................................................................120
4.5.2 Dimensional Appraisal Theory: Lazarus .................................................................125
4.5.3 Would Dimensional Appraisal Theories of Emotion Suffice for Norm-Expressivism? ..............................................129
4.6 Solomon’s Engaged Judgmentalism: Emotion as Judgments of the Body ................................................135
  4.6.1 Would Solomon’s Engaged Judgmentalism Suffice for Norm-Expressivism? ........................................142
4.7 Greenspan’s Perspectivism .................................................................145
  4.7.1 Would Greenspan’s Perspectivism Suffice for Norm-Expressivism? ..............................................150
4.8 Conclusion ........................................................................................................154

5. Somatic Theories of Emotion ........................................................................156
  5.1 Introduction ........................................................................................................156
  5.2 The Beginnings of Embodied Appraisal Theories in Science .....157
  5.3 Robinson’s “Process Theory” of Emotion (Weak Somaticism) .................................................................164
    5.3.1 Nature of Modular Affective Appraisals .............................................167
    5.3.2 Robinson & Lazarus .............................................................................169
    5.3.3 Robinson’s Account of Emotion Typology .........................................173
      5.3.3.1 Robinson’s Account of Moral Emotions......................................176
    5.3.4 Would Robinson’s Process Theory Suffice for Norm-Expressivism? .......................................................180
  5.4 Prinz’s “Perceptual Theory” of Emotion (Strong Somaticism) .................................................................183
    5.4.1 Emotion as Perception ........................................................................185
      5.4.1.1 Perceptions of Core Relational Themes (Emotions as Representations) ...................................................189
    5.4.2 Prinz’s Account of Emotion Typology ..............................................191
      5.4.2.1 Prinz’s Account of Moral Emotions ............................................193
      5.4.2.2 Perceptualism & Cognitivism: Shared Compatabilities for Norm-Expressivism ............................................200
    5.4.3 Would Prinz’s Perceptual Theory Suffice for Norm-Expressivism? .......................................................201
  5.5 Conclusion .........................................................................................................204

6. Conclusion ........................................................................................................207

Works Consulted ......................................................................................................213
Ethics students frequently insist that moral judgments (somehow) come down to our emotions and yet, at the same time, they provide reasons for their own ethical viewpoints that they think others should adopt upon hearing them. I have long found this fascinating. It seems difficult for many people to shake the notion that moral judgments are grounded in both reason and emotions in some way. It’s as if the moral life seems to push us toward these two convictions as a matter of instinct. Gibbard’s metaethical analysis of the moral claim, proposing that *a moral claim just is a judgment about what it makes sense to feel*, offers a possible account of this instinct. Nonetheless, many facts have yet to be established about the nature of such emotions in our lives in relation to reason, as well as (and perhaps most importantly) in relation to social behavior. This work is an inquiry into one necessary condition for the cogency of norm-expressivism – whether the nature of emotion would render it possible for guilt and anger to work in the way in which Gibbard alleges. I’m happy for this work to contribute toward even that much insight into the plausibility of Gibbard’s account.

I am thankful to the Department of Philosophy at Stockholm University for its hospitality and patience in working with me from afar on this long-term goal. I thank the Department for reading drafts, and providing comments, on this work. Ragnar Ohlsson, Björn Eriksson, and later Åsa Carlson supervised this work. Ragnar Ohlsson first welcomed me into the Department of Philosophy (in their Practical Philosophy division) many years ago, and I appreciate his support as I begun this project and for many years thereafter. Åsa Carlson and Björn Eriksson have also generously advised me on many chapter drafts, and I thank them for sharing their time and insights.
I am also grateful to my employer, Anoka Ramsey Community College (Coon Rapids, MN), for sponsoring my travel and much needed time in order to complete this work.

May, 2012
Kristin Klamm-Doneen
1. Introduction

Norm-expressivism is a contemporary metaethical theory, developed by Allan Gibbard, which aims to propose a meaning for moral claims based largely upon the way in which humans employ such judgments in their lives. Gibbard proposes a meaning for moral claims, and he intends for this meaning to reflect the evolutionary role of moral judgments in the human life. He argues that the moral claim can be understood as the expression of a specific state of mind. As we will see, Gibbard takes the moral claim to express a state of mind in which one accepts norms governing moral emotions. Because his theory explains the meaning of a moral judgment with reference to moral emotions (guilt and anger in particular), and because he takes this meaning to be derived from the way in which moral judgments actually function in our moral lives, it would be important that moral emotions themselves are explainable in a way that is compatible with the nature of such moral judgments. As we will see, norm-expressivism has been criticized (in part) due to its implications about the workings of both moral emotions and moral emotion concepts. I find, though, that these criticisms are overstated. There are, to be sure, some hanging empirical questions that undermine the cogency of norm-expressivism and, yet, it is quite plausible that the norm-expressivist has things right in these regards.

I will argue that, upon surveying Gibbard’s theory and upon surveying theories of emotion, there are several theories of emotion that turn out to be consistent with norm-expressivism. Still, many of the theories that turn out to be consistent with norm-expressivism will show to be compatible only through (plausible) modifications of, or embellishments to, their theories. Thus, of those theories that are consistent with norm-expressivism, some of them make for a better fit – just as they are – than do others. Certain theories of emotion, that is, require more or less modification and/or embellishment in order to meet the norm-expressivist’s needs. Even with respect to the theories which fit quite well (as they are, with little or no amendments), these theories are only compatible with norm-expressivism if certain empirical facts turn out to be the case. It is plausible that such facts are the case, though these facts are also questionable. Thus, the cogency of norm-expressivism is in the end left de-
pendent upon such hanging empirical questions more so than an ability to theoretically account for moral emotions. There are, as we will see, several theories of emotion that provide frameworks to explain moral emotions in ways consistent with norm-expressivism.

Once compatible theories of emotion for norm-expressivism are identified, the merits of norm-expressivism (with respect to the role of emotions) would then depend upon the independent merits of those theories of emotions and the empirical facts around emotion types in such contexts. It is beyond the scope of this project to fully analyze the independent merits of all compatible theories of emotion and the empirical facts around the workings of the moral emotions. This project aims to provide norm-expressivism with the necessary conditions to support the norm-expressivist’s claim about the relationship between a moral judgment and moral emotions, and it aims to identify (among some of the major contenders in emotion theory today) the theories most coherent with norm-expressivism. Noting the contingency of such theories’ compatibility (with norm-expressivism) upon these empirical facts, we can then hypothetically grant a recognition of some theories of emotion which are quite fruitful for norm-expressivism. This investigation will lead to the conclusion that there are versions of behaviorist, cognitivist, and somatic theories of emotion that are all compatible with norm-expressivism.

The compatibility of these theories with norm-expressivism is due to the philosophical framework that these theories of emotion provide for emotion studies, although (as noted) philosophical frameworks alone will not determine the compatibility of norm-expressivism and the moral emotions. Though there are several theories of emotion that turn out to be compatible with norm-expressivism (certainly more than Gibbard anticipated in his original proposal), these same theories could also be consistent with an explanation of the moral emotions that turns out (upon an empirical investigation) to contradict norm-expressivism. That is, the compatible theories of emotion are consistent with norm-expressivism even though they do not logically imply (i.e. necessitate) the norm-expressivist’s views of the moral emotions. While these theories of emotion are compatible with norm-expressivism, in that they provide a framework that is consistent with the norm-expressivist’s understanding of the way in which the moral emotions must work, each of these theories can only support norm-expressivism in the end if the moral emotions turn out in fact to operate in specific ways. Empirical data is, then, quite relevant to solidifying the capacity of these theories to complement norm-expressivism. Thus, while we can survey theories of emotion and find that several of them are consistent with norm-expressivism,
their capacity to support/evidence norm-expressivism is ultimately dependent upon specific empirical facts which (though plausible) are left hanging. While the empirical nature of emotions is certainly relevant to the merit of norm-expressivism, the norm-expressivist will nonetheless find herself hard pressed to answer some of these most critical questions. For this reason, I find it fruitful to at least identify the theories of emotion which can complement norm-expressivism.

An example may illustrate the relevance, and yet challenges, of empirical claims with respect to norm expressivism. Shaun Nichols has critiqued norm-expressivism with respect to the implications of empirical facts around moral emotions. He cites empirical research demonstrating (on his interpretation anyhow) that children make distinctions between moral and non-moral/conventional offenses prior to understanding moral emotion concepts (Nichols 90). That is, small children can determine that an action is morally wrong, in contrast to just conventionally wrong, prior to understanding the concept of guilt. If this is so, then it would seem that moral judgments cannot require that one comprehend moral emotions, since children make moral/non-moral distinctions prior to grasping moral emotion concepts. If these findings are correct, then moral judgments cannot be about moral emotions as purported by Gibbard. Nichols cites a study by Gertrud Nunner-Winkler and Beate Sodian, in which they report on children’s inability to anticipate the guilt that offending children feel in response to their misdeeds (Nichols 90; Nunner-Winkler and Sodian 1329). This proves, in Nichols’ view, that these children do not understand guilt concepts, even though he notes, per studies by Judith Smetana and Judith Braeges, that young children can already distinguish between moral and non-moral/conventional offenses (Nichols 10, 90). Smetana and Braeges’ research framed questions in terms of a moral/conventional distinction; on this view, judgments which were taken to be more serious, pan-cultural and independent of authority were taken to reflect a moral judgment, whereas judgments that were taken to be less serious, culturally determined, and dependent upon authority were taken to reflect a conventional judgment. Smetana and Braeges inferred their conclusion from the way in which children responded to questions about “this school” and “another school”; the children generally recognized the type of action that is wrong in a conventional way (noting that it would be wrong at their own school, but not at another one where it was allowed) whereas they identified actions typical of moral transgressions (that involve harming another) to be wrong at another school too (342). This is taken by Nichols to prove that children can make moral judgments at a young age.
It is unclear, though, whether these studies have conclusively determined that children lack such emotion concepts at the time in which they (seemingly) make moral judgments. Though Nunner-Winkler and Sodian report that children may not attribute these emotions to others who are purported to have done misdeeds, there are other studies that critique such conclusions. S.J. Murgatroyd and E.J. Robinson, for example, report on their research which set out to “examine contradictions between three published accounts of age-related changes in children’s judgments of the emotion experienced by a wrongdoer” (873). They report that their research contradicts Nunner-Winkler and Sodian’s findings, in that Murgatroyd and Robinson’s research demonstrated that the incidence of failures to attribute guilt to wrongdoers “did not decline with age” (873). Perhaps some people just tend not to assume that others who do wrong will feel guilty about it. Even for these individuals, though, this need not imply that they fail to understand the feeling of guilt. (Maybe they’re just cynics, or unbelieving anyhow, with respect to others’ responses. Perhaps all people just periodically see things in this way, at roughly the same frequency.) Studies with young children are especially challenging because young children may, of course, understand more than they can articulate, and it is challenging to “show” an understanding of guilt without relying upon verbal skills. If Nunner-Winkler and Sodian’s studies fail to prove such conclusions, then they may not evidence the critique of norm-expressivism put forth by Nichols.

Gibbard could also respond to such data reported by Smetana & Braeges, and cited by Nichols, by arguing that such studies fail to show that the children in these studies are really making moral judgments. Whereas some studies have reported similar findings on the capacity of young children to make the conventional/moral distinction, others have conducted studies demonstrating flaws in the conventional/moral task paradigm used in such research. One of these latter is Kelly, et al. whose research concludes, “If, as our results suggest, the moral/conventional task is not a good assay for the existence of a psychologically important distinction, then the reasoning behind those conclusions merits careful scrutiny indeed” (130). Their research suggests that, when expanding the type of questions (beyond what they call “schoolyard” harms), they find that the criteria used to make the moral/conventional is not found to hold (Kelly et al. 129). The subjects in their studies do not consistently judge

---

1 Smetana, in “Understanding of Social Rules,” reviews literature on the capacity of children to make such a distinction.

2 Also see Turiel et al. for exceptions to the hard distinction between moral/conventional judgments based on the cluster criteria. These authors maintain that the criteria are good for the most part, though they recognize exceptions.
actions according to clusters of seriousness, universality, and independence of authority, on the one hand, and then less seriousness, culture, and authority on the other hand. If this is so, then we cannot measure the capacity for moral judgments according to moral/conventional tasks because people do not, in fact, make judgments of these sorts according to the domains which serve as the criteria for this distinction. Kelly et al.’s studies found that, when they provided questions that depict a much wider range of harm transgressions, “they do not evoke the signature pattern of responses found in studies using only schoolyard transgressions” (117). The studies concluding that young children make moral judgments are inconclusive, in my view, due to the limited nature of the circumstances that they judge. As we will see (in Chapter Two), Gibbard’s analysis of the moral judgment suggests that a moral judgment need not refer to a unified cluster of criteria, and thus the norm-expressivist would likely not find such studies (as cited by Nichols) to reflect any clear facts about the way in which people make moral judgments at all. Whatever kind of judgments these young children are making, it is clearly a different sort of judgment than the moral judgment as depicted by Gibbard.

My point here is that, while empirical data on the moral emotions (and the understanding of moral emotion concepts) may be relevant to critiques of norm-expressivism, the interpretation of this data is by no means conclusively solved. The interpretation of such research will depend upon the way in which we conceptualize the moral judgment and the moral emotion itself. Thus, because our inquiry into the empirical facts about emotion and moral judgments would seem to presuppose a theory of emotion and a metaethical theory (which depicts the nature of a moral judgment), and because the data itself is still far from finding a consensus in psychology, I find it important to the merits of norm-expressivism to consider whether there are plausible philosophical frameworks for use in the norm-expressivist’s analysis of moral emotions. Although empirical findings matter, the norm-expressivist cannot simply look to the sciences alone to indicate what emotions are and how they interact with moral judgments. A conceptual framework of emotions and moral judgments is needed for this analysis to take place. Thus, the norm-expressivist can best establish the compatibility of moral emotions with norm-expressivism through a survey of theories of emotion. Furthermore, since there are competing theories of emotion in emotion studies today, it will be good news for the norm-expressivist if her theory is compatible with more than one of these theories. The more compatible theories of emotion, the better for norm-expressivism. Thus, my work here will be dedicated to surveying and looking for compatible theories of emotion for norm-expressivism.
Chapter Two will introduce norm-expressivism. It will explain, in some detail, the meaning of a moral claim in Gibbard’s view, as well as his argument for proposing this meaning. Gibbard takes moral judgments to be evolved capacities, capacities that human beings avail in order to promote social coordination. A moral judgment is ultimately said to express our acceptance of feeling norms, and our acceptance of the norms expresses our view that such norms are rational. Gibbard proposes a very broad meaning for judgments of rationality. As we will see, “rationality” (in norm-expressivism) is taken to imply an endorsement of a position or commitment in response to reasons.

Gibbard proposes that moral claims are naturally employed to express judgments about the rationality of, specifically, guilt and anger. The moral claim is then interpreted (upon this theory) to be the claim that it makes sense to feel angry at others who violate a moral norm and that it makes sense to feel guilty whenever it is oneself who violates the norm. If moral judgments are explained by what we think it makes sense to feel in terms of guilt and anger, then guilt and anger cannot be moral judgments. Any acceptable theory of emotion will need to avoid such circularity. If moral emotions just were moral judgments (or if they were necessarily caused by, or defined in terms of, moral judgments or even moral concepts), then Gibbard’s proposed meaning for the moral claim becomes circular. Also, if Gibbard’s evolutionary argument for this understanding of the moral claim is correct, then these moral emotions (anger and guilt, as he sees it) would need to be subject to the influence of such moral judgments. As we will see, Gibbard notes that the right theory of emotion for the norm-expressivist would need to explain the emotion in light of empirical data and evolutionary theory, while avoiding this circularity challenge and allowing for their normative governance.

Norm-expressivism has been criticized for failing to meet such necessary conditions. Walter Sinnott-Armstrong has critiqued norm-expressivism on the basis of its claims with respect to the moral emotions (308-310). As we will later see in greater detail (Chapter Four), Sinnott-Armstrong argues that moral claims could not be judgments about the rationality of guilt and anger, as Gibbard maintains, because we can only reference a guilt concept (when making a moral judgment) through an understanding of moral concepts themselves. Short of recognizing a moral emotion in terms of its implicit moral judgment (identifying guilt, for example, as a response to the finding that one has committed a moral transgression), we could not even have a concept of guilt. Sinnott-Armstrong maintains that I can only recognize guilt, in contrast to other emotions like embarrassment, by recognizing the behaviors that elicit guilt as moral failures rather than failures in etiquette or some other sort
perhaps. If my moral claim refers to an emotion that itself refers to a moral judgment, though, then the norm-expressivist’s account of the moral judgment is circular. Thomas Hill has expressed this same criticism of norm-expressivism as follows:

Gibbard’s strategy is to understand judgments that acts are wrong by appealing to norms for the aptness of feelings of guilt, and so he must explain feeling guilty as something other than judging that one has done wrong. Though we often feel residual, misplaced, or indeterminate guilt, I still think that we understand these deviant cases by reference to the standard case in which a crucial part of feeling guilty is judging that one is guilty, i.e., that one has done wrong (960).

If norm-expressivism is to offer a cogent analysis of the moral judgment, then this critique with respect to the workings of the moral emotions must be refuted. This will require that there is some way in which moral emotions are identifiable, and definable, without the use of moral concepts. Any theory of emotion that fails to provide for this would undermine the credibility of norm-expressivism by rendering it circular.

Through an understanding of the central tenets of norm-expressivism in Chapter Two, I will propose seven conditions for a compatible theory of emotion. That is, I will identify conditions that would need to be met by any theory of emotion that is consistent with the workings of our moral lives according to norm-expressivism. Should a theory of emotion fail to meet all of these criteria, then the theory of emotion is inconsistent with norm-expressivism. Should no theory of emotion prove to be consistent with these conditions, then norm-expressivism would lose credibility due to its expressivism. While it is possible that norm-expressivism has things right with respect to the moral emotions, and that nonetheless no current theory of emotion can account for these facts, an investigation into the merits of norm-expressivism should include an inquiry into what the best theories of emotion today offer in the way of confirming such implications. Should no theory of emotion be consistent with norm-expressivism, this leaves the credibility of norm-expressivism highly suspect. Should the nature and typology of moral emotions turn out (on all accounts) to be inconsistent with Gibbard’s proposal for how they function in our moral lives, then Gibbard’s account of the moral claim would have to be wrong. It is important, then, that the moral emotions (of guilt and anger in Gibbard’s view) turn out to be explainable and analyzable in terms consistent with the main tenets of his theory. Chapter Two will explain these central tenets of norm-expressivism and
it will provide conditions to guide our survey of possible theories of emotion.

A theory of emotion that meets all seven conditions for a compatible theory of emotion (as proposed in Chapter Two) will go far in addressing these above critiques. Even then, while I do not find Nichols’ empirical critiques to be conclusive blows to norm-expressivism by any means, I admit that critical empirical questions (for the time-being anyhow) remain. If empirical data somehow falsified the claim that moral emotions can occur independently of moral judgments, that would clearly discredit norm-expressivism. Since, though, I have yet to find any such conclusive data, and since I find that such studies themselves presume a theory of emotion (and even imply metaethical claims) in order to proceed in recording the way in which these emotions operate, I suggest that the norm-expressivist is best to begin addressing the nature of emotions by identifying compatible theories of emotion. Thus, I move on to survey diverse theories of emotion.

Chapter Three explores the compatibility of behaviorist theories of emotions and norm-expressivism. This chapter will introduce philosophical behaviorism (as developed by Gilbert Ryle) and psychological behaviorist theories of emotion (as developed by John Watson and B.F. Skinner). Included in psychological behaviorism, and yet distinct from the “radical behaviorism” of Watson and Skinner, will also be a discussion of a more recent “paradigmatic behaviorist theory of emotion” (as developed by Arthur Staats and Georg Eifert). Paradigmatic behaviorism identifies emotions as central nervous system responses analyzable according to a heuristic that draws upon many fields of science. I will argue that a paradigmatic behaviorist theory of emotion is consistent with the conditions for a compatible theory of emotion for norm-expressivism, although the theory’s account of emotion typology is in need of amplification via other theories of emotion.

Chapter Four will consider the compatibility of cognitivist theories of emotion and norm-expressivism. Cognitivist theories purport that emotions either are, or are necessarily caused by, some sort of cognition. These theories are quite varied – from attributional theories, to Martha Nussbaum’s pure cognitivism, to dimensional appraisal theories (Andres

---

3 Arthur Staats has recently shared, via email, that he now prefers for his “Paradigmatic Behaviorism” to be called “Psychological Behaviorism”. That being said, I will refer to it as Paradigmatic Behaviorism, in Chapter Three, in order to reflect his specific writings on it. In calling it psychological behaviorism now, though, Staats still maintains the same central tenets put forth in his writing on paradigmatic behaviorism. He intends to propose a behaviorist theory of emotion that clearly distinguish itself from Watson and Skinners’ “Radical Behaviorism.”
Ortony et al. and Richard Lazarus), to Solomon’s engaged judgmentalism, to Patricia Greenspan’s perspectivism. These theories may offer the advantage (over behaviorist theories) of accounting for the meaning (or content) of emotion concepts, and they are supported (to varying degrees) by work in experimental psychology. These theories do not, though, connect in any specific way with the neurological findings regarding the mapping of bodily changes in emotion nor the distinct levels of emotional processing. Still, that is more of a challenge (if it is one) with respect to the independent merits of these theories. In terms of meeting the conditions for a compatible theory of emotion, some of these theories will turn out to be compatible with norm-expressivism. For those cognitivist theories that employ a sense of “cognition” which is especially broad (including unconscious evaluations, modular evaluations, and even evaluative abilities shared with other animals), the moral emotions need not be (nor be caused by) moral judgments in Gibbard’s sense even if they are evaluations which respond to objects of moral concern. It is plausible that we may even conceive of these emotion concepts in non-moral terms. I will argue that all cognitivist theories would allow for moral emotions to be elicited in ways that respond to their objects independent of any moral judgment, so long as these theories are either embellished or modified in specific ways.

Chapter Five will consider somatic theories of emotion. Jenefer Robinson’s process theory and Jesse Prinz’s perceptual theory both maintain that emotions are embodied responses. That is, emotions are appraisals that entail either bodily responses or at least the perception of such bodily change. Robinson argues that, through a modular and non-cognitive appraisal of an object, we experience bodily responses that are then monitored through cognitive reappraisals. As a non-cognitivist, Prinz argues that the body can respond directly to the perception of objects, and thus individuals (on his view) have the ability to track objects in light of themes that will define emotion types even when the individual fails to grasp (in any cognitive way) the emotive concepts of which these changes represent. Both of these theories purport that emotions can be appraisals of objects that take place without higher-level thought processes of the sort involved in moral judgments. In the end, I argue that both of these theories meet all of the norm-expressivist’s conditions, and they do so without modifying their claims about the moral emotions. Prinz’s theory, in particular, would not even require embellishment in order to be compatible with norm-expressivism. That, then, would leave the question as to whether his theory of emotion is in fact the best one (or an equally good one) in terms of its independent merits, but with respect
to my task, this ideal fit otherwise would be good news for the norm-expressivist.

While some cognitivist theories of emotion may meet the conditions of a compatible theory of emotion for norm-expressivism, the paradigmatic behaviorist theory of emotion and the somatic theories emphasize the empirical data with respect to bodily changes in emotion more explicitly than do the compatible cognitivist theories. Still, the difference between the compatible cognitivist theories and the somatic theories seems to be largely one of how they define “cognition.” On many points, I am not convinced that they are as far apart as they sometimes claim to be from one another. Still, I try to consider these theories in their own terms (rather than critiquing them or modifying them) in order to consider how their frameworks would pair up with norm-expressivism. All acceptable theories of emotion for the norm-expressivist, upon my analysis in this project, account for the fact that emotions are about both their particular objects and their more generalizable themes (e.g. one’s joy at getting a raise at work is about “this raise” being “a fortuitous event”), and they account for the phenomenological (and bodily) quality to emotions in some way or another. Most critically for the norm-expressivist, though, these theories (in their own ways) each meet the seven conditions for compatibility with norm-expressivism.

If I were to place a bet on the best suited theory of emotion for norm-expressivism at this time, I would recommend the somatic theory of emotion put forth by Prinz. The somatic theories of emotion better account for the empirical research found largely in neuroscience today regarding the bodily changes in emotion studies than do cognitive theories, and they offer an account of emotion typology that addresses (or even completes) many issues left open in paradigmatic behaviorism. I suspect that the benefits of somatic theories of emotion, for norm-expressivism, are found in the independent merits of these theories, specifically in their holistic account of the ways that multiple fields of science may be used in the fullest explanation of emotion (and emotions). I suspect that their interdisciplinary account of emotion would sit well with Gibbard’s naturalism. Whether or not somatic theories conceptualize “cognition” in the best way (which is indeed questionable), they offer sophisticated accounts of the body’s role in emotion, the evolution of emotion types, and the distinctly different ways in which people go about evaluating their environment. These somatic theories seem to offer much of the intent found in paradigmatic behaviorist theories of emotion, by offering a theory of emotion that integrates many fields of research together. The somatic theories, though, would seem to be directly enriched by some of the philosophical arguments unavailable to behaviorists.
suggest, then, that somatic theories not only meet the norm-expressivist’s conditions, but they may also (in fact) offer a strong theory of emotion as well. Furthermore, whereas the other theories of emotion are taken (in my work) to be compatible with norm-expressivism given modified or additional (but consistent) arguments, Prinz alone offers these needed arguments in explicating his own theory. In this way, his theory takes the lead in this quest for a compatible theory of emotion for norm-expressivism.

Clearly, this project does not include a survey of every theory of emotion. Other theories of emotion may turn out to offer a match for norm-expressivism as well. I focus upon theories of emotion that are (based upon my research) main contenders recognized in contemporary emotion studies. Whatever else such additional theories of emotion may offer to the norm-expressivist, I hope this project demonstrates that there are some major theories of emotion that are quite consistent with norm-expressivism. Clearly, this project is not a thorough defense of each surveyed theory of emotion, but it does suggest that (should these theories turn out to be sufficiently cogent) the norm-expressivist can rest assured that the nature of moral emotions need not threaten the credibility of norm-expressivism.
2. Norm-Expressivism

2.1 Introduction

In his book *Wise Choices, Apt Feelings*, Allan Gibbard proposes a novel metaethical theory which he coins “norm-expressivism” (8). The purpose of this work is, largely, to propose a semantic theory for moral language which is informed by empirical data. The purpose of this chapter is to explain, not analyze or critique, this theory so that we can then (in later chapters) consider compatible theories of emotion. In doing so, I will here explain aspects of norm-expressivism and identify conditions for a compatible theory of emotion in light of these features.

As an expressivist, Gibbard aims to explain the meaning of moral terms like “ought” and “wrong” through a description of one’s “state of mind” (41, 45, 9). Gibbard proposes a meaning for moral judgments based upon the state of mind that he takes such judgments (or moral claims) to express. *Norm-expressivism*, as we will see, identifies this state of mind as the “acceptance of norms” which governs specific moral emotions (9). That is, when one makes a moral claim (such as “Stealing is wrong”), one expresses one’s acceptance of norms governing specific moral emotions and this acceptance will, in turn, motivate behaviors. If we are to understand Gibbard’s theory, though, much needs to be explained about the nature of these moral emotions, the nature of this acceptance, and the nature of such norms. This chapter will explain central tenets of norm-expressivism, noting the ways in which these aspects are relevant to my project of seeking a compatible theory of emotion. Below, I will explain the norm-expressivist’s account of the meaning of a moral claim in more detail, as well as Gibbard’s understanding of moral emotions, what it is to be an accepted norm, the types of moral norms we purportedly accept, and the way in which moral judgments are judgments of rationality. In doing so, I will identify criteria (or conditions) for a

---

4 In this chapter, unless otherwise noted, all page citations are from Gibbard’s book *Wise Choices, Apt Feelings* (Harvard University Press, 1990).
well-suited theory of emotion for norm-expressivism. Throughout this discourse, we should note the important ways in which the norm-expressivist’s central claims are taken to connect with broad empirical sciences.

While Gibbard’s theory aims to provide a theory of meaning for moral language, his theory is often noted for its naturalist bent. As he says, “I treat meanings as a part of nature, and so factual” (“Reply to Blackburn, Carson, Hill and Railton” 972). Gibbard’s point here is not that moral claims are factual claims, but rather that the meaning of a moral claim can be derived from the facts about how such claims operate in our moral lives. As we study his theory, we should understand that his “goal...(or one of them), is to try out a way of conceiving meaning and normative life as part of nature, and to see how much that might explain” (35). The empirical sciences will, in his view, help us to correctly ascert

tain the meaning of moral claims. Though we may well describe the meaning of our own claims differently, Gibbard is interested in the way in which our moral claims promote social coordination, and this is the key to an understanding of their factual meaning in his view. As he explains it, “Normative judgments are to be explained psychologically, and so I need a psychology” (26). His psychology draws upon a matrix of studies into the human life. Thus, he claims that “we should work for as richly successful theories as we can get – theories that combine biological rigor with the humanistic eye for the complexities of the human psyche. We should draw on the anthropologist’s sense of the scope of cultural variation and the typical patterns of human life, and on analytical tools from the social sciences” (27). In the end, he believes that the sciences, and evolutionary theory in particular, provide a picture of the way in which moral claims operate in our lives and thus the states of mind they imply. A compatible theory of emotion for norm-expressivism, then, will explain the nature of moral emotions in ways consistently supported across broad empirical science. By understanding the way in which moral claims function in our lives, we are prepared to engage in the conceptual analysis needed to articulate the meaning of these moral claims. The meaning of a moral claim will need to be fully consistent with scientific evidence regarding the moral emotions.

Below, I will detail Gibbard’s theory. As I proceed to explain the major tenets of norm-expressivism, we should note that all of these details are taken by Gibbard to proceed from his naturalism. Thus, if the facts around our moral lives turn out to be significantly different from the ways in which Gibbard supposes, then norm-expressivism loses credibility. This project will suggest that, insofar as a theory of emotion is con-
cerned anyhow, the sciences would actually lend credibility to norm-expressivism.

2.2 Meaning of a Moral Claim: What Does a Moral Claim Express?

According to norm-expressivism, moral claims express our acceptance of moral norms. Moral norms are taken to prescribe specific moral emotions in response to behaviors. As Gibbard puts it, “moral norms in particular are norms for the rationality of guilt and resentment” (47). For example, the moral claim that “stealing is wrong” is taken to express the view that “one ought to feel guilty for stealing and others ought to feel angry at those who steal.” Gibbard proposes that moral claims express our acceptance of norms governing the specific moral emotions of “guilt” and “resentment” (42). Nothing in the moral judgment tells us how strongly to feel, and this is because the moral claim expresses judgments about these emotions from a particular standpoint. Because moral norms “set aside the question of how fully engaged one’s feelings are to be,” Gibbard claims that “they are norms for how to feel given full, impartial engagement” (127). Our example may be clarified then: the claim that “stealing is wrong” should be taken to express my acceptance of the norm that “from a standpoint of full engagement one ought to feel guilty for stealing and from a standpoint of full, impartial, engagement one ought to feel resentful at those who steal.” The fact that we make moral claims is taken, in norm-expressivism, to indicate our capacity to make such standpoint dependent judgments; that is, when making a moral judgment, we adopt the standpoint of impartial engagement. Still, while the norms are about guilt and resentment from this standpoint, we find these norms to influence (and even guide) our actual emotional responses in many circumstances. In judging an act to be wrong, one judges that (from this standpoint) it makes sense to feel resentful at those who commit the act and it makes sense to feel guilty if one commits such an act. We do not judge how strongly one should feel until we know one’s actual standing.

A central point in all of this is that, in accepting moral norms, we find that it is “rational” to feel guilt and resentment in response to the circumstances in question (9). Below, I will discuss what Gibbard means

---

5 Gibbard later acknowledges, though, that “guilt” is prescribed from a standpoint of full engagement only. Guilt is judged to be rational from a position of full engagement, though it is not impartial since it is a response to one’s own behavior (Wise 127).
by “rational” in more detail, but for now we can note Gibbard’s claim
that to call something rational is to say that the moral emotions “make
sense” in response to the behavior (49). In expressing my acceptance of
the norm that “one ought to feel guilty for stealing and others ought to
feel angry at theft,” I am expressing my judgment that these emotions
make sense in response to theft. A moral claim, then, expresses a judg-
ment of rationality for it is the judgment that the moral emotions make
sense (and thus we accept norms that call for them).

While (on Gibbard’s view) the moral claim is taken to express a
judgment about the rationality of moral emotions, moral judgments are
taken to indirectly govern behavior as well. While Gibbard claims that
“Narrowly, we might try saying, morality concerns moral sentiments: the
sentiments of guilt and resentment and their variants,” he also admits that
“Broadly the moral question is how to live” (6). Because the norms di-
rectly prescribe emotions (even if to uncertain degrees), they indirectly
motivate behavior as well. Thus, moral claims should be taken to express
our acceptance of norms that directly govern moral emotions and indi-
rectly govern behavior. Because of Gibbard’s naturalism, this would
imply that our moral judgments do in fact have some tendency to moti-
vate such emotions and behaviors.

For now, though, let us proceed to understand his argument for
the moral emotions that he identifies, as well as his argument for identi-
fying these particular moral emotions in his analysis of the moral judg-
ment. We will then proceed to an explanation of their acceptance, the
normative system that produces them, and the sense in which they are
taken to be “rational.” An understanding of this will help us to identify
the conditions for a theory of emotion friendly to norm-expressivism.

2.3 Moral Emotions: Why Guilt and Resentment
(Anger)?

Gibbard maintains that a moral judgment expresses our acceptance of
norms governing guilt and resentment. Given his naturalism, his point is
not that moral claims should express judgments in terms of guilt and
resentment, but rather that moral claims do express judgments that can be
interpreted in such terms. As he sees it, a study into the way in which
moral claims function in our lives will reveal this to be the case. If I
make the moral claim, for example, that “pet starvation is wrong” (or that
“you shouldn’t starve a pet”), then I express my acceptance of the norm
that “one should feel guilty about starving a pet and others should feel
resentful at one who starves a pet.” Gibbard suggests that “resentment”
may not be quite the right term, though. He explains that resentment “suggests feeling harmed or offended, whereas the moral feeling can be impartial” (126). I need not know the starved pet, of course, in order to find moral emotions to be warranted in response to its mistreatment. Gibbard provides the following explanation for regarding anger and guilt as the main moral emotions:

Resentment, indignation, and the like seem to be kinds of anger. Perhaps, then, we should stick to talk of anger: a person is to blame for something he has done if it makes sense both for him to feel guilty for having done it and for others to be angry with him for having done it. Moral convictions, we are then saying, consist in norms for anger and for the first-person counterpart of anger, guilt. (126)

Thus the primary moral emotions, or the emotions referenced in a moral claim, are taken to be guilt and anger. These are the emotions directly relevant to the moral judgments we express.

Important to our purpose, though, we should understand why Gibbard takes the moral judgment to be interpreted in terms of guilt and anger. This will have implications upon the right theory of emotion for norm-expressivism. In doing so, we will gain an understanding of Gibbard’s parsimony around the moral emotions at large. Gibbard recognizes that many emotions shape our moral lives, and yet he believes that moral claims express our acceptance of norms particularly governing guilt and anger.

2.3.1 Other Moral Emotions Shape, Yet Are Insufficient Terms For, a Moral Judgment

One might wonder why the moral judgment must be conceived of in terms of guilt and anger rather than other moral emotions. While Gibbard believes that “Moral concern involves feelings,” he admits that there are many “broadly moral sentiments” (255). He notes that there are many feelings that “constitute sources of moral concern, and many of them we can count as broadly moral sentiments: benevolence, a sense of fairness, respect, feelings of worth, and the like” (256). These feelings seem to carry their own rationale (that is, they seem to make sense in certain circumstances more than others), as do guilt and anger, and they tend us toward action in response to their rationales. Gibbard argues, though, that the broadly moral sentiments serve a secondary role in our moral lives. Broadly moral sentiments stand in contrast to the narrow moral emotions
of anger and guilt, in that only guilt and anger function properly as the terms of a moral norm.

Gibbard regards the broadly moral sentiments (e.g., pity, benevolence, a sense of fairness, respect) to be broadly moral sentiments because they provide “a rationale for giving the narrowly moral emotions some particular shape. They induce us to govern anger and guilt with one set of norms rather than another” (256). For example, assuming that people have some natural tendencies to feel benevolence toward one another, we may then find it rational to feel guilty about denying simple assistance to those in serious need. That is, if I naturally feel some benevolence toward others, then I may find that it makes sense to feel guilty about denying such assistance to those in need. In this case, though, Gibbard explains that “Benevolent feelings . . . push us toward benevolent norms for guilt and anger – norms, perhaps, whose acceptance would foster benevolent goals” (256). Broadly moral emotions feed into norms for guilt and anger; they may have independent values of their own in terms of evolutionary advantage, but in terms of a moral judgment they play a supportive rather than constitutive role.

The broadly moral emotions do not serve as the terms of a moral judgment because their rationale does not necessarily translate into the rationale for a moral judgment. One may, for example, find that giving to charity is the benevolent thing to do and yet not find that abstaining from a donation is morally wrong; I may think that a donation would be a benevolent act, and yet think that one need not feel guilty about not giving. It seems intelligible enough to say something like “Feelings of benevolence would promote helping, but you’re not morally wrong to abstain from this charity.” Benevolence, then, cannot provide the rationale by which we make moral judgments. The same could be said of any of the broadly moral sentiments; they are not sufficient for providing the terms of a moral judgment because one may find that those emotions make sense in terms of their own rationale, and yet find that one is not morally obliged to act in accordance with those feelings. Only a finding of rationale for guilt and anger will imply, in Gibbard’s view, a moral judgment. Whereas benevolence does not suffice for a moral judgment, it is odd indeed (or so Gibbard thinks) to say that “one should feel guilty for stealing, but it is not wrong to steal.” A finding of rationale for guilt and anger imply a moral judgment. Other moral emotions appear, then, to serve a supportive role in choosing one’s norms for guilt and anger.

Gibbard argues that there is probably an evolutionary explanation for the secondary role of these emotions. Moral judgments have not come to be defined in terms of positive emotions (like benevolence calling us to care for others or pity calling us to at least relieve their suffer-
ing) because positive feelings, he suggests, are “too weak to account for a socially effective morality” (260). Evidence, as he sees it, just does not suggest that positive emotions are capable of motivating action with the same force as negative emotions. Thus, Gibbard suggests that, “When love fails, bad feelings about bad things we might do can still restrain us” (295). Perhaps one does not want to honor a commitment out of love (or pity or fairness, etc.) alone in a given moment, but in such cases then guilt may suffice. A child, for example, may well love his parent and, yet, still be quite unmotivated to stop begging for things that he knows the parent can’t afford. If Gibbard is right, then the child may be more inclined to stop this harassment, of sort, if guilt kicks in. This is not to deny that one could feel guilty and still do it. The child may feel guilty and keep begging because the child is in the grip of other norms or desires. Gibbard’s point, though, seems to be that guilt and anger are the moral emotions which carry the strongest motivational force in general, and thus (in part for this reason) they are ultimately the moral emotions to which we refer when making a moral judgment. Thus, he finds that there is a biological explanation for why we make moral judgments in terms of negative emotions. From an evolutionary perspective in his view, moral judgments would seem to function best if they primarily express our acceptance of norms governing negative emotions because negative emotions are more powerful in guiding actions.

Still, we might wonder whether other negative moral emotions might not equally serve this role in the moral judgment. Gibbard argues, though, that the rationale for guilt and anger just is the rationale for a moral judgment, and that guilt and anger mesh particularly well with one another in order to support “cooperative schemes” (138). For these reasons, as we will now see, guilt and anger are taken to be the emotive references of a moral judgment.

---

6 Indeed, this would be a critical empirical claim to further investigate in the analysis of norm-expressivism. Gibbard admits that “nothing I say disproves the claim that all our moral concern stems from wishing everyone well” (Wise 260), and yet he suspects that feelings of benevolence alone are not sufficient as the reference of a moral judgment. He claims that, “much of life suggests that when benevolence does spur action, it does so with the help of other sentiments” (Wise 260). All things considered, the most likely (truly sufficient) references for the moral judgment in his view are guilt and anger. Should the empirical sciences disprove this, though, then the narrow moral emotions in norm-expressivism would surely need revision.
2.3.2 Moral Judgments Identify Blameworthiness in Terms of Guilt & Anger

Gibbard argues that, when we judge an act to be morally wrong, we judge it in terms of “prima facie blameworthiness” (47). The warrant of guilt and anger, he purports, is the only rationale for blameworthiness. Thus, an act is taken to be morally wrong whenever it would make sense to feel guilt on the part of the agent and anger by others whenever relevant extenuating circumstances are lacking. Thus, Gibbard states that “to think an act wrong is to accept norms for guilt and resentment [or anger] that, prima facie, would sanction guilt and resentment [or anger] if the act were performed” (47). Karen, for example, may make the moral judgment that “It is wrong to abuse others.” Thus, she accepts that it makes sense for people who abuse others to feel guilty in ordinary circumstances, and she accepts that it makes sense for others to feel angry at abusers in any ordinary circumstances. In this case, if Peter abused his wife one day because he lacks the motivation to treat her humanely, then Karen’s norm would imply that it is rational for others to feel angry at this behavior, and she would find it rational for Peter to feel guilty. In accepting this norm, Karen would logically regard the abuse as wrong and Peter as actually blameworthy. If, on the other hand, Ted abused his wife because of a brain injury he received when forced into battle, then Karen may assess his actual blameworthiness differently. If Ted’s condition is taken to constitute extenuating circumstances, then her norm does not direct her (even from a fully engaged standpoint) to actually blame him for this behavior. The behavior is still wrong according to her moral judgment because, had he been of the right mind, then others should be angry at him and he should then feel guilty; thus the action is wrong. Ted is not actually blameworthy though because anger and guilt are not in fact prescribed for extenuating circumstances; if his abuse were truly outside his control, then no one ought to be angry at him, he ought not feel guilty, and he is not in fact blameworthy. Gibbard maintains that, regardless of one’s judgment about actual blameworthiness, moral judgments are made in terms of guilt and anger because they are made in terms of prima facie blameworthiness.

Gibbard thus explains that “to think an act morally reprehensible [or blameworthy] is to accept norms that prescribe, for such a situation, guilt on the part of the agent and resentment [or anger] on the part of others” (47). Regardless of the actual blameworthiness, norms for wrongness appear to govern prima facie guilt and anger. Thus, in this sense, Gibbard maintains that “Norms for wrongness are thus explained in terms of norms for guilt and resentment [or anger]” (47).
Other negative emotions would not suffice for the purpose of a moral judgment. We would not say, in Gibbard’s view anyhow, that an agent is blameworthy whenever it makes sense for the agent to feel any negative emotion about oneself (such as embarrassment or shame). An agent is only (prima facie) blameworthy when it makes sense for the agent to specifically feel guilty about a behavior and for others to specifically feel angry at the agent’s behavior. While other negative emotions function to support social coordination, Gibbard argues that guilt and anger are particularly well suited to be the objects of moral norms because only their rationale indicates a judgment of blameworthiness. He notes that shame and disdain, for example, are closely related to guilt and anger. He considers the possibility, for example, that “Shame stems from things that indicate a lack of the abilities, powers, or resources one needs if one is to be valued for one’s cooperation and reciprocity. Guilt stems from things that indicate insufficient motivation” (138). If shame and disdain do have their own rationale (circumstances which seem to warrant them), then this is not the same as the rationale for guilt and anger. We may judge that it makes sense to be angry at someone who purposely does a poor job at work, and that it makes sense for this person to feel guilty, but it would not follow that we should also feel disdain toward this person and that the person ought to feel ashamed about that which s/he could control. Perhaps someone could think that it makes sense to feel ashamed about doing a poor job after trying your best, but even if someone thought this, it would not follow that the person is “blameworthy” for the performance. I would suggest, too, that the rationales for guilt and shame may even be independent from one another and that they do not merely differ in terms of the agent’s motivation. While Karen’s norm above, for example, may imply that Peter ought to feel guilty about his abuse, it does not necessarily imply that she thinks that Ted ought to then feel ashamed about his abuse in extenuating circumstances; the emotions may even have quite independent norms governing them.

Again, Gibbard provides an evolutionary account on this point about guilt and anger. He admits that it might be possible that there is some other emotion than guilt which might serve its role (295-296), though he suggests that the feeling of guilt is a most likely candidate for such a role in our moral deliberation. His reasoning is that, while “it does not always reform a person…it does confine itself fairly well to the things he can be motivated to do or not to do. Other self-directed bad feelings – shame, fear, disgust, embarrassment, humiliation – are less discriminate” (297). A moral judgment will be made in terms of negative emotions which most motivate socially coordinating behavior. Shame and disdain, he notes, are closely related moral emotions to guilt.
and anger; nonetheless, the rationale for guilt and anger alone serves as a rationale for a finding of blameworthiness and thus this rationale motivates behavior within one’s direct control. But for having the emotions of guilt and anger in our repertoire, we would not (in the norm-expressivist’s view) be able to assess blameworthiness or make the narrowly moral judgments which so promote social coordination.

2.3.3 Moral Judgments Promote Social Coordination Through Meshing Guilt and Anger

Gibbard argues that evolutionary theory would suggest that we have a capacity for adopting moral norms (and, indeed, for experiencing their justifying emotions) because they tend to promote compliance that aids in social coordination. The human capacity to accept moral norms and experience the moral emotions allows us to coordinate behavior with others, and this ultimately benefits the individual by promoting inclusion in cooperative relationships (64-68). Gibbard states that “Propensities well coordinated with the propensities of others would have been fitness-enhancing, and so we may view a vast array of human propensities as coordinating devices. Our emotional propensities, I suggest, are largely the results of these selection pressures, and so are our normative capacities” (67). Gibbard proposes that we make moral judgments, and thus experience the motivated emotions and behaviors, because moral emotions gel with one another (67). The narrow emotions of guilt and anger particularly mesh in this way. As Gibbard explains, “guilt is coordinated with anger in a special way: it aims to placate anger [i.e., indignation], and it is governed by the same norms as govern anger” (139). Arnold may believe, for example, that he shouldn’t keep secrets from his wife Maria. If norm-expressivism is right, Arnold would then think it makes sense for Maria to be angry at him about his secret life, and thus he also thinks it makes sense to feel guilty. If he is truly “in the grip” of this norm, he would then display guilt over his secret life (69). Just as remorse tends to be relevant in criminal sentencing, displays of guilt tend to appease anger in private matters as well; anger and guilt gel to create pressures for honesty, since these are emotions to be avoided due to the misery they bring. They also gel to provide a means for reconciliation in that guilt can appease anger, at times much better than a mutual display of anger. Thus, it is suggested that evolution has left things this way. We have a natural capacity to experience the moral emotions of guilt and anger, and these emotions gel with one another in order to promote social coordination whenever the parties share moral judgments.
Of all the moral emotions, then, Gibbard finds that the moral judgment expresses one’s acceptance of norms governing guilt and anger. One’s acceptance of a moral norm is taken (in norm-expressivism) to entail one’s judgment that guilt and anger are rational responses to the conduct from an impartial and fully engaged standpoint. We will move on to discuss Gibbard’s understanding of this acceptance and rationality (based upon an impartial and fully engaged standpoint), but we should now understand Gibbard’s arguments (above) for identifying the terms of a moral judgment to be the rationality of anger and guilt in particular. The judgment that it is rational to feel angry at an agent and that it is rational for the agent to feel guilty is taken, by Gibbard, to provide necessary and sufficient standards for making the moral judgment because these emotions are believed to in fact carry the strongest motivational force for action compliance of any moral sentiment (even though other moral sentiments can help to shape them), their rationale is a judgment that acts are within the agent’s direct control, and these narrow emotions mesh with one another to promote social coordination.

2.3.4 The Nature of Moral Emotions: The Circularity Problem

Because we judge blameworthiness in terms of the rationality for anger and guilt, anger and guilt themselves cannot be, nor necessarily be responses to, judgments of blameworthiness. Gibbard casts moral judgments as a consideration of moral emotions, and if this is so then moral emotions cannot be moral judgments without making norm-expressivism circular. For example, if I perceive a deception and then feel angry, my anger cannot be the judgment that this deceit is a “moral failure” to be honest or that the mistruth is some kind of “wrongdoing”; if this was so, then my emotion itself would be a moral judgment, rather than (as Gibbard proposes) the moral judgment being about the emotion. Likewise, my perception of a mistruth could not somehow imply the judgment that the mistruth is “wrong,” for this too would make the norm-expressivist account of a moral claim circular. If anger and guilt either were, or were responses to, moral judgments, then the claim “It is wrong to do X” would express the claim that “It makes sense to judge that X is wrong.” Thus, Gibbard makes the point that anger and guilt cannot require us to think of situations in terms of blameworthiness. Rather, he suggests that circumstances are enough to cause these emotions through some other means.

Gibbard suggests that, in fact, emotions may respond very immediately without any thought as to the types of stimuli which elicit dif-
different emotions. An adolescent could feel ashamed of her/his shabby clothes, he notes, without deliberating about the way in which these clothes will impact one’s role in cooperative schemes; shabby clothes are enough to cause the shame (in some cases probably) because these clothes just may tend to have a real impact upon one’s standing in cooperative schemes (137). Shabby clothes may have come to elicit shame because of the way in which they might impact one’s place in cooperative schemes, but the emotion itself need not involve such thoughts. We may have learned early on the impact that attire can have upon the way others engage with one socially, and thus shabby clothes have become a rather immediate trigger for shame. Likewise, one can feel guilty about lying without “thinking” about the fact that now others are less inclined to trust one in cooperative schemes; feelings of guilt in response to lying would seem to be another very immediate, and yet learned, response in our society. Gibbard clarifies this point as follows:

none of this means that the person who feels guilty or ashamed thinks in any such terms. None of this means that I have specified the properties a circumstance must have if a person is to feel guilty over it or ashamed over it. Specific sorts of things will make a person feel guilty, and specific sorts of things will make him feel ashamed. What biological speculation suggests is a reason why these particular types of things should be lumped together when they are otherwise varied. (138-139)

It is important that moral emotions could work in such a way for the norm-expressivist. If my anger consisted in thinking of someone as “lacking” effort toward “proper” coordination, then this could arguably implicate a moral judgment (about what this person should have been doing) already. When people make moral judgments about specific conduct, they express the view that the relevant moral emotions make sense. Thus, anger and guilt can respond to things that indicate a failure in cooperative schemes, even though anger and guilt are not judgments that such failures exist.

2.3.5 Why This Matters

We can now derive some conditions for a theory of emotion compatible with norm-expressivism. The right theory of emotion for norm-expressivism must explain how the broadly and narrowly conceived moral emotions function in our lives, and it must do this in a way compatible with Gibbard’s characterizations above. Anger and guilt will need
to be distinct from one another, and they will need to be distinct from
shame and disdain and all other broadly moral emotions. Thus, we can
let Condition One for a compatible theory of emotion be:

C1: The theory of emotion must provide a framework capable of
distinguishing between guilt and anger, as well as shame and
disdain (and other broadly moral emotions).

Condition Two could be the more substantive claim that:

C2: The theory of emotion must provide a framework capable of
explaining how some emotions (such as anger and guilt in norm-
expressivism) tend to be stronger motivators of action than are
other moral emotions.

To prove that guilt and anger are, in fact, the most motivating moral
emotions is an empirical question that goes beyond a theory of emotion.
What a theory of emotion can do is provide a conceptual framework for
emotions so that, if anger and guilt carry the motivational force that
Gibbard suggests, then his theory will mesh with the nature of emotions
at large.

The theory of emotion would need to explain how the broadly
moral emotions could serve a secondary role to moral judgments by
shaping the norms for guilt and anger. In doing so, it is more credible to
then suppose that specific ones (like guilt and anger) are the specific
terms of a moral judgment. In order to avoid the circularity challenge,
though, we cannot distinguish guilt and anger in terms of moral judg-
ments; guilt and anger cannot be nor require judgments of blameworthi-
ness (or a lack of sufficient motivation). Even defining the moral em-
otions in terms of moral concepts would render norm-expressivism a cir-
cular meta-ethical account for the meaning of moral terms. Thus, let
Condition Three be:

C3: The theory of emotion cannot imply that anger and guilt are,
nor that they are necessarily responses to, moral judgments, nor
can moral emotions be defined in terms of moral concepts.

Moral emotions must be responses to a cluster or circumstances that
serve in some evolutionary capacity, and yet these emotions cannot be a
judgment specifically about such circumstances (due to their immediacy,
in order to avoid making the theory circular to norm-expressivism).
Thus, C3 implies that the theory of emotion must provide a framework
for explaining how anger and guilt respond to a cluster of circumstances without implying that these emotions are responses to a finding of rationale.

While it is an interesting question whether or not Gibbard has correctly identified the moral emotions, he admits that if this is not the case then other cultures may have their own moral emotions that serve to motivate social cooperation in similar ways. If, in fact, guilt and anger are not universal emotions, or if they are not even the most motivating ones in support of cooperative schemes in general, then we cannot use evolutionary arguments to support the view that they comprise the terms for (all) moral judgments. If other cultures do not experience guilt for example, and Gibbard admits that this may be questionable, then “we cannot interpret any of their talk as narrowly moral …. If they accept norms as governing the emotions they recognize and experience, they can then entertain questions that are somewhat like our own narrow moral questions” (150). Though we can note such possibilities, my project will focus upon a theory of emotion capable of explaining anger and guilt in a way compatible with norm-expressivism. We can note, though, that a desirable trait of a theory of emotion may be to explain all moral emotions in such a way that Gibbard’s secondary emotions (or even culturally distinct moral emotions) may function as the narrow moral emotions in other places. In doing so, norm-expressivism may offer an account of moral judgments that offers a more pluralistic account of the moral judgment in terms of culturally distinct moral emotions. If emotions other than guilt and anger function as do guilt and anger in Gibbard’s society, and if evolution did somehow of course allow for the development of culturally distinct moral emotions, then norm-expressivism may still advance meaningful moral dialogue for these societies (Gibbard, Wise 150). Should a theory of emotion turn out to offer an account of anger and guilt, only, in ways compatible with norm-expressivism, then norm-expressivism would offer at least a meta-ethical analysis for use in this society.

2.4 Accepting a Norm

As discussed, norm-expressivism maintains that a moral claim expresses one’s acceptance of norms that govern moral emotions. Because Gibbard’s theory is derived largely from his naturalism, this would imply that (if his theory’s account of the moral claim is correct) the “acceptance” of a norm must be a real state of mind. According to Gibbard, “Accepting a norm in whatever psychic state, if any, gives rise to [the]
syndrome of avowal of the norm and governance by it” (75). Should there not turn out to be any such psychic state, allowing for the acceptance of norms as he describes this syndrome, then his theory would need significant revision. Gibbard’s theory presumes, that is, that there is such a psychic state. He explains that, “The state of accepting a norm, in short, is identified by its place in a syndrome of tendencies toward action and avowal – a syndrome produced by the language-infused system of coordination peculiar to human beings” (75). To understand the acceptance of a norm, then, we need to understand avowal and governance. Because avowal, as we will see, requires that one is prepared to sincerely declare one’s support for a moral norm in discussion with others, an understanding of acceptance will benefit from an understanding of normative discussion as well. Thus, in order to understand what it is to accept a norm, we will need to understand what avowal, via normative discussion, and normative governance all are in Gibbard’s view. This understanding will prepare us to identify additional criteria for a theory of emotion compatible to norm-expressivism.

2.4.1 Avowal

Gibbard notes that, minimally, “to accept a norm is to be prepared to avow it in normative discussion” (73). If I avow a moral norm, such as that “one should feel guilty for killing innocent people, and others should be angry at this person,” then I would be prepared to declare this position in sincere discussion with others. Gibbard explains that avowal is not just a willingness to defend a position to others; we can, after all, insincerely defend a position. “Avowal,” as he uses the term, refers to our disposition to share one’s position when speaking freely with the highest degree of sincerity – that is “without the psychic complications of self-censorship” (74). The sincerity he refers to here is not the “deliberate” sort of sincerity found whenever one is consciously “holding oneself to standards of honest avowal” (74). Rather, avowal is said to require a type of sincerity more likened to “childlike openness” – prior to self-censorship (74). Gibbard suggests that the sincere avowal of a moral norm requires “a disposition to avow spontaneously” (74). If we cannot realize our own mental constraints (or self-censoring) at times, though, then this would imply that we could ultimately accept norms that we would avow only in ways which are perhaps quite unknowable to ourselves at times. This implies, I believe, that one’s awareness of the acceptance of a moral norm exists on a spectrum. That is, we may accept moral norms that we are more or less fully able to know that we accept.
One’s acceptance of a moral norm implies that we would avow it, that we would freely affirm it in response to demands for consistency within unconstrained normative discussion. Because it is difficult to know to what degree our own self-censoring is the cause of such constraint in discussion with others, I would find there to be epistemological difficulties in knowing when one avows a norm. Nonetheless, Gibbard proposes that the meaning of a moral claim should be taken to imply an avowal of this sort.

2.4.1.1 Normative Discussion

Gibbard states that the acceptance of a norm is “in part to be disposed to avow it in unconstrained normative discussion, as a result of the working of demands for consistency in the positions one takes in normative discussion” (74). The acceptance of a moral norm implies the sincere avowal of the moral norm in discussion with others in one’s community, though Gibbard takes such discussion to be both “actual and imaginary” (75). Though normative discussion does not require that such considerations are realized through actual conversation with others, normative discussion does imply a state of mind which takes the norm to be at least hypothetically defensible to others should such unconstrained conversations take place. The acceptance of a norm, as Gibbard says, stems from a “language-infused system of coordination peculiar to human beings” and normative discussion relies upon such an ability for dialogue (75). I must be prepared to sincerely avow the moral norms I accept in the midst of (but imagined) discussion with others. In doing this, I must be prepared to declare this norm to others in my community while allowing others to demand my consistency and while allowing for the possibility that others’ views may warrant influence upon the norms I accept. Normative discussion provides opportunities to “think together on absent situations,” and it thus assumes a linguistically infused deliberation with others (72). Thus, Gibbard explains that, in avowing a moral norm to others through even an imagined discussion, “One tends to be influenced by the avowals of others, and to be responsive to their demands for consistency” (75).

Again, Gibbard speculates on an evolutionary account of this characterization. Normative discussion is possible because one is both open to mutual influence between the discussants and one places a value upon consistency, and normative discussion is a necessarily implied element of the accepted moral norm because this is needed in order for moral claims to function in support of cooperative schemes. If I refuse to give others’ views any possible authority, or if I refuse to care about glaring inconsistencies, then this is not the making for a fruitful conver-
sation in terms of cooperative schemes. Gibbard argues that normative discussion functions to “coordinate acts and feelings” because it “tends toward consensus” (73). Thinking through things with others, we consider other positions (influencing one another with them at times), we check each other for consistency, and we move toward consensus. This interaction is explained, by Gibbard, in terms of evolutionary considerations. Both real and imagined normative discussion provides an opportunity for individuals to hash out, weigh out, and plan for the consequences of various norms. As Gibbard puts it, “shared evaluation is central to human life…because it serves biological functions of rehearsal and coordination” (72). Thus our (sincere) moral claims should be taken to imply such an avowal in response to demands of normative discussion.

Normative discussion is a natural way in which we promote consensus in our moral lives, thus promoting social coordination. In saying that a moral claim expresses our acceptance of a moral norm, and that the acceptance of a moral norm thus implies normative discussion, Gibbard suggests that our (sincere) moral claims do in fact imply our consideration of discourse with others in our community.

2.4.2 Normative Governance

A community’s consensus around a moral norm would help to support cooperative schemes only if the acceptance of moral norms makes a real difference upon how we do, in fact, feel and then act. If a group agreed that “one should feel guilty for stealing,” and yet this had no tendency to make people in fact feel guilty, then theft is in no way reduced by the moral judgment. Gibbard refers to the influence of one’s avowal upon what one does in fact feel and do as “normative governance” (72). Again, Gibbard provides a biological argument – now for normative governance. Normative governance is the mechanism which allows the avowal of moral norms to actually motivate our feelings and behavior. Gibbard explains the biological case for normative governance as follows:

Working out in community what to do, what to think, and how to feel in absent situations, if it has these biological functions, must presumably influence what we do, think, and feel when faced with like situations. I shall call this influence normative governance. It is in this governance of action, belief, and emotion that we might find a place for phenomena that constitute acceptance of norms. (72)
**Normative governance** then, refers to the tendency for avowed norms to actually motivate us to feel and act in accordance with our moral judgments.

Though normative governance inclines us to comply with the norms we avow, we ought not overstate this. As Gibbard notes, “normative governance … is like political governance. Governments do not always prevail: what a government commands, people do not invariably do. Normative governance is, in Gibbard’s view, a “mitigated” one, for we are often faced with competing norms (77). A youngster may accept that “it’s wrong to be unkind” and yet succumb to peer pressure and join in bullying a classmate; he may not even feel guilty about it at the time, but rather excited to be in the group. It may be that the youngster breaks his moral norm because he accepts another norm (such as that, “one should feel ashamed of refusing to back up one’s team”) that conflicts with his moral norm in this situation, or it may be that some other psychic mechanisms (such as a rather innate tendency, perhaps, to join the enemy in order to save oneself) will outweigh the normative control of his moral judgment in this situation. Gibbard maintains that, generally speaking though, one’s moral feelings will tend to be impacted by the moral norms we accept about them. As he says, “what a government commands has some influence on what people do, or it is no government” (77). If one accepts the norm that “one ought to feel guilty for bullying another and angry at those who bully,” then one will tend to feel these ways when fully aware of these circumstances (such as, perhaps, later in the principal’s office). Because the acceptance of the moral norm tends to promote actual feelings of guilt for bullying others, one will also tend to avoid bullying. Gibbard’s point here is that the acceptance of a moral norm implies, through normative governance, some motivation toward actually having the moral feelings in response to such circumstances and, indirectly, the related behavior as well in these circumstances – even though the tendencies are mitigated ones.

With regard to normative governance, we should perhaps recall that a moral norm prescribes emotions from a certain standpoint – one of full, impartial engagement. Gibbard explains the standpoint of a moral judgment as follows:

Moral norms tell how to feel only given a special standpoint – one it does not always make sense to take …. A person’s feelings will not often be impartial and fully engaged, but they do

---

7 Gibbard provides the example of the Milgram experiment, in which subjects of a study went along with torturing others in the study (or so they thought) at the directive of those in charge of the study. 1990, 58-61.
need to be coordinated to some degree with the feelings of others. This coordination may be fostered by a common judgment of how it would make sense to feel from a standpoint of full and impartial engagement. (127)

Gibbard provides an example about a stolen camel. While we may find that “it is wrong to steal (a camel),” one would not therefore find that it makes sense to experience the same degree of anger in all such circumstances. He explains that “If my camel is stolen, I will be outraged, and we may all agree it makes sense for me to be outraged. If, though, I were outraged at the theft of a stranger’s camel far away, that would be strange. It doesn’t make sense, you might tell me, to fume over a camel theft far away – there are so many worse things happening in the world” (126). Thus, Gibbard claims that “How fully to engage one’s feelings in a situation, then, is a separate question from how to feel if one’s feelings are fully engaged” (127). Moral judgments incline us toward guilt if we commit a moral offence, and they incline us toward anger at those who commit the transgression, but these judgments will not implore us to engage these emotions fully in actual circumstances in which we are little engaged (unless the offence is so great as to then warrant fully engaged emotions). We can note, then, that we accept norms for what it makes sense to feel from this moral standpoint, and that other norms govern their degrees of engagement.

In accepting a moral norm, I am actually motivated toward such feelings and action tendencies, even if this motivation is at times insufficient to cause these emotional experiences and even if the degree to which one actually engages these emotions is caused independently. The normative governance of emotions (and the actions inspired by them) is mitigated in these ways. Nonetheless, the acceptance of a norm entails that I am prepared to sincerely avow it in unconstrained normative discussion, and that I am motivated to actually experience anger and guilt (and the actions that follow from these) in response to the moral norms I accept.

2.4.3 Why This Matters
Gibbard’s understanding of the acceptance of a moral norm will impact the criteria for a theory of emotion. Through normative governance, the acceptance of a moral norm will motivate guilt and anger, and these emotions will in turn motivate action tendencies. Thus, we can let Condition Four be:
C4: The theory of emotion must explain how specific types of emotions (e.g. guilt and anger) motivate behavior.

Because Gibbard takes the avowal of a norm to assume a level of sincerity that surpasses any self-censorship, this means that one’s avowal (and thus acceptance) of a moral norm may depend upon a subconscious response to influences within one’s community. Let, then, Condition Five be:

C5: The theory of emotion must explain how emotions (e.g. anger and guilt) can be caused by the acceptance of norms (and the fact that one has accepted norms) that we are more or less aware of at the time.

We can also now recall (per Section 1.3.1) Gibbard’s argument that broadly moral emotions help to shape the norms for anger and guilt. Because normative governance implies that the acceptance of norms does lead to the emotions that the norms endorse, this means that the secondary emotions do in fact contribute to causing anger and guilt. Thus, we can let Condition Six be:

C6: The theory of emotion must provide a framework that explains how broadly moral sentiments shape guilt and anger.

2.5 Types of Moral Norms

Though we often have incomplete sets of norms, which can explain our ambivalence or uncertainty on moral dilemmas, Gibbard argues that individuals each have a system of norms which facilitates them making moral judgments whenever they make them. Even though moral judgments express an acceptance of specific norms governing anger and guilt in specific circumstances, these attitudes are taken to be responsive to higher order norms. Thus, we can speak of the higher and lower order norms together as comprising a “system of norms” (87). We can distinguish between higher and lower order norms by both what they govern and what shapes them. In this section, I will explain the types of norms proposed by Gibbard. This will not result in conditions for a theory of emotion itself, though it will provide for an understanding of the condition to follow in the next section.

Higher order and lower order norms are about different things; that is, they have different objects. Lower order norms direct our actions, beliefs, and feelings in response to facts. A lower order norm could be,
for example, “One ought to feel guilty for taking what belongs to another without the owner’s consent, and one ought to be angry at those who do this.” Whether someone takes what belongs to another without this person’s consent is an issue of fact. The norm directs our feelings in response to this fact – telling us to feel guilty for engaging in such behavior and telling us to feel angry at others who engage in such behavior. Moral norms are lower order norms because they specify specific circumstances for which guilt and anger are prescribed. Higher order norms are taken to “govern the acceptance of other norms” (168). Thus lower order norms are shaped by higher order norms. Gibbard’s point is not that we literally think about, or are conscious of, higher order norms. His point seems to be that higher order norms must be part of our mental architecture, and that lower order norms do thus represent (or logically imply) them.

Gibbard explains that higher order norms fall into “norms of warrant” and “norms of rationale”; norms of warrant “say how to recognize a good judgment,” whereas norms of rationale state “a rationale for the norms we should accept” (213). Norms of warrant direct us on the formal coherence of lower order norms as well as other epistemological parameters by which we judge the warrant of a lower norm (282). For example, a norm of warrant may provide epistemological criteria for choosing moral norms by telling us to choose lower order norms “only when one has considered it in a calm state of mind, and thought vividly about everything that might move him toward accept it or not” (168). On this account, it would seem to be plausible that a norm of warrant could even be to “choose moral norms only when in a benevolent state of mind.” Norms of rationale prescribe values which can justify a lower order norm; norms of rationale provide the standards for the types of reasons that count in support of a moral norm. Gibbard provides the example of a norm of rationale being: “Accept a norm if its acceptance, in one’s community, would most enhance a sense of meaning in life” (213). Likewise, I would think that a lower order norm of rationale could be “Accept a norm if its acceptance, in one’s community, would most promote feelings of mutual benevolence.” In place of benevolence, of course, we could substitute sympathy or a sense of justice or any other broadly moral sentiment that turned out, in fact, to function in this way within one’s moral appraisals. Thus, because higher order norms may stipulate rationale and warrant in terms of such broadly moral sentiments, it is possible that higher order norms provide ways in which the broadly moral sentiments help shape the narrow moral emotions of guilt and anger.
Now, let us understand what *causes* higher and lower norms. Lower order norms are, as we know, regulated by higher order norms which provide for their rationale and warrant. Gibbard discusses the “two prongs” of moral inquiry, explaining that (as we see above) “we start with tentative norms and work from there” (282). The norms that we start with, the norms that help shape other norms, are higher order norms. The two “prongs” which ground these higher order norms are said to be “loosely intuitionistic” and “pragmatic” considerations (282-284). He explains that “we might call loosely intuitionistic [the attempt to] think through examples, discuss together, confront puzzles and inconsistencies” (284). Pragmatic considerations would respond to “thought about gains and losses” (284). While higher order norms need not be consciously deliberated, our mental architecture provides for such representations in shaping our lower order norms. That is, higher order norms represent a consideration of logical coherence (intuition) and real benefits for human life. Because a beneficial belief can sometimes be inconsistent with other beliefs, these two prongs will not always promote the same higher order norms. (Thus, they do not always promote the same lower norms resulting from them). I may, for example, find that it is beneficial to hold the belief that one should prioritize the interests of their own immediate family (as a pragmatic consideration), and yet also believe that justice calls for equal rights for all (as a matter of logical coherence). Such higher order norms tend toward consistency. Gibbard explains, that the two prongs are “welded together by consistency” (285). Still, we can imagine how, if one is in a position of political power to benefit one’s own family members for example, that these prongs could lead to conflicting norms (wherein one considers giving unequal benefits to one’s own family). This consistency, as we will see though, is a rather general one. Thus, Gibbard suggests that moral judgments are shaped by one’s higher and lower order norms.

The suggestion that moral norms and emotions are judged for their rationality (in terms of warrant, plausibility and a general consistency in some manner) seems quite fruitful as an ethical analysis in the main. Higher and lower order norms appear, generally, to account for a quite realistic way in which we naturally come to endorse beliefs, actions, and emotions. This is not to say that reasoned normative systems are the only causes of such, but such a system would seem to be quite capable of influence. It appears to be quite cogent, indeed, to infer that what all moral norms have in common is the implication that a specific emotional response is taken to be warranted whenever the norms are violated – and that a system of norms lies behind such judgments.
2.5.1 Why This Matters

A moral judgment is the judgment that it is rational to feel guilty for doing something and that it makes sense to be angry at those who do this. In order to understand the sense in which it is judged to be rational, we need an understanding of higher and lower order norms. An understanding of the relationship between higher order and lower norms can also help us to anticipate ways in which broadly moral emotions could help (though need not) shape norms for anger and guilt through the implication of these broadly moral emotions in higher order norms.

2.6 What it Means to Think Something “Rational”

When one makes a moral claim, one expresses one’s view that the moral norm is rational and, in doing so, one has expressed one’s view that it is rational to feel guilty and angry in those circumstances. If I say that “theft is wrong,” I express my acceptance of the norm that “one ought to feel angry at others who steal, and one ought to feel guilty for stealing”; I think it is rational to accept this norm. In accepting the norm, I think it is rational to feel angry and guilty in such prescribed ways. Here, we will focus on what it means to find the norm, or feelings, to be “rational.” As Gibbard says, “moral norms in particular are norms for the rationality of guilt and resentment …. An observer thinks an act blameworthy, or morally reprehensible, if and only if he thinks it is rational for the agent to feel guilty over the act, and for others to resent the agent for it” (47). Recall, again, that as a meta-ethical theory, Gibbard is aiming to describe the meaning of moral language. Norm-expressivism does not aim to determine what does (or what should) make something rational, but rather what it is to think one’s claim rational. As Gibbard says, “My own analysis … is not directly a hypothesis about what it is for something to be rational at all. It is a hypothesis about what it is to think or believe something rational, to regard it as rational, to consider it rational” (46). Thus, the acceptance of a moral norm implies that one thinks it is rational to have these emotional responses to the circumstances.

Because of normative governance, our actual experiences of anger and guilt would be caused by the finding that it is rational to have these emotional responses to the circumstances. We judge a moral norm to be rational, and thus the feelings it governs to be rational, whenever we would endorse the norm in response to three basic criteria.
2.6.1 Rationality as Endorsement

To think that a norm (or even a specific act, feeling, or belief) is rational, in Gibbard’s view, is to “endorse” it based upon one’s own deliberation (6). When one deliberates over the thing to do (or the thing to believe or feel), one need not reflect upon specific sorts of reasons as we have seen. Whatever reasons are found to warrant the thing to do (or believe, or feel), those will be received as the grounding reasons for the person. Gibbard notes that, in defining “rational” in terms of such an endorsement or recommendation, he is not offering a traditional meaning for the term “rational.” Gibbard explains his view of rationality as follows:

Other phrases [than “the rational thing to do”] may capture this notion better. I have freely substituted talk of what it “makes sense” to do, to think, and to feel about things; that might be my best canonical phrase. Alternatively, we might talk of what one “ought” to do, think, or feel, and explain that the “ought” is not the moral one. With feelings and beliefs, we can talk of what states of mind are “warranted,” “well grounded”, or “apt” …. We might talk simply of …“the thing to feel” about something. If a flavorless recommendation on balance can be found in any of these terms then this is what ‘rational’ shall mean in this book. (49)

When one decides upon what it makes sense to feel in response to circumstances, one has decided upon what it is rational to feel. A moral judgment is the judgment that anger makes sense in response to the circumstances from the position of full, impartial engagement, and it is the judgment that it makes sense to feel guilty in response to the circumstances from the position of full engagement. Though rationality, in Gibbard’s sense, may be a rather loose concept, there are criteria that apply to any judgment about the rationality of a norm. In finding a moral norm to be rational, our judgment is that the norm is endorsable in that it satisfies these criteria. Because of normative governance, this means that anger and guilt are subject to the influence of such criteria.

2.6.2 Criteria for Rationality: Coherence, Plausibility, and Grounding Reasons

Gibbard argues that we, in fact, regard a moral norm to be rational whenever we endorse it in response to three criteria: we find the norm to be coherent with our other held beliefs and norms, we find the norm to be plausible, and we find the norm to be warranted by grounding reasons.
2.6.2.1 Coherence

The judgment that it is rational to accept a norm implies one’s judgment that the norm is coherent with “one’s beliefs and normative judgments” (157). This is a minimal criterion for rationality for, as Gibbard explains, “coherence is simply consistency, whether one’s judgments be well- or ill-founded” (157). A moral norm is judged to be rational only when it is consistent with one’s other accepted norms and relevant beliefs. Clearly, for example, the moral norm that “one ought to be angry at abortion providers” is inconsistent with the norm that “one ought never feel angry.” Beliefs can also logically imply inconsistencies in one’s normative claims. If, for example, one believes that the Plan B pill (i.e. the Morning After pill) is an abortifacient, then it would be inconsistent to accept that “one should be angry at abortion providers” and yet that “one should feel angry at those who oppose the Plan B pill.” As Gibbard explains it, “Coherence … is a matter of formal, internal consistency in one’s beliefs and normative judgments. An ideally coherent person could accept the logical consequences of everything he accepts without falling into logical contradiction” (157). Of course, people are not ideally coherent – nor (as Gibbard admits) does biological evidence suggest that we strive to be ideally coherent. Nonetheless, we find that (in Gibbard’s view) a general coherence matters in any judgment about rationality. That is, when we judge a norm to be rational (when we endorse it, or find that it makes sense), we find the norm to be generally consistent with our other norms and beliefs. If we take a moral norm to be rational, then we take it to be generally consistent with other moral norms and our beliefs about the facts.

The judgment that a moral norm is coherent with one’s system of norms will imply that it is consistent with one’s higher order norms as well as one’s other moral (lower order) norms and beliefs. As Gibbard says, “If a person is fully coherent, then he accepts something as an objective matter of rationality only if the higher order norms he accepts ascribe to it a standpoint-independent validity” (193). If, for example, I accept a higher order norm of rationale that says “One ought to choose moral norms that promote the improvement of all persons,” then it would be inconsistent to have a moral norm that says, “One ought not provide rehabilitation services to criminals” (or, in other words, that “one should

8 This is an issue of heated debate in medical ethics. The curious reader may contrast the position put forth by Pharmacists for Life with other empirical studies that refute the claim that such methods can terminate a pregnancy after conception (Glacier). The American Pharmacist Association’s professional Code of Ethics includes a controversial “Conscience Clause” (2004) that reflects such a debate as well. Glaiser, A. “Emergency Postcoital Contraception.” New England Journal of Medicine 337 (1997): 1058-1064.
feel guilty for providing rehabilitation services to criminals, and one ought to feel angry at those who provide such services”). Likewise, if I accept a higher order norm of warrant that states “One should choose moral norms from behind a veil of ignorance,” then it would be inconsistent to also accept a higher order norm for rationale that says “One should choose norms for punishment that vindicate victims”; in promoting the vindication of victims, one may accept norms which (at times) conflict with norms for punishment that would be warranted from behind a veil of ignorance. (If one did not know if one was the victim or perpetrator, one might endorse a punishment for a crime which is contrary to one endorsed from the standpoint of a victim only. Thus, these higher order norms have logical consequences that contradict one another.) When we accept a moral norm, thus finding a moral judgment to be rational, we must find the moral norm to be consistent with other beliefs and our entire normative system.

Here again, Gibbard speculates that evolutionary theory would support his claim that rationality implies a finding of coherence. Considerations of coherence promote normative discussion, for even hypothetically conceived discussions can pressure us to explore the consistency of our commitments. As Gibbard sees it, “These pressures [from my community for consistency] are part of a psychosocial mechanism by which discussion tends toward consensus” (230). In meeting the demands for consistency in normative discussion, we weed out inconsistent norms and beliefs, narrowing those that are worth considering with others. This social coordination offers a mutual benefit, and thus promotes our own fitness. More importantly, though, Gibbard explains that “Inconsistency lays us open to a special kind of self-frustration” (289) – one in which we cannot identify clear commitments needed to guide action. In checking our norms for consistency, we better pin down specific commitments to the exclusion of others; through this, our actions are better determined. The “special value” of consistency as he sees it is that it promotes action-guidingness; with greater consistency, I need not be paralyzed in my decision making by conflicting norms (289). Thus, the demand for coherence promotes consensus and action guidingness; for these reasons, Gibbard suggests that coherence has become a criterion for endorsing a norm (thus judging it to be rational).

Nonetheless, the consistency of norms is a qualified one. It is not an absolute consistency, but a rough consistency. In keeping with Gibbard’s naturalistic argument, we can note that issues are sometimes better solved through a consensus that does overlook inconsistencies in one another. We may have a mix of people with different normative systems (such as some reflecting more of a natural law theory and others
reflecting more of a utilitarian analysis of the issue at hand), but these people may be able to agree upon their moral commitments on certain topics (such as the view that murder is wrong) even without trying to reconcile one another’s inconsistencies. The fact that we can all oppose murder, when living alongside one another for example, is more of a pressing issue than is each person’s internal consistency with respect to their broader normative systems. If we delay accepting norms against murder, for example, because we can identify some inconsistent norms within one another’s normative systems (and we decide to push for reconciling these inconsistencies prior to reaching an agreement about murder), then time goes by without any real rejection of murder in our society. Likewise, a thorough consistency can drive away consensus; if a nihilist is thoroughly consistent, then he may not be inclined toward consensus with others at all and, thus, inconsistency on his part would better support his self-interests in yielding the benefits that an inclusion in society may offer. Thus, Gibbard explains the qualified value of consistency as follows:

Even if consensus were all that mattered, responsiveness to demands for consistency would not always promote it. If some are especially consistent, that might drive them away from consensus with the rest. Jointly inconsistent norms may each have a strong natural appeal. When that is so, there may be more consensus to be gained by sweeping inconsistencies under the rug than by exposing them and trying to root them out. (288)

While consistency promotes consensus and action-guidingness, inconsistency can also be fitness-enhancing at times. Gibbard notes that the judgment that X is rational only implies the finding that the norm is on the whole consistent with one’s other beliefs and norms (both higher and lower). As he says, “There is a special way in which inconsistency matters, both with norms for action and with norms for feelings …. it will not justify seeking consistency at all costs” (291). In noting this rough consistency, Gibbard recognizes that we at times have norms that are not consistent with one another – times in which (perhaps if these inconsistencies have little practical conflict in our lives and provide good individual guidance to us) we are willing to sweep the inconsistencies under the rug. Thus, the judgment that it is rational to feel angry or guilty does not imply that one takes the feeling norms to perfectly consistent with all other norms we accept for anger and guilt. It just implies that we take the norm to be consistent on the whole. We may even, as noted, have higher order norms that stipulate a stronger or weaker demand for consistency,
and perhaps we could even have higher order norms that stipulate certain (and more flexible) terms for the consistency of norms governing emotions. The norm-expressivist will require that moral norms (governing anger and guilt, for example) are accepted only when they are taken to be consistent, but this consistency can be a flexibly defined one. It would seem, then, that anger and guilt are caused (in part at least) by the acceptance of norms that are by and large consistent.

2.6.2.2 Plausibility
Gibbard stipulates that norms are regarded as rational only when they appear plausible as well. Gibbard finds that “It might turn out that we simply find implausible the ways of thinking it would serve us best to find plausible, and that nothing can change our minds” (227). That is, it may be possible to find a norm, or even a belief of course, quite sensible in many ways and yet nonetheless find that “the view is still crazy” (227). In order to ultimately endorse a norm, finding it to be rational in his sense, we must find the norm to be plausible.

Gibbard suggests that nature itself has likely provided us with a sense of plausibility, which can guide our positions even in the face of other evidence. For example, even if there were logical evidence that it is both possible and ultimately beneficial for parents to love all children as much as their own, the idea (to me) is just implausible. (While it is impossible to know what I would think in this bizarre situation, where such evidence exists, favoritism for one’s own children seems to be driven by natural instincts rather than evidence.) It just seems crazy to suggest that I, for one, should try to do this. Our capacity to reject implausible judgments offers the advantage of promoting stability by rejecting wild ideas. Through this capacity, we can reject wild ideas even when they are proposed by a fabulously prepared proponent. Since we do in fact sometimes just run up against those who are more articulate than oneself, it is probably helpful to sometimes hold fast to one’s own gut reaction in spite of the debate. Gibbard is not suggesting that one’s gut reaction, or (as I take it) an assessment of plausibility, is sufficient to cause endorsement; we are certainly swayed by reasons as well, in addition to such considerations of plausibility. A finding of plausibility, though, is a necessary condition for the judgment that something is rational. Typically, of course, we find well supported claims to be quite plausible. In keeping with his evolutionary account of the moral judgment, Gibbard notes that “if advantage and plausibility … coincide fairly well … that can be no evolutionary surprise” (229).
2.6.2.3 Grounding Reasons

The third condition for the rationality of something (or the judgment that something makes sense, anyhow) is that there must be grounding reasons for it. Gibbard argues that moral claims are believed (by the one who makes the claim) to be rational only when they are supported by the “preponderance of available grounding reasons” (162). He argues that, whenever we judge something to be rational (whether a norm, a belief, or an emotion), we endorse it based upon our perception of its “warrant” (37). People, at large, take their moral judgments to be warranted by broad reasons. The type of reason that motivates our endorsement of a moral norm (whether deontological, utilitarian, etc.) has no bearing upon whether the moral claim expresses a judgment that it is rational to endorse the norm. What is important is that one finds reasons that (in one’s own view) call for, or recommend, the norm. In keeping with his observations about our moral nature, Gibbard finds grounding reasons to be broad in kind. Our reasons for such judgments about the rationality of a norm may derive from rationales, for example, that involve faith, pragmatic consequences, strong feeling, and/or even spontaneity. Gibbard notes, for example, that in some realms, “the thing to do is what faith picks out, and not what reasoning would pick out – even the best of reasoning” (50). What he means here is that “the thing to do” actually just is “the rational thing to do” (50). If I decide that faith knows best, then (according to norm-expressivism) I regard the rational thing to do as being whatever it is that faith recommends. If one believes that faith knows best, then the dictates of faith are the thing most warranted. Faith is, then, my grounding reason.

Grounding reasons are explored through normative discussion. As we know, the acceptance of a moral norm requires normative discussion. Through normative discussion, one judges the moral norm to be rational in light of one’s own reasons which are subject to the influence of others in discussion and also subject to demands for (general) consistency. Nonetheless, we do not merely adopt the norms for guilt and anger that our culture adopts by consensus; we also think independently about what one ought to feel based upon our own reasons. Through normative discussion, we consider competing reasons (being open to mutual influence) and we are responsive to the demands for consistency. The most rational norm (in one’s own view), though, is the one ultimately endorsed in the end and its grounding reason just is whatever reason actually warranted our endorsement of it (162, 50).

In emphasizing the need for grounding reasons, Gibbard contrasts the endorsement of a norm found to be rational to one that is merely found to be pragmatic. When we judge a norm to be rational, we do
not mean that it would be merely prudent to endorse it. We are rather saying that there are reasons that actually warrant endorsing the norm; it would make sense to believe that the norm is correct. Some of those reasons may be pragmatic (and they may indirectly even all be of some pragmatic value in evolutionary terms), but my judgment that it makes sense for others to feel angry and for one who does X to feel guilty is not a judgment that “it would be useful for people to feel in such ways” (222). If I consider it rational to endorse a norm (such as “one ought not lie”), I am not thinking that this norm makes sense because of the gains it may offer. We do not only endorse what we find to be pragmatic. If I am a deontologist, for example, then I may decide that it makes sense to be honest, in spite of my awareness that nothing but trouble will result. We often regard acts and emotions to be warranted, even when they are not of a direct pragmatic value, even though (as Gibbard notes) there is ironically an “indirect pragmatism” to all of this (222). Again, Gibbard provides an evolutionary argument for why we should interpret rationality in this way: It is a good thing that norms for feelings and conduct are often driven by considerations of warrant in contrast to a direct appeal to pragmatic considerations. As Gibbard explains it, “To think “Grieve only when it does you good” would drain grief of significance…. There are gains to be had from accepting non-pragmatic norms” (222). If one feels guilty only when one finds that it would be useful to feel guilty, then the sincerity of guilt may tend to be quite questionable; guilt appeases anger because it indicates regret, not strategic analysis. Though evolution has made it such that judgments about the rationality of an emotional response tend to have pragmatic benefits, these judgments about the rationality of anger and guilt are not best understood (in Gibbard’s view) as claims about pragmatic value. The norms we endorse may find their drive, indirectly, in some sort of biological disposition to pursue pragmatic benefits, although the judgment about the emotions is not that they are useful but that they are indeed warranted by grounding reasons.

2.6.3 Why This Matters

A moral claim expresses a state of mind that thinks it is rational (that it makes sense) to feel anger and guilt in response to specific circumstances. Because of normative governance, as discussed above, this judgment would need to then actually cause such emotions. The judgment could not be a necessary condition for the emotion (for, as we have seen above, this would be problematic to norm-expressivism due to the circularity concern), but the norms must be among the possible causes for the emotion. We now know that norm-expressivism takes a rational judgment to
imply a finding of plausibility, a finding of grounding reasons, and a finding that the moral norm is generally consistent with one’s other norms (including higher norms, which may give place to other broadly moral sentiments). When we find a norm which governs an emotion (like guilt and anger) to be “consistent,” we need not judge the consistency in the same way in which we judge the consistency of ordinary beliefs; higher order norms may allow for more flexible terms when determining the logical structure of norms, or reasons, for emotion. Nonetheless, if norm-expressivism is correct, then emotions would need to respond to judgments that they are rational, and these judgments would need to in fact represent a rough coherence with one’s entire normative system. Their consistency would imply that, with respect to one’s system of norms, the norms do not (on the whole) rule one another out. This, too, would have to be considered.

Furthermore, we can recall (from Section 2.4) that the acceptance of a norm is taken to be an actual state of mind. In this state of mind, one’s avowals govern anger and guilt even if they do so in a mitigated way. These avowals will respond to cultural considerations as well as demands for consistency in normative discussion. That is, the moral emotions are in fact caused (if but in part) by our preparedness to respond in a language-infused way to demands for consistency and cultural influences. Because normative discussion entails a language-infused consideration of cultural perspective and demands for consistency, normative discussion would then require higher level thinking. That is, normative discussion requires thought processes that separate us from other species. Thus, let Condition Seven be:

C7: The theory of emotion must explain how guilt and anger respond to considerations regarding their consistency with other norms, their grounding reasons and their plausibility. (That is, anger and guilt can be caused by judgments that are derived from language-infused thought processes in consideration of logical consistency and cultural influence.)

2.7 Conclusion
Understanding, now, the central tenets of norm-expressivism has allowed us to identify seven conditions for a compatible theory of emotion. Gibbard explicitly states that the right theory of emotion for norm-expressivism will need to do two things: 1) “explain particular emotions without invoking normative concepts like being to blame,” and 2) “allow for normative governance of emotions” (148). Above, I have fleshed out features under both of these points, so that our conditions for a compati-
ble theory of emotion are even more specific. Conditions 1-2 and 6, below, identify the need to explain particular moral emotions, and the way in which broad and narrow moral emotions may interact according to Gibbard’s proposal; without doing so, the argument that a moral claim expresses one’s acceptance of norms governing anger and guilt, in particular, looses credibility. Condition 3 addresses his circularity challenge (also mentioned here by him in statement number one above). Conditions 4, 5 and 7 are all implied through the normative governance of moral emotions; normative governance entails that the (even unconscious) judgment that it is rational to experience anger and guilt will actually cause one to experience these emotions as well as their corresponding behaviors. As we move onto survey theories of emotion, we will consider these seven conditions:

<table>
<thead>
<tr>
<th>Seven Conditions for a Theory of Emotion Compatible With Norm-Expressivism</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The theory of emotion must provide a framework capable of distinguishing between guilt and anger, as well as shame and disdain (and other broadly moral emotions).</td>
</tr>
<tr>
<td>2. The theory of emotion must provide a framework capable of explaining how some emotions (such as anger and guilt in norm-expressivism) tend to be stronger motivators of action than are other moral emotions.</td>
</tr>
<tr>
<td>3. The theory of emotion cannot imply that anger and guilt are, nor that they are necessarily responses to, moral judgments, nor can moral emotions be defined in terms of moral concepts. (The theory of emotion must provide a framework for explaining how anger and guilt respond to a cluster of circumstances without implying that these emotions are responses to a finding of rationale.)</td>
</tr>
<tr>
<td>4. The theory of emotion must explain how specific types of emotions (e.g. guilt and anger) motivate actions.</td>
</tr>
</tbody>
</table>
5. The theory of emotion must explain how emotions (e.g. anger and guilt) can be caused by the acceptance of norms (and the fact that one has accepted norms) that we are more or less aware of at the time.

6. The theory of emotion must provide a framework that explains how broadly moral sentiments shape guilt and anger.

7. The theory of emotion must explain how guilt and anger respond to considerations regarding their consistency with other norms, their grounding reasons, and their plausibility. (That is, anger and guilt can be caused by judgments that are derived from language-infused thought processes in consideration of logical consistency and cultural influence.)

We will now move on to explore the compatibility of major theories of emotion today in light of these conditions. The theories will be broken into three camps (and thus chapters) – behaviorist, cognitivist, and somatic theories of emotion. We will consider each of these theories in light of the conditions above, arriving at the most compatible theory of emotion to support the psychology implied by norm-expressivism.
3. Behaviorist Theories of Emotion

3.1 Introduction

It is because emotions cash out in action that, in Gibbard’s view, they promote social coordination in our moral lives. As we move on to survey theories of emotion for norm-expressivism, we will find different accounts of the emotions’ action-guidingness. It seems most obvious, perhaps, to then consider behaviorist theories of emotion. Perhaps the explanation for the action-guidingness of emotions could be as direct as that emotions just are these actions (or dispositions toward these actions). Though there are different varieties of behaviorist theories, they share the claim that an emotion is correctly defined in terms of behaviors and that an emotion is properly identified with the behaviors (or behavioral dispositions) that it manifests. In this chapter, we will consider how both philosophical and psychological behaviorist theories of emotion address the conditions for norm-expressivism.

It can be argued that philosophical behaviorism and psychological behaviorism are unrelated, in that philosophical behaviorism proposes a meaning for emotion words based upon a logical analysis of the way in which words are used and psychological behaviorism offers a meaning and explanation of emotions based upon empirical observation. Both theories, though, reject attempts to identify or study emotion as mental phenomena and both promote an understanding of emotion in terms of behaviors (or at least, in the case of Ryle’s philosophical behaviorism, in terms of dispositions toward behaviors). Since both theories offer accounts of emotion in terms of behavior (or tendencies toward it), we should consider whether they would meet the needs of norm-expressivism in accounting for normative governance and, particularly, the action-guidingness of moral emotions. Ultimately, these theories are compatible with norm-expressivism only if they meet all seven conditions for an acceptable theory of emotion.

In this chapter, I will consider Gilbert Ryle’s philosophical behaviorism and then two forms of psychological behaviorism (radical
behaviorism and paradigmatic behaviorism). In the end, I find that although the more recent paradigmatic behaviorist theory best meets the norm-expressivist’s conditions (per Chapter Two), it falls short by itself of providing a yet definitive account of emotion types. Thus, while its definition and account of the moral emotions may be consistent with norm-expressivism, the norm-expressivist should continue seeking more specified theories.

3.2 Philosophical Behaviorism: Gilbert Ryle

Ryle’s behaviorist theory of emotion derives from his rejection of Cartesian dualism. His theory of emotion stems from his critique of what he calls “the dogma of the Ghost in the Machine” (15-16). In disputing metaphysical dualism, Ryle denies that there is any immaterial thinking “thing” that we ought to conceive of as mind; there is no immaterial mind at all (19). In keeping with this, thoughts and emotions are not to be explained as immaterial substances or processes. Rather, he argues that an analysis of language (and mental terms in particular) will provide a behaviorist theory of mind. He argues that the meanings of emotion words are explainable in terms of dispositions toward behaviors. As explained in the Stanford Encyclopedia of Philosophy, Ryle’s philosophical behaviorism (also regarded as “analytic behaviorism”) is a theory “about the meaning or semantics of mental terms or concepts. It says that the very idea of a mental state or condition is the idea of behavioral tendencies, evident in how a person behaves in one situation rather than another” (Graham). Just as Gibbard defends a conceptual analysis of normative terms in light of indirect empirical data (given his naturalism), Ryle also engages in a conceptual analysis of emotion terms that is grounded in his understanding of the way in which emotion concepts are actually employed in our lives.

In this section, I will explain Ryle’s theory of mind as it pertains to his behaviorist theory of emotion, a theory that is then detailed in his particular taxonomy of emotions. It will be argued that his taxonomy of emotion terms and his general premise (that emotions are dispositions toward behaviors) fail to meet the norm-expressivist’s conditions for a theory of emotion.
3.2.1 Ryle’s Behaviorist Theory of Mind Leads to a Theory of Emotions

Though Ryle refutes the immaterial mind, he provides a behaviorist theory of meaning for terms used to designate “mental” activities such as “thinking”; frames of mind such as the “temporary moods”; and other more enduring mental dispositions such as “motives” as he calls them (34, 95, 84). Ryle purports that the meaning of such concepts can be explained through an account of the behaviors and dispositions toward such behaviors which we reference. Thinking, for example, is explained as the act of talking to oneself; when one stews over an offence, for example, one is talking to oneself (though perhaps silently, without moving one’s lips). While one’s silent thoughts may, in Ryle’s view, be explained as acts of talking to oneself through “auditory word-images, instead of spoken words,” other aspects of the mind (such as moods and motives) are not acts but rather “dispositions” to act in certain ways in response to specific circumstances (35, 85). Because dispositions themselves are not acts or events (but rather tendencies toward acts), Ryle argues that we should not take them to be a proper cause of behavior (86-87).

Ryle explains that, “The traditional theory of the mind has misconstrued the type-distinction between disposition and exercise into its mythical bifurcation of unwitnessable mental causes and their witnessable physical effects” (33). He argues that there are not unwitnessable mental acts; even thoughts are witnessed (as self-talk) by those having them in his view. We ought not confuse, though, mental acts for dispositions. While mental acts are witnessable, dispositions are neither witnessable nor unwitnessable. Dispositions are not acts or substances at all. Thus, while mental acts can be causes of other acts, dispositions are explanatory and, yet, not causal accounts of behavior. Though dispositions are not witnessable or unwitnessable substances, they can be evidenced in observable ways; we can tell, for example, that someone has a vain disposition because we see that s/he (fairly regularly) acts in such ways (89-90). Emotional dispositions are tendencies toward behavior (and this behavior may be a disjunctive set of behaviors) in response to certain stimuli. They are not, in Ryle’s view, either witnessable nor unwitnessable mental causes of behavior.

Still, while dispositions do not cause behavior, Ryle maintains that they do explain actions in a different sense. In his own example, Ryle explains that a disposition toward particular actions in response to certain circumstances is likened to the brittle constitution of a glass, which explains why the glass breaks upon being hit by a stone, but even this disposition itself is not an act or proper cause (88-89). The glass’s
brittleness is not an occurrence or an act, and thus it is not the cause of the breaking glass; the cause would be an act or occurrence, such as an object striking the glass (89). Ryle argues that the brittleness of the glass just is a law-like disposition to behave in certain ways in response to certain circumstances; its brittleness just is the tendency to break upon being struck (89). Likewise, he explains that if a person is motivated by “vanity,” one will tend to boast and be most disappointed at a lack of others’ acknowledgement *whenever* such opportunities (to boast) and circumstances (of others’ failure to acknowledge one’s importance) occur (89-90). The disposition to act just is the tendency to act in at least one of several loosely associated ways; if one feels guilty, for example, one may tend toward apology or avoidance (or both). Ryle proposes that certain types of emotion words are used to denote tendencies toward certain behaviors in this way, and that such emotions explain their behaviors by way of denoting these dispositions to act in such ways whenever such circumstances are present.

Within this context, Ryle argues that “the word ‘emotion’ is used to designate at least three or four different kinds of things,” which he calls “‘inclinations’ (or ‘motives’), ‘moods’, “agitations” (or ‘commotions’) and ‘feelings’” (83). As we will see, Ryle later explains that agitations are a subset of moods (93). Through his analysis of language, Ryle proposes that each of these emotion words has a distinct behaviorist meaning. We can now consider whether (any of) these emotion words could denote emotions in terms compatible with norm-expressivism.

### 3.2.2 Ryle and Feelings

Ryle notes that we often speak of emotions as if they were feelings. Though emotion words can be taken to mean “feelings,” Ryle argues that this meaning of feelings words implies no explanation for behavior. By feelings, Ryle means “the sorts of things which people often describe as thrills, twinges, pangs, throbs, wrenches, itches, prickings” and sort; we do, of course, sometimes describe our emotions in such terms as when we claim to experience “a throb of compassion” (84). He argues that, though we speak of emotions in such ways and even though there may be some interaction between feelings, motives and moods, emotions do not universally entail feelings. Feelings, in Ryle’s view then, are neither necessary nor sufficient for moods or motives. Ryle explains that “feelings” are not what we mean when we refer to emotions, for the following reason:
To say that a person is happy or discontented is not merely to say that he has frequent or continuous tinges or gnawing; indeed, it is not to say even this, for we should not withdraw our statement on hearing that the person had had no such feelings, and we should not be satisfied that he was happy or discontented merely by his avowal that he had them frequently and acutely. (100)

Feelings are but “specific bodily sensations,” in Ryle’s view, and he argues that we should not equate emotions with such feeling. As he says, “compassion is not to be equated with a throb or a series of throbs, any more than … fatigue in his gasps” (84). Motives and moods, as we will see, help explain behavior, but moods and motive words do not refer to feelings.

3.2.2.1 Would Rylean Feelings Suffice for Norm-Expressivism?
Ryle argues that the concept of “feeling” offers no such explanatory power. Feelings, in his view, do not explain why we act, for feelings (as bodily occurrences) are unnecessary to action. If we are looking for a concept of emotion to explain behavior, then “feelings” are not it in his view. Ryle claims that:

There are two quite different senses of ‘emotion’, in which we can explain people’s behavior by reference to emotions. In the first sense we are referring to the motives or inclinations from which more or less intelligent actions are done. In the second sense we are referring to moods, including the agitations or perturbations of which some aimless movements are signs. In neither of these senses are we asserting or implying that the overt behavior is the effect of a felt turbulence in the agent’s stream of consciousness. (114)

Thus, feelings are not taken to affect behavior, and they then offer no explanation of behavior in his view. Given C4, the norm-expressivist needs a theory of emotion that explains how emotions motivate behavior. If feelings offer no explanation of behavior, then this concept is of no use to the norm-expressivist.

3.2.3 Ryle on Moods and Agitations
Ryle explains that mood words refer to “temporary conditions” - also referred to as “frames of mind” (83). He claims that, “In saying that [one] is in a certain mood we are saying something fairly general; not
that [one] is all the time or frequently doing one unique thing, or having one unique feeling, but that [one] is in the frame of mind to say, do and feel a wide variety of loosely affiliated things” (99). Thus, while moods themselves are not actions, they are temporary “propensities” (i.e. “tendencies” or “liabilities”) to act in certain (loosely affiliated) ways (83, 85). Moods are, according to his analysis, the tendency “to act or react in some or other of certain vaguely describable, though easily recognizable, ways, whenever junctures of certain sorts [of circumstances] arise” for the duration of the mood (96).

Ryle notes that an analysis of language (the way we use mood words) shows that moods are taken to “monopolise” a person’s frame of mind; we are not taken to be in two moods at once, for “To say that he is in one mood is to say, with reservation for complex moods, that he is not in any other. To be in the mood to act and react in certain ways is also not to be in the mood to act and react in a lot of other ways” (99). Though mood words reference a tendency toward actions and feelings, Ryle argues that the mood words denote propensities toward actions more than feelings. While he admits that “to be in a particular mood is to be in the mood, among other things, to feel certain sorts of feelings in certain sorts of situations,” he argues that “we are not thinking primarily of these feelings when we say that we feel lazy; in fact, we seldom pay much heed to sensations of these kinds, save when they are abnormally acute” (103, 104). Thus, he claims that “mood words” are of “use in characterizing the total ‘set’ of a person [(of a person’s tendencies toward actions, that is)] during that short term” (100). In this way, mood words may explain a temporary pattern of behavior in terms of one’s liability to behave in such ways. If I am in a lazy mood, then I am liable not to do much that day, and this liability explains my lying on the couch. Mood words can then, in Ryle’s view, help to explain behavior in this way. (In a similar way, as we will see, motive words also explain behavior in terms of propensities.)

Ryle argues that agitations are a subset of moods (93). Whereas some mood words like “tranquil” and “jovial” do not indicate agitations, other mood words like “harassed” and “homesick” stand for agitations (97). Ryle explains that, “To be anxious, startled, shocked, excited, convulsed, flabbergasted, in suspense, flurried and irritated, are familiar kinds of agitation. Whereas agitations can be “violent or mild disturbances,” Ryle claims that other moods (such as laziness) are not conceived of in such terms (93). Agitations, he argues, are distinct from other moods in that agitations are tendencies to act in spite of one’s inability to purposefully control the actions (97). He explains that:
though agitations, like other moods, are liability conditions, they are not propensities to act intentionally in certain ways …. The agitated person cannot think what to do, or what to think. Aimless and vacillating behavior, as well as paralysis of behavior, are symptoms of agitations in a way in which making a joke is not a symptom but an exercise of a sense of humour. (97)

He provides the example of wringing one’s hands, in that such acts may be part of the behaviors that manifest a mood of anxiety; one is not intending to wring one’s hands, but one rather does so without thinking and, thus, anxiety is an agitation (97). Thus, though moods are temporary propensities toward certain acts, they may or may not be propensities toward intentional acts.

3.2.3.1 Would Rylean Moods Suffice for Norm-Expressivism?

If the norm-expressivist’s moral emotions (anger and guilt) were moods in the Rylean sense, then it might seem that anger is likely to be a form of agitation. Arguably, the word “anger” denotes a sort of upset or commotion – one that admits of varying degrees of violence perhaps. If anger is taken to be an agitation, though, then the concept of a “mood” would imply that the angry person cannot think what to do or what to think. While this may be true at times, it would seem to be an overstatement to claim this just is the very nature of anger – that anger is taken to always work this way. One may certainly be blamed for making the wrong choice when angry, even by those who do not deny that the blameworthy person was angry; a court may well understand the anger that one feels toward a neighbor, and yet decide that one’s reaction was nonetheless unjustified. The judge need not be misusing the word “anger” when he claims that the person should have chosen otherwise in spite of the understandable anger. Thus, our concept of anger does not imply that one takes the person to be unable to think what to do or what to think. Indeed, if one’s anger and guilt were to effectively gel (in order to promote social coordination) in the way Gibbard supposes, then it would seem important that those dispositions allow us to think about what to do. Given C7, the theory of emotion must explain how anger and guilt respond to considerations regarding their consistency with other norms, their grounding reasons, and their plausibility. If anger were taken to prohibit any rational response on this theory (assuming that it was an agitation), then the theory would fail to explain how anger responds to judgments regarding its rationality.

Ryle could surely respond that the word “anger” is sometimes meant to indicate an agitation and it is sometimes meant to indicate a
non-agitating mood. If it is meant to refer to a non-agitating angry mood, then it denotes a temporary disposition toward (a disjunction of loosely associated) behaviors which allow for controlled thought and actions. In this case, the moral judgment (according to norm-expressivism) could just refer to anger in the sense of being a non-agitating mood. The moral judgment that X is wrong could then be taken to express the claim that, “It makes sense to adopt a temporary disposition toward purposeful actions (of the angry sort) whenever another does X, and it makes sense to adopt a temporary disposition toward purposeful actions (of the guilty sort) whenever one has done X.” Still, this would run into a problem meeting C7.

Recall that moods, according to Ryle’s analysis, monopolize the person’s frame of mind. As he says, mood words are of “use in characterizing the total ‘set’ of a person [(of a person’s tendencies toward actions, that is)] during that short term” (100). That is, we cannot be in two moods at once. Thus, if anger is a mood in Ryle’s sense, then when I say that it makes sense to feel angry in response to certain circumstances, I would have to mean that anger is the only justified emotion for the time being. If I said that it makes sense to be angry whenever a spouse is unfaithful, then (assuming that anger is a mood in Ryle’s sense) this would mean that it makes sense to feel only anger at such times. This seems implausible. Ryle admits that “complex moods” also occur, but that seems to be an exception as he regards it and it would be odd to think of the norm-expressivist suggesting that moral claims express judgments about the rationality for “complex moods.” (If that was the case, after all, then why would Gibbard choose anger and guilt as the narrow moral emotions?) To say that one is in a mood does not imply that one is in the mood to do or feel one specific thing, for moods imply “a frame of mind to say, do, and feel a wide variety of loosely affiliated things.” Moods are states of mind to behave in a number of loosely associated ways that reflect (in a disjunctive sort of way) such emotional states, but moods do “monopolise” (99). It’s hard to deny the possibility that our normative systems may indicate that other emotions (moods) are also called for at such times. When one says that it makes sense to feel angry at a betrayal, one does not necessarily mean that this is all the person should feel; though one need not think so, it is certainly conceivable that one also thinks it makes sense for the person to feel guilty about other things (like neglected communication) at the same time. C7 requires that the theory of emotion explain how anger and guilt respond to consistency with other norms. Thus, if anger and guilt are moods that are caused by the acceptance of norms (per C5, if the concept of “acceptance” could somehow be behaviorally accounted for) and if this acceptance implies that
one finds the accepted norm to be consistent with other norms (per C7), then C7 would only be met if feeling norms never endorsed more than one mood for the same time period. Given the implausibility of this (which I suspect would likely be noted by norm-expressivist, given their naturalism), this theory of emotion would fail to meet the norm-expressivist’s C7 if we treat the moral emotions as Rylean moods. If we try to save this possibility by recasting Ryle’s non-agitation “moods” as non-monopolizing, then moods appear quite similar to what Ryle calls “motives.” The only difference between moods and motives, in this case, seems to be that motives are longer-lasting dispositions than moods. Let us consider whether the moral emotions might be better explained as motives for the norm-expressivist.

Above, we note a challenge particular to the attempt to cast anger and guilt as moods, and we will now continue on in our understanding of the more generalized challenges. What we find, I believe, is that whether emotions are long-term or short-term dispositions toward behaviors, several of the norm-expressivist’s conditions will go unmet.

3.2.4 Ryle on Motives

Motives, Ryle explains, are also called “inclinations” in that they denote long-term propensities to act (83). He argues that “inclinations are the motives by which people’s higher level behavior is explained” (84-85). When we inquire into one’s motive for a behavior, Ryle claims that we inquire “into the character of the agent which accounts for his having acted in that way on that occasion” (89). Like moods, motives are “propensities, not acts or states” (83). Ryle maintains that moods are, thus, taken to provide an “explanation” of our behavior but not in the “causal sense” (88). Ryle treats cause, here, as an “event” which triggers an effect rather, and he regards a disposition as an explanation that is distinct from cause. (Others, of course, might treat the disposition which helps to explain an event as a sort of formal cause perhaps – in the spirit of Aristotle. We can note the way in which Ryle means it.) Motives explain one’s behavior in that they just are the dispositions to behave in certain (though roughly associated) ways over long periods of time. If Cilla is motivated by her kind-heartedness, then she just tends to act in kind-hearted ways (which are numerous of course) – not just today, but over a long enough period of time so as to become her character. One’s motives, on this view, just are the dispositions to behave as one generally does behave.

Some central differences between motives and moods are that, whereas mood words denote temporary tendencies toward action which
monopolize the person’s frame of mind (and thus one cannot be in two moods at once), motive words denote long-term tendencies toward specific actions in response to specific circumstances and they need not monopolize one’s frame of mind (100). One can have many motives, which at times compete with one another. Ryle argues that, “A person’s momentary mood is a different sort of thing from the motives which actuate him. We can say of a person that he is ambitious, loyal to his party, humane and interested in entomology, and that he is all of these things, in a certain sense, at the same time …. Moods, on the contrary, monopolise” (99). Motives, then, are long-term dispositions to act and they may co-occur with other motives.

This is not to say that all long-term dispositions toward certain behaviors are motives in his view. As with the non-agitating moods, Ryle argues that motive words are used to denote “purposeful,” and not “merely automatic,” ways of acting (110-111). Unlike long-term habits, then, which are also dispositions toward specific behaviors in response to certain situations, Ryle explains that motive words denote:

that he is acting more or less carefully, critically, consistently and purposefully, adverbs which do not signify the prior or concomitant occurrence of extra operations of resolving, planning or cogitating, but only that the action taken is itself done not absent-mindedly but in a certain positive frame of mind. (111)

Since motives are both purposeful and non-monopolizing, they would seem to better meet the norm-expressivist’s conditions than do moods. Still, even if we regard anger and guilt as “motives” in the Rylean sense, this theory will struggle to meet several of the norm-expressivist’s conditions.

3.2.4.1 Would Rylean Motives Suffice for Norm-Expressivism?

Even if anger and guilt are taken to be motives in the Rylean sense, the theory cannot meet the norm-expressivist’s conditions for a theory of emotion. If motives are dispositions, and thus they cannot be said to trigger (cause, in Ryle’s sense) behavior, then we have not provided a clear account for the way in which they explain behaviors. I will argue that the theory’s explanation of behavior is not sufficient to meet C4 (explaining how specific types of emotions motivate actions) and C5 (explaining how emotions can be caused by the acceptance of norms that we are more or less aware of at the time). Additionally, Ryle’s behaviorism is challenged to identify behaviors that evidence distinct emotional dispositions, behavior with recognizable and distinct patterns in behavior. The
theory’s inability to, in my estimation, indicate such defining behavior renders it unable to meet C1 (distinguishing between moral emotions), C2 (explaining how it is that some emotions tend to be stronger motivators of action than are other moral emotions), and C6 (explaining how broadly moral sentiments shape guilt and anger).

Richard Wollheim criticizes Ryle’s theory for its lack of explanatory power, and this criticism highlights the theory’s inability to meet C4 and C5. Wollheim explains the lack of explanatory power as follows:

[A] consequence of the Rylean view of mental dispositions is that the emotions lose whatever explanatory value they can ordinarily claim as far as what the person feels, or the way in which the person views the world, or how the person acts. The most that the attribution of an emotion can achieve for our understanding is that, by subsuming what the person does on one occasion under the larger pattern of what he ordinarily does, it takes away some of what might otherwise strike us as the unfamiliarity of the behavior. (23)

This critique would pertain, of course, to moods as well as motives. Whether the moral emotions are taken to be long-term or short-term dispositions toward behaviors, it is unclear how these dispositions to act in certain ways explain why we are so inclined to behave in these ways. There may, generally, be some ways in which angry or guilty people typically behave, but we cannot understand what led to such dispositions toward this behavior upon Ryle’s view. A long-term disposition just is the inclination to behave in such ways, and thus it does not explain how (or why) this disposition came about. Dispositions, like laws of nature, may provide a sort of causal explanation for an event (though Ryle would seemingly avoid the word causal here) in that they just are the regularities by which relevant events occur. When we’re considering human behavior, though, we want to know what triggers these law-like patterns in an individual. Ryle’s discussion of motives seems to miss this point, or it may better to say that his theory is not even intended to address this point. Dispositions aren’t the sort of thing to be put in place, or triggered by another thing (for dispositions are not things at all), and thus Ryle’s “motives” cannot explain the way in which one’s character (or mood) comes into place. While it may be quite true that dispositions are triggered by other factors (such as one’s environment or genetic make-up for example), Ryle’s account would not allow for such explanations.

Because dispositions are not events, states or processes, they cannot trigger any occurrence and, likewise, they cannot be caused in
this sense. We can try saying that dispositions explain our behavior just as the brittle disposition of a glass affects the likelihood of the glass breaking (when a stone hits it). This may, though, be a weak analogy. The brittleness of a glass is arguably manifest in a material form, whereas (on Ryle’s view) a disposition to act is not to be of this sort. Arguably, there is a material manifestation of some sort that constitutes “brittle” (such that soft things are not brittle), and thus this may be a weak analogy when compared to the human’s disposition to act in certain ways. If we try to save this analogy by saying that one’s disposition to act in a certain way may also have a material manifestation of some sort (perhaps of a neurological sort, which Ryle would have no way of knowing at this time), then we undermine Ryle’s rejection of the “mythical bifurcation” between seen and unseen causes that his argument rests upon; his point is that some concepts, and some explanations, are just not of the sort to be seen nor unseen. If the emotion is not, itself, an act or occurrence, then it cannot be triggered by anything; to seek the cause of it is a mistake in logic, as the disposition just isn’t the sort of thing to cause or be caused. To say that the disposition toward such behaviors provides a causal explanation of one’s actual behaviors is to say that a non-physical state (such as a disposition, which is neither material or immaterial) explains physical occurrences. Likewise, then, it cannot cause anything either. If this is so, though, then emotions cannot be caused by the acceptance of norms. Thus, even if moral emotions are motives, Ryle’s theory does not meet the norm-expressivist’s C5 because it does not explain how the dispositions can be caused by the acceptance of norms. It is not only unclear how the acceptance of norms could motivate, or cause, anger and guilt, but it is additionally challenging to even provide a behaviorist account (in the Rylean sense anyhow) of the acceptance of a norm. The acceptance of a norm is not, on Gibbard’s view, just self-talk or a disposition; it is a psychic state that endorses norms upon judgments of rationality, which then causes normative governance. Normative governance is taken, by the norm-expressivist, to result from the acceptance of norms. Perhaps the norm-expressivist could try to use a Rylean concept of moral emotions (as moods) anyhow, but if so then it would seem that the norm-expressivist needs to find his or her own philosophical framework for explaining the acceptance of norms and normative governance. I think this problem would also challenge the Rylean’s ability to meet C5, in that Ryle’s theory seems unequipped (at least at first glance) to utilize Gibbard’s distinct concept for the “acceptance” of a norm. Likewise, because the disposition is not an act (or occurrence) in Ryle’s view, his theory fails to fully explain how the disposition motivates actual behavior in that it does not provide for explaining a way in
which reason, enculturation or even nature itself enters into such a causal chain of events. Surely, we can explain that Cilla acts in kind-hearted ways because that’s just the kind of person she is, but this does not explain how she got to be that way (and, specifically, how the acceptance of norms contributed to this). Thus, even if moral emotions are motives, Ryle’s theory does not meet the norm-expressivist’s C4 because it does not provide a framework to explain in any very useful way how dispositions motivate behavior.

All we can know on Ryle’s account, as Wollheim points out, is that one’s behaviors are in line with other behaviors of this sort for the person over an extended period of time; they are familiar, but that is the only afforded explanation (22-23). As Ryle put it, “To say that he did something from that motive is to say that this action, done in its particular circumstances, was just the sort of thing that that was an inclination to do. It is to say ‘he would do that’” (93). Ryle’s theory intends to provide a meaning for emotion words while avoiding Cartesian dualism. It is not intended to provide for a psychological explanation with respect to the controls which can change or trigger such behavioral dispositions, and thus his analysis of emotion words lacks some features of explanation that norm-expressivism requires. What the norm-expressivist needs is a psychology, in order to explain (not just that we have patterns of behavior) but the way in which different events and even psychic states (like the acceptance of norms) can trigger such behaviors.

We should also note that, while we find specific challenges in trying to use Ryle’s behaviorism with norm-expressivism using distinct parts of his taxonomy (such as looking just to moods or motives), a more general concern for his theory is that behaviorism is challenged to meet C1, C2, and C6. With regard to C1 (which requires that anger and guilt be distinguishable from one another), we can note that if anger and guilt are to be distinguished from one another, then they will be distinguished by their dispositions toward distinct behavior. This, as we will see below (in psychological behaviorism), is a challenge for behaviorism in general. Whether an emotion is taken to be a disposition toward silent self-talk or outward behavior, or some disjunctive combination of the two, it is difficult to clearly identify what type of behavior distinguishes the emotion typology. It is challenging, and in my view (as we will see in Section 3.3.1) not yet accomplished, to demonstrate a universal behavioral framework capable of identifying anger and guilt. If we cannot

---

9 While Ryle distinguishes between purposeful and yet not necessarily cogitating behaviors in the “motive,” above for example, it seems rather unclear as to what sort of difference there really is between these two mental behaviors. What sort of mental behavior, exactly, is voluntary and yet not cogitated?
clearly distinguish between anger and guilt, then we have a problem meeting C2 and C6. If I can’t say exactly what anger and guilt are, then I cannot proceed to explain how they are stronger motivators of other behaviors than are other emotions (per C2); I cannot identify the emotions to compare, and I cannot explain any causal relationship between emotions (as dispositions) and behaviors at all. For the same reasons, I cannot explain how broadly moral sentiments shape guilt and anger (per C6).

The only condition that Ryle’s behaviorism would seem to clearly meet is C3, in that dispositions to act would not be moral judgments. Ryle’s behaviorism has difficulty meeting the majority of the norm-expressivist’s conditions: C1, C2, C4, C5, C6 and C7. We should note, though, that Ryle’s behaviorism is not intended to provide a psychological account of emotion. It is intended to provide a theory of meaning for emotion words. Thus, the fact that Ryle’s theory fails to meet the norm-expressivist’s conditions need not be a demerit in Ryle’s theory; his theory just does not aim to provide a psychological account of emotion needed by the norm-expressivist. Nonetheless, for our purpose, we need a theory of emotion that does provide a psychological explanation of the way in which emotions interact with behaviors and normative judgments. Psychological behaviorism, as we will see, aims (among other things) to offer a theory of emotion that explains how emotions arise and how emotions can cause consequent behavior. This theory will also purport that emotions can be defined in behaviorist terms (as behaviors, rather than even dispositions), but they will aim to provide causal explanations for the emotions as well. Thus, we should consider whether psychological behaviorism may better meet the norm-expressivist’s needs.

3.3 Psychological Behaviorism: Radical Behaviorism

Psychological behaviorism is particularly concerned with studying the impact of one’s environment upon behavior. With respect to such theories of emotion, psychological behaviorism identifies emotions as behaviors that can be empirically studied by others through observation.

Methodological behaviorism is a normative theory recommending that psychology focus its study on observable behaviors; it may also include a form of analytic behaviorism, taking emotions to be behaviors themselves (rather than taking them to be dispositions toward behaviors). John Watson developed methodological behaviorism, and then Watson and B.F. Skinner together advocated psychological behaviorism (largely through Skinner’s “radical behaviorism”; About Behaviorism 16). Psychological behaviorism can include both methodological and analytic behaviorism. Skinner’s radical behaviorism is a form of this (About Behaviorism 18).
Whereas Ryle’s analytic theory recommends a theory of meaning for emotion terms or concepts based upon a philosophical analysis of the way in which emotion words are used, radical behaviorism would recommend a different methodology for identifying the meaning of emotion words. Radical behaviorism, as a form of psychological behaviorism, is here taken to purport that an emotion is a distinct and observable (though perhaps disjunctive) set of behaviors which are learned through interaction with one’s environment. This section will consider whether radical behaviorism might provide a psychological theory of emotion which is consistent with the norm-expressivist’s conditions.

John Watson’s work in psychological behaviorism was furthered by B.F. Skinner’s radical behaviorism (Graham). Resulting from Watson’s experiments in experimental psychology, he argued that all behaviors (including emotions) were learned in response to stimuli, and that new stimuli could be associated with prior stimuli to elicit the same behaviors. As a result of further studies, Skinner went on to argue that, while some behaviors are learned through such a stimulus-response conditioning, other behaviors were learned responses to stimuli due to the reinforcements that the behaviors receive. A defining principle of radical behaviorism is its claim that we can experience a response to our environment without any internal mental processes. In his article “Why I am Not a Cognitivist Psychologist,” Skinner argues that we need not project mental constructs in order to explain the learning of behaviors (including the behaviors that constitute emotion); we need not, he claims, assume that animals (nor we) make associations between various stimuli, nor that we make emotional appraisals, in order to explain our behavioral responses. It is enough to know that observable behaviors occur in response to stimuli; that, alone, explains the learning of behaviors (and, thus, emotions). Emotions should not, then, be regarded as “internal surrogates” of which we can study in order to explain observable behaviors (Skinner “Why I am Not a Cognitivist Psychologist” 1); that is, we cannot study internal states in place of the actually observable ones. S

Thus, Skinner denies that the study of “cognition” holds explanatory power for human behavior. He explicitly dismisses explanations of human behavior that derive from attributions of conceptual understanding to the agent, claiming that “Children do not go around forming concepts … and consequently behaving in special ways …. Behavior changes because the contingencies [i.e., environment] change, not because a mental entity called a concept develops” (2). Skinner, then, would deny that Mary’s anger at the televised news is some sort of internal process or state – nor is a disposition to act. Her anger at the news just is whatever observable response she has to the news being broadcast. If we can iden-
tify the type of behavior that constitutes anger, then we can define the emotion in these terms and we can explain how this behavior is caused by stimuli.

Thus, unlike Ryle’s theory, radical behaviorism purports to explain how emotions are caused and how they, in turn, cause other behaviors in that these behaviors now become part of the environment. We can consider, then, whether radical behaviorism might better meet the norm-expressivist’s conditions for a theory of emotion. I will argue that, although radical behaviorism purports that emotional behaviors can cause and be caused, radical behaviorism offers no better address of the norm-expressivist’s conditions for a theory of emotion than does Ryle’s theory.

According to radical behaviorism, moral emotions (such as guilt and anger for norm-expressivism) would be but learned behaviors developed in response to stimuli and reinforcements. Anger and guilt would be behaviors that respond to certain stimuli because these behaviors either avoid negative reinforcement or produce positive reinforcement. In this case, anger and guilt are learned, and observable, behaviors.

3.3.1 Would Radical Behaviorism Suffice for Norm-Expressivism?

Radical behaviorism would be limited to describing and explaining emotions in terms of outwardly observable behaviors that respond only to environmental learning. This brings challenges in meeting the norm-expressivist’s conditions.

As with Ryle’s behaviorism, radical behaviorism is challenged to meet C1. It will be difficult to identify any universally observable behaviors that distinguish and fully constitute anger and guilt. While we need not, in the case of radical behaviorism, try and identify dispositions, we do need to identify observable behaviors themselves in order to distinguish emotion typology. If emotions are behaviors, then each emotion will need to have a distinct type of behavior; that is, there must be some distinct set of behaviors which constitute all moral emotions. In my review of literature, though, there appears to be a shortage of evidence to sufficiently prove that even the narrow moral emotions of anger and guilt exhibit their own defining behavior. Because radical behaviorists emphasize directly observable behaviors as being the criteria for any psycho-

---

11 Skinner, himself, originally denied (Science and Human Behavior 160) that emotional behaviors offered any causal explanation of other behaviors, but we will see that radical behaviorists need not accept this view.
logical phenomena, it is critical that there are in fact observable behaviors that correspond to any alleged emotions. Such data is hard to find. Because this is such a critical point to the norm-expressivist’s use of radical behaviorist theories of emotion, we should consider the evidence of behaviorally defined emotions.

Perhaps Paul Ekman and Wallace Friesen’s famous study with the Fore tribesmen in Papua New Guinea comes close to evidencing universal behavior for some emotions (“Constants Across Culture” 125). This study is known for evidencing an ability to recognize universal facial expressions in response to similar types of stimuli; regardless of culture and language, people in his study overwhelmingly picked the same facial expressions from photographs in response to similar stimuli (such as stories about great losses, threats, fortunate news, etc). Ekman’s conclusion was that there are some “basic emotions” (as coined in his article “An Argument for Basic Emotions”) that can be identified by their facial expressions, and these basic emotions include “happiness anger, sadness, disgust, surprise and fear” (“Constants Across Culture” 127). Ekman was not claiming that emotions were these facial expressions; he argued that “Each member of an emotion family shares certain characteristics, for example, commonalities in expression, in physiological activity, in nature of the antecedent events which call them forth, and perhaps also in the appraisal process” (“An Argument for Basic Emotions” 172). The basic emotions, in his view, were each a family of related characteristics. Still, if the radical behaviorist was to define the emotion in terms of outward behaviors only (rather than any internal processes), it seems that s/he could only use such defining characteristics as the facial behavior. This, though, would not be sufficient for the norm-expressivist.

For one reason, if facial expressions really are the behaviors that constitute the emotions, then radical behaviorism might imply the unreasonable conclusion that emotions last only as long as their distinct facial expressions (or perhaps the other occurrence states like physiological activity, though the radical behaviorist would seemingly be unable to accept this). This was not a conclusion that Ekman favored; he claimed that “Basic emotions can occur without any evident signal. This may be due to deliberate or habitual attempts to inhibit the appearance of a signal. Also, a threshold may need to be crossed to bring about an expressive signal, and that threshold may vary across individuals” (“An Argument for Basic Emotions” 177). Furthermore, while there may be some emotions with recognizable facial expressions per Ekman’s studies, even he acknowledged that other states that we tend to call emotions (like jealousy, hatred, and parental love) have no recognizable expressions. These types of experiences are referred to as “emotion complexes” (Expression
of the Emotions 60, 213). Unlike the universally recognized basic emotions, in his view, emotion *complexes* are taken to have the distinct ability to last over long periods of time although Ekman seems at a loss to then tell us how to identify them (*Expressions of the Emotions* 83). If the norm-expressivist considers that guilt and anger are at times long-lasting emotions (as I believe the norms for their engagement would imply), then they will not manifest always in a distinct facial expression; this is not to say that long-lasting emotions are so long lasting as to be characterological of the person, but they must be capable of being more or less enduring. Even if the radical behaviorist could account for enduring moral emotions through accepting that there is the disposition of sort toward such facial behaviors, it has not been proven that guilt even has a distinct facial expression. Though Ekman proposes that guilt may “be found to share the characteristics [which distinguish basic emotions from one another and from other affective phenomena],” he admits that “evidence is certainly not available now” (“Basic Emotions” 55). Furthermore, as seen above, many of these characteristics are internal processes. Without any proof that guilt is also a recognizable behavior, we find an asymmetry between the moral emotions of anger and guilt. Surely, these emotions cannot coordinate with one another in order to explain a moral judgment if only one of them exists. If one is to try and adopt radical behaviorism and look to facial expressions for the emotion’s behavior, then the norm-expressivist’s account of the moral judgment in terms of guilt and anger is undermined. While it might be responded that, given Ekman’s proposal here, it is conceivable that guilt and anger could both be defined in terms of facial expressions upon further studies, it is interesting to note that Ekman himself finds that facial expressions are unnecessary to the emotion even if they provide evidence in identifying it; it is for this reason that he identifies so many other internal characteristics in his criteria for the basic emotion, and thus it seems unlikely to think that such studies in facial expressions alone will evidence the distinct emotion of guilt.

Of course, it may be that the radical behaviorist could define guilt and anger in terms of other behaviors, and yet this seems even more unsubstantiated. It is very unclear what outward behavior, other than one’s facial expression, would come close to being both sufficient and necessary for feelings of anger and guilt. Cannot one ever be angry without acting toward revenge? Cannot one ever have the emotion “guilt” even though one feels no inclination to do anything about it? The radical behaviorist would deny, of course, that these emotions can exist without the inclination (in some sense) to outwardly behave in some way. This, though, seems intuitively questionable: Wollheim goes so far as to argue
that emotions are attitudinal, and thus “emotions are not directive” for “someone angry with the world is not necessarily inclined to change it, and might well get even angrier with someone who was so inclined” (28). It is unfathomable, to me anyhow, to identify a specific (even disjunctive) macro-level behavior that always characterizes guilt and anger. Even if guilt and anger were specific behaviors, then it is quite a question as to what these behaviors would be. Of course, the behaviorist could agree with this and thus deny that anger and guilt are genuine emotions. This, though, would be another counterintuitive conclusion for behaviorism, one we should probably resist if other theories of emotion fare better in explaining these emotions. If a theory of emotion requires us to deny that guilt and anger are genuine emotions, then this would not be a theory for the norm-expressivist in that case. It would clearly violate C1, in that radical behaviorism does not provide a framework capable of distinguishing between the moral emotions.

For other reasons, too, norm-expressivists would need to reject the proposal that emotions must be identified by such macro-behavior. Norm-expressivism would accept that, although emotions tend toward action, they need not result in action and should not therefore be identified with the behaviors toward which they tend. Whereas Gibbard noted that emotions cash out in action, and thus can support social coordination, he also noted that we need not act on every emotion, for we may be in the grip of other norms. These exceptions, after all, may even be explained through our system; we do have, according to norm-expressivism, other norms that determine the degree (of engagement) to which we experience our feelings, and we have other norms that silence our emotional expressions in some circumstances. A child may be angry at his soccer coach and not act upon it, out of respect for authority. I see no reason to assume that the subjects of Milgram’s experiments, who believed they were electrocuting others, felt no guilt at the time about the very actions they engaged in at the time; they may have proceeded to harm others because they were in the grip of norms to obey authority, even though they felt guilty about it while acting to the contrary. Clearly, an emotion cannot consist in one’s macro-behavior. It might be that some show of guilt seeped out in the facial expressions of those who obediently (and seemingly) electrocuted others, even though they still engaged in the gross behavior, but even this behavior is (as seen above) undetermined.

It is unclear, then, how radical behaviorism can identify anger and guilt, through either facial expressions or other macro-behavior, and this has several implications upon norm-expressivism. Radical behaviorist theories would fail to address C1 in my project; this theory does not
provide a framework with the promise of clearly distinguishing between guilt and anger. Additionally, since we cannot even clearly identify the specific emotions of guilt and anger, then (as was the case when considering Ryle’s behaviorism) we cannot then show how they are stronger motivators of behavior than are other moral emotions or how the broadly moral emotions help to shape the narrow ones; in this case, radical behaviorist theories fail to meet C2 (showing how some emotions are stronger motivators of action than are others) and C6 (showing how moral emotions help to shape the narrow ones of guilt and anger). It is possible that the emotional behavior (of, say, some universal facial expressions like that of guilt and anger if they both exist) could turn out to be higher motivators of a more macro-level behavior and that the broadly moral emotions (if we could recognize such facial expressions for them) might motivate the facial behaviors of guilt and anger. All of this, though, seems quite unsubstantiated at the time; there is no scientific consensus (or anything close to it in my review of literature) regarding universal facial behaviors for the broad moral emotions.

Radical behaviorism is also often challenged for its silence on internal processes and this will raise problems with the theory’s ability to meet C4 (Graham). The theory may seem especially hard to reconcile with any theory (such as norm-expressivism) that implies such concepts as “motivation.” If we are to meet C4, then emotions must motivate behaviors (behaviors that are particularly useful in social coordination). A common criticism of radical behaviorism is that, because the theory denies that private inner experiences cause other behaviors, we cannot make use of concepts such as “motivation” or “feelings.” According to Dale Brethower, though, a practicing radical behaviorist today, such depictions do not deny our ability to know about processes that we may typically regard as internal (such as motives and emotions). Brethower advocates radical behaviorism, though he is seemingly of a softer sort than were Skinner and Watson. Brethower argues that radical behaviorists do not deny motivation or feelings as they are (frequently) accused, but that they just identify (and, indeed, define) them in terms of the very behaviors that indicate them to us in our own observations of others. In his view, “the point is that what is "radical" about radical behaviorism is that it rules out no psychological phenomenon while analyzing all psychological phenomena in terms of behavior.” Attitudes and intelligence would be described in terms of behavioral responses to one’s environment. Of course, this brings us back to the same challenge: there does not

---

12 Dale Brethower is with the International Society for Performance Enhancement at Florida State University.
seem to be any well definable behaviors that are necessary to constitute and causally explain anger and guilt.

Along the same lines, the radical behaviorist’s rejection of internal mental processes will impact the theory’s ability to meet Conditions Five and Seven. As noted (Chapter Two), C5 states that a theory of emotion must explain how the acceptance of a moral norm comes to cause guilt and anger. As with Rylean behaviorists, the radical behaviorist is unable to utilize the “acceptance of norms” (which Gibbard calls a “psychic state”) in its causal explanation of the emotion. Gibbard intends the “acceptance of a norm” (as discussed in Chapter Two) to be a psychic state that entails a willingness to defend a norm in normative discourse; it is inconceivable, to me, how one might discuss a willingness to do something if uncensored in strictly behaviorist terms. (What behavior could demonstrate such a secret willingness? If one always feels censored, then we would never know what their accepted norms are, and thus on the radical behaviorist model such undemonstrated psychic states do not exist.) Any theory of emotion which is compatible with norm-expressivism must allow for higher level thought processes to contribute to the cause of guilt and anger – a thought process that, if uncensored (perhaps even from one’s own recognizing mind), finds a norm to be grounded in reason and consistent with other norms per C7. Again, it is hard to know how a radical behaviorist would define secret, private, thought processes. The norm-expressivist (unlike the radical behaviorist) does not take the environment alone to cause my anger at theft; the cause of moral emotions must also, at times, be my acceptance of the norm calling for this emotional response (per C5), and my acceptance of this norm is a response to higher level reasoning (per C7). If a theory of emotion cannot account for the acceptance of a moral norm in response to higher level thought, then norm-expressivism collapses as an account of the moral judgment. Radical behaviorism would not explain how this person’s emotions are, in part at least, caused by his/her judgment that anger and guilt responses “make sense” in terms of grounding reasons and consistency. Thus, radical behaviorism fails to meet Conditions Five and Seven for norm-expressivism in that it fails to explain any causal connection between the acceptance of norms, judgments of rationality, and emotional behavior.

Later psychological theories of emotion have proposed that stimulus-response (and stimulus-response-reinforcement) models of understanding human behavior fail to account for the role of the organism in determining this behavior; educators, and those who teach on educa-
tion such as Jonathan Atherton\textsuperscript{13}, will often note that an understanding of human behavior requires a “SOR” (stimulus-organism-response) model instead of “S-R” model. SOR theories purport that there is much in the individual’s internal milieu that mediates the individual responses to stimuli. Even Ekman argued that the “characteristics which distinguish basic emotions” included “automatic appraisals (emphasis added) … distinctive thoughts, memories images [and] distinctive subjective experiences” (“Basic Emotions” 56). Such S-O-R theories would better account for the impact of accepted norms and higher level reasoning. Noam Chomsky’s criticism of Skinner’s theory of language acquisition perhaps reinforces this criticism of a Skinnerian theory for the learning of emotion (“A Review of B.F. Skinner’s Verbal Behavior” 142-143). As Chomsky demonstrated with regard to language acquisition, the mind is possessed of its own capacity for learning such behaviors, and the environment alone is not sufficient for explaining this capacity. Even though Chomsky directed his critique at behaviorist speculations as to the nature of “language” and “higher mental processes,” I suspect that his criticisms hold true with regard to the mental aspects of emotion as well as language (142). Emotions may entail learned responses, but this learning requires a mental processing that is provided for by the person’s internal milieu (not merely the environment itself). Marc Hauser, an evolutionary biologist, argues that Chomsky’s argument for the necessary capacity of the mind prior to language acquisition applies to moral judgments as well; in an interview with American Scientist magazine, Hauser argued that, “in the same way that we are endowed with a language faculty that consists of a universal toolkit for building possible languages, we are also endowed with a moral faculty that consists of a universal toolkit for building possible moral systems” (qtd. in Ross). Bryce Huebner, Susan Dwyer and Marc Hauser argue that their research in current neurological, behavioral, developmental and evolutionary sciences evidences that emotions motivate morally relevant action after a moral judgment is made (2). (It may turn out that moral emotions also focus attention to morally salient features, when making a judgment, but they maintain that this is not known and this would not mean that these emotions cause the judgments.) These studies track the relationship of emotion as an internal process to the outward behaviors that they motivate. Thus, while these arguments are not conclusive, it seems that radical behaviorism has significant challenges in terms of its own merit, which may of course additionally recommend that the norm-expressivist seek another theory of emotion irrespective of the failure to meet the norm-expressivist’s ex-

\textsuperscript{13} James Atherton is with DeMont Fort University.
explicit conditions. Still, as we will see, not all forms of psychological behaviorism are radically behavioristic.

Radical behaviorism cannot offer a theory of emotion that is compatible with norm-expressivism. Such theories fail to meet Conditions One, Two, Four, Five, Six and Seven. As with Ryle’s behaviorism, this theory only satisfies the norm-expressivist’s C3. (Emotions, on this theory, need not be, nor be caused by, nor need be defined by moral judgments or moral concepts.) In spite of looking for a psychological theory to compliment norm-expressivism in such behaviorist terms, and in spite of avoiding some of the problems found in Ryle’s dispositional account, radical behaviorism fails to meet the same conditions. Let us now see if, perhaps, another sort of psychological behaviorism might better address the norm-expressivist’s conditions.

3.4 Psychological Behaviorism: Paradigmatic Behaviorism

Though Ryle’s behaviorism and radical behaviorism are not compatible with norm-expressivism, we can consider whether a more recently developed “paradigmatic behaviorist” theory of emotion better addresses the norm-expressivist’s needs (Staats and Eifert 539). Arthur Staats and Georg Eifert look to the internal biological processes to define emotion, although they also aim to contextualize these processes in terms of other empirical data. According to this branch of psychological behaviorism, an emotion is a process of “central nervous system responses that have to some extent been localized in brain areas … often referred to as the “the limbic system.” These central nervous system responses result from the perception of stimuli (which may be a sensory object or a thought) and, in turn, are taken to often motivate overt behavior. What is especially unique about paradigmatic behaviorism, is that while it defines an emotion as a central nervous system response, the theory includes a heuristic in order to “incorporate and link systematically the biological foundation

---

14 Per recent email correspondence, Arthur Staats now prefers that his theory be called “Psychological Behaviorism” although he still maintains his central tenets here otherwise. He aims to especially distinguish his theory from Radical Behaviorism.

15 The limbic system is “localized in brain areas such as the hypothalamus, amygdala, hippocampus, septum, and cingulated gyrus of the prefrontal cortes” (Staats and Eifert 541). Together, these areas function to “regulate autonomic and endocrine function, particularly in response to emotional stimuli,” they set off levels of physiological arousal in response to emotional stimuli, and they play critical roles in the formulation of particular types of memory (Swenson).
of emotion, basic behavioral study, basic human learning study, personality and psychological measurement, abnormal personality, and clinical treatment study.” The theory provides a heuristic in order to provide “a level-by-level analysis where the findings and principles of one level are the basis for elaboration at the next higher level,” and Staats and Eifert claim to provide “a bridge from biological to psychological study by its definition of emotion that incorporates and utilizes both biological and behavioral findings and is heuristic with respect to each” (Staats and Eifert 541). Thus, while they maintain that an emotion is a central nervous system response (mainly in the limbic system), they also purport that these other levels of emotion study will provide an account of other aspects (such as emotion typology).

Arthur Staats and Georg Eifert recommend what they take to be, a “heuristic guide” capable of unifying the work of diverse sciences into a comprehensive and behavioristic account of emotion (540). Staats maintains that his theory is behaviorist in that it claims that the emotion is an empirically observable behavior (changes in the central nervous system, or probably even the limbic system which can be picked up using fMRIs). He also maintains that we can speak meaningfully about emotion typology in light of the emotion’s causes and effects when he claims the following:

In the PB theory of emotion there are not a large number of different emotions that operate via different principles. Rather, there are only two kinds of emotional responses – positive and negative .... PB theory states that what is referred to in enumerating different emotions is that different stimulus situations are involved and different types of behaviors. (Staats, Behavior and Personality 231)

The emotion itself is, then, the change in the central nervous system and, yet, we can identify a meaning for the emotion concept in light of the causes and behaviors that result from these changes. In that emotion typologies are identified by their stimuli or behaviors, we can note that a full account of an emotion includes an explanation of its stimulus situations and resulting behaviors. I will argue that this theory, as developed by Staats and Eifert, may provide a framework by which to avoid many of the challenges faced by these other forms of behaviorism, although in doing so it is perhaps rather misleading to call this theory a form of “behaviorism” then at all. Whatever the name of this theory, though, I find it to offer a framework for defining and explaining moral emotions that is quite consistent.
Staats and Eifert explain that a classic challenge, for behaviorists, has been to relate behavioral principles, in the study of emotion, to principles obtained through other sciences—such as studies of emotion that take place in biology (particularly with respect to central nervous system responses in the limbic system) and psychotherapy for example. These sciences often pursue their studies in ways that exclude the data of other fields, and thus “schisms” arise. Whereas (as we have seen above) radical behaviorists traditionally explain any apparent mental phenomena in terms of learned outward behavioral responses to stimuli and perhaps reinforcement, Staats and Eifert suggest that biology highlights inborn (or later occurring) influences upon this very learning process; the “basic biological conception of emotions” takes emotions to be “central nervous system responses that have to some extent been localized in brain areas such as the hypothalamus, amygdala, hippocampus, septum, and cingulate gyrus of the prefrontal cortex, all of which comprise what is often referred to as “the limbic system”” (Staats and Eifert 541). Staats also notes elsewhere that paradigmatic behaviorism should recognize data regarding the “biological inheritance” of intelligence, and thus this biological account of ways in which emotions are learned would include such characteristics as individual inheritance in contrast to a mere environmental exposure (“Paradigmatic Behaviorism and Intelligence” 8).

Staats and Eifert recommend a holistic theory that connects not only central nervous system responses with studies on cognition (such as the impact of intelligence upon emotion and vice versa), but they propose a (somehow) behaviorist theory of emotion that includes studies on the subjective feeling of emotions (per self-reports) as well. They report that the “notion [that the subjective feeling is the experience of stimuli] is important because it provides an avenue for integrating subjective, phenomenological, and cognitive concerns into an objective behavioral formulation, although further analyses and empirical support are needed” (544). They aim for a “unified theory of emotion,” a theory under which the different sciences would agree upon terminology (defining their terms in consistent ways in terms of observable conditions) and, thus diverse studies might all come together to share in an empirically evidenced and most complete account of emotion (540). Staats and Eifert put it as follows:

Experimental study has produced important knowledge of basic principles. Clinical study has produced important knowledge and concepts of human behavior. Each by itself is incomplete and when cut off from other knowledge, and inappropriately generalized, produces a misleading conception. What is needed are the
necessary theory bridges by which the several areas of knowledge can be productively unified ... An important example is the linkage of emotional learning and emotion-instigated behavior with the language-cognitive processes and other essentially human mechanisms that are responsible for producing, expressing, and communicating emotions. (548)

Such a theory would seem to better meet the norm expressivist’s conditions for a compatible theory of emotion per Chapter Two. It would allow for broader means by which to identify and define emotion types, which in turn provide for the possibility to explain their relationships to one another in ways that radical behaviorism cannot.

Staats and Eifert argue that, in keeping with both our knowledge of the biological foundations of emotions and behavioral principles, we should expand a behaviorist account of emotion and move toward defining “emotions as central nervous system responses that must be distinguished from the physiological indices commonly employed to measure emotions” (539). Rather than looking at emotional behavior in terms of facial expressions, body posture, or even physiological signs (like heart rate and skin conductance), we can consider activity in the central nervous system (the brain, brainstem and spinal cord) or perhaps even (more narrowly) activity in the limbic system. As noted above, though, we can do so in light of psychological studies around language-cognition processes and phenomenological reports as well. Such unified approaches are intended to bridge the more traditional behaviorist descriptions of emotion with findings in biology. Thus, they aim to find one unifying theory of emotion that uses behaviorist observations of both internal processes and outward behaviors in order to empirically explain the multi-facilitated nature of emotions.

Not only do Staats and Eifert recommend a very broad form of behaviorism (one that, through their heuristic, offers a “multi-level theory of emotion” which includes studies of emotion and its types in terms of the biological and cognitive sciences), but they also distinguish themselves from classic behaviorism in that they treat emotion as a cause of the more outward macro-behavior (542). As they say, “it is because an emotion has stimulus properties that it can further elicit overt motor behavior” (Staats and Eifert 544). Thus, on this behaviorist account, emotion is not overt behavior or a disposition toward such behavior. Though the emotion may be defined in terms of empirically supported data (in addition to the stimulus circumstances which cause the emotion), the emotion is not the macro-behavior that it elicits on this view (nor is it even outward behavior of another sort, such as a facial expression or
reflex alone). The emotion is rather an entire process of central nervous system responses linked to cognition and phenomenological experience.

3.4.1 Would Paradigmatic Behaviorism Suffice for Norm-Expressivism?

The skeptic might suggest that classic criticisms of behaviorism would likewise challenge the paradigmatic form of behaviorism. If all meaningful concepts to be employed in the sciences must be defined in terms of observable states, then it may be difficult to define some of the more central concepts within paradigmatic studies. The very concept of a “stimulus” is difficult to define apart from the internal and subjective response of the subject; in this case, then, there is no stimulus but just a response (the only observable part). Staats and Eifert, though, would respond that subjective responses (such as feelings and thoughts) are known through behavioristic means after all, in that we have empirical data evidencing these subjective states. They claim that, “we may take the universality with which subjects report feelings of emotion as evidence for the stimulus aspect of emotion” (544). That is, reports of feelings may be taken as evidence of such feelings on this view. Thus, not only does Staats and Eifert’s behaviorist theory purport that concepts can be defined through means of broad empirical support, but they suggest that this more inclusive behaviorist methodology avoids the criticisms often facing classic behaviorism in that these classic behaviorist theories are often posed as the result of their exclusion of other sciences. I find that this theory meets the norm-expressivist’s seven conditions, although it may be that other theories meet these conditions with greater specificity.

With respect to C1, the theory would need to distinguish between moral emotion types. Because paradigmatic behaviorism proposes a unified theory of emotion, this theory allows for us to distinguish between emotion types through either their causes (stimulus circumstances), their resulting behaviors, or even their central nervous system responses (and ultimately, then, the relationship between them). As discussed earlier, I do not find any outward behaviors by which we can define (or identify) moral emotions. Thus, we should consider ways in which moral emotions are distinguishable from one another according to either their biological processes or their stimulus.

Should the central nervous system (and, in particular, the limbic system) processes themselves be considered as a necessary component in the identification of emotion types, then neuroscience may offer some hope in distinguishing emotion types. For example, the neuroscientists
Jorge Moll et al. conclude, in response to functional magnetic resonance imaging (fMRI scans) of moral and nonmoral emotions in response to stimuli, that “Our results show that viewing moral and nonmoral unpleasant visual stimuli activates a common network of brain areas that includes the amygdala, insula, thalamus, and upper midbrain”; these are distinct networks which contrast to those activated by stimuli that was “interesting,” “pleasant” or “neutral” (2733, 2731). Pictures portraying “physical assaults, poor children, abandoned in the streets, war scenes” were used to measure responses to negative visual stimuli that was specifically moral in nature (2731). There were statistically significant differences between the emotional response to the “unpleasant” and “moral” stimuli; there were statistically significant differences in that the moral stimuli was reported to have higher negative valence (higher degrees of aversiveness) and less arousal (less excitability) than did the unpleasant stimuli (2731). Additionally, Moll et al. confirmed their hypothesis that moral emotions “would share common neural substrates with tasks that invoke social schemas and behaviors, as well as inferences about the mental states of others” (2733). This study did not evidence in what way negative moral emotions (responses to moral offences) compare to other negative social emotions (like embarrassment), and so these findings by themselves would not be conclusive (according to my read) for the norm expressivist in meeting C1 in that such findings do not pinpoint distinctions between moral emotions (and even other negative moral emotions). Still, the possibilities for neuroscience to identify guilt and anger in part, at least, by such biological processes may be promising.

If paradigmatic behaviorism were to expand the constitution of emotion to include brain processes beyond the limbic system, then additional studies might even better evidence biological distinctions between the moral emotions as well. Luke Chang et al. have published interesting results with respect to guilt in relation to cognitive neuroscience and economics. They conducted fMRI scans to identify regions of the brain associated with “guilt aversion” and they explain that in their results “we observed evidence suggesting that when participants chose whether or not to honor an investment partner’s trust distinct neural systems are involved in the assessment of anticipated guilt and in maximizing individual financial gain, respectively” (563, 566). Chang et al. find that “These results provide converging psychological, economic, and neural evidence that a guilt-aversion mechanism underlies decisions to cooperate” (566). Through their fMRI scans, they claim to have identified “distinct neural systems” which are involved in guilt-motivated cooperation (565-566). It is not only conceivable, but seemingly possible, for the paradigmatic behaviorist to then be able to use neuroscience to identify
guilt in terms of the central nervous system. If this is possible with guilt, then it is especially possible with anger—a “basic” emotion. As Damasio puts it, “most of what we know about the neurobiology of emotions comes from studying the primary emotions” (45).  

Thus, there may be some viable evidence forthcoming that we can identify emotions in terms of neural pathways and activity, in addition to defining them by their elicitors and resulting outward behaviors. Still, this research is still rather new and we do not yet have clear evidence on the neural pathways for distinct emotions. This research on “guilt-aversion,” for example, is on the mechanisms employed in aversion, not the identification of the guilt emotion. It would seem that there is “some” relationship between the pathways for guilt aversion and the experience of guilt, of course, but cognitive neuroscience is not yet able to pinpoint answers to many distinct questions about our moral psychology. As Huebner et al. note, “existent neurological data are insufficiently precise to demonstrate the causal or temporal role of emotion in moral psychology…. these data fail to isolate the precise point at which emotion has a role in our moral psychology” (2). In Huebner et al.’s own research on emotion and moral judgments, they note activity in different regions of the brain, but they make no attempt to call one brain activity “guilt” and another “sympathy” for example; they do not name the brain activities at all, and rather talk about “strong emotional responses” (3). Perhaps this is why paradigmatic behaviorism maintains that there are only two emotions at the biological level of analysis—positive and negative. Further levels of analysis are likely needed in order to explain emotion typology in greater detail on this view.

Because paradigmatic behaviorism would also allow for distinguishing emotion types according to their causes, I find that paradigmatic behaviorism would meet C1. We need not distinguish between moral emotions through distinct biological processes for each moral emotion. In this case, though, the distinction between emotion types would require that we pinpoint causal circumstances that distinguish moral emotion types from one another. There would need to then be some factual circumstances (or “stimulus situations”) that cause distinct emotion types (Staats, Behavior and Personality 231). Paradigmatic behaviorism does not narrow down what sort of stimulus circumstances serve as the distinguishing cause of emotion types. Via its heuristic, the theory would leave it up to psychology to identify stimulus circumstances which cause distinct emotion types. Thus, while it may be that the stimulus for a given emotion type is determined (by some psychological theory) to be a sort

---

16 See Damasio’s long list of references for neurological studies on the basic emotions, page 303.
of cognition (such as a moral judgment itself), it may also be that such psychological accounts explain the response to circumstances without implying moral judgments. On my view, nothing in the field of psychology (as we will see in Chapters Four and Five) need imply that moral emotions respond only to moral judgments. Certainly, moral emotions need not be caused by moral judgments in the norm-expressivist’s sense; we need not accept norms that we would defend in normative discussion if only uncensored (Chapter One). Moral emotions may be accounted for as responses that occur prior to any such judgments, occurring at times prior to any higher level thought processes of the sort of thought processes needed for moral judgments. This topic will be further explored in Chapter Four, wherein we survey cognitivist theories of emotion (some within in psychology), and Chapter Five wherein we study somatic theories. There are evidence and analyses which purport that even moral emotions may be responses to circumstances independent of higher level thought about the circumstances at all. We need not be able to identify emotions by neurologically distinct processes, though this possibility would seem to offer the paradigmatic behaviorist yet another possibility for defining them in conjunction with stimulus situations. Thus, so long as there are some psychological theories (whether cognitivist or somatic) which distinguish between moral emotions without implying moral judgments, then paradigmatic behaviorism provides a framework for distinguishing between moral emotions and thus meets C1. The ability of paradigmatic behaviorism to meet C1, then, will ironically depend upon the ability of other psychological theories of emotion to meet C1. Although paradigmatic behaviorism meets C1 then, it does so in spite of hanging questions.

The paradigmatic behaviorist theory would be consistent with the possibility that emotions respond to considerations regarding their consistency with other norms and their grounding reasons (per C7). Due to the theory’s heuristic, which includes studies in the cognitive sciences at a level which they call “Human Learning and Emotion” (Staats and Eifert 542), paradigmatic theories would meet C7. The theory recognizes that central nervous system responses frequently (though not necessarily) relate to “language and cognition” (542). Even so, the theory would not limit the emotion to being elicited by consciously accepted norms. As Ralph Adolphs in the Department of Neurology in the University of Iowa reports, cognitive neuroscience offers “the realization that its neural regulation reflects both innate, automatic and COGNITIVELY IMPENETRABLE mechanisms, as well as acquired, contextual and volitional aspects that include SELF-REGULATION” (165). Given that the emotion is taken to include (at times at least) both acquired, volitional aspects
(in contrast to cognitively impenetrable mechanisms) as well as (at times at least) automatic cognitively impenetrable mechanisms, it would seem that emotions may be influenced by one’s acceptance of norms that we are more or less aware of at the time (thus meeting C5).

While it would be quite an ambitious endeavor, it is nonetheless conceivable that the paradigmatic behaviorist theory of emotion could address Conditions Two, Four and Six. It might be more challenging for the paradigmatic behaviorist theory to address C2, although it seems to be a rather empirical rather than philosophical challenge. The paradigmatic behaviorist would need to explain how it is that some emotions are stronger motivators of behavior than others. (Otherwise, the norm-expressivist could not justify claiming that moral judgments are about the rationality of guilt and anger in particular.) If we could distinguish between different emotions according to the combination of their distinct central nervous systems and the circumstances which cause them, then we would have a means by which to study the motivation of the different emotions toward outward behavior. (It does seem, after all, that the norm-expressivist takes guilt and anger to cash out in outward behavior more so than do broad moral sentiments, since only this type of behavior can really comprise social coordination.) The paradigmatic behaviorist could conceivably demonstrate how some emotions (such as anger and guilt, if Gibbard is correct) are stronger motivators of outward behavior. It has been shown, for example, that functional magnetic resonance imaging (fMRI) studies demonstrate stronger activity in the brain upon exposure to some moral dilemmas in contrast to other moral dilemmas (Cohen et al. 2105-2107). Perhaps, as this science becomes even more refined with respect to moral emotions in particular, the paradigmatic behaviorist could study the correlation between these emotional responses (and perhaps even in the Limbic system alone) to our more outward behavior in order to demonstrate that anger and guilt are, in fact, the strongest motivators for social coordination. (Even a correlation, though, might not be enough to show cause, but it is perhaps a natural place to start.) Still, should the results contradict Gibbard’s position on this, then this could potentially undermine the credibility of norm-expressivism. The paradigmatic behaviorist would seem open to the possibility of measuring the motivational impact of certain emotions in terms of other mediums as well; if, for example, moral emotions are identified according to their distinct circumstances, then perhaps those could be compared to overt behavior in measuring the motivational force of certain emotion types. If the paradigmatic behaviorist did pursue such studies, and if they did get the answers that the norm-expressivist wants (that is, that guilt and anger are the strongest motivating moral emotions), then this theory
would also meet for C4 as well (providing a framework to explain how emotions motivate behaviors). By demonstrating that anger and guilt provide such an exceptional motivation for socially coordinating behavior, we would certainly then demonstrate that anger and guilt motivate behavior.

It is also conceivable that the study of fMRI’s for distinct emotions, coupled with the studies of their corresponding cognitions and affective states if this can all be done, might evidence ways in which the broad moral emotions help to affect the way in which we reason about moral judgments (about guilt and anger); here too, though, the empirical findings would need to come out confirming Gibbard’s hypothesis that (per C6) broadly moral emotions shape the narrow moral emotions.

C3 requires that anger and guilt are not, nor are they necessarily responses to, moral judgments, nor can moral emotions be defined in terms of moral concepts. This theory could also address this condition. If moral emotions are, on this view, defined according to either their central nervous system processes and/or their stimulus conditions, then this theory does not necessarily imply that anger and guilt result from (nor are) moral judgments. It is unlikely, since Staats takes there to really only be two emotions (in terms of the limbic system), that moral emotions could be identified (or defined) in such terms alone. It is more likely that guilt at least (and perhaps even anger) would need to be defined in terms of their elicitors. If it is possible that the stimulus circumstances for guilt and anger do not require any moral judgments (even implicitly), then this theory meets C3. Still, while the theory is consistent with this condition, is it unclear given paradigmatic behaviorism how we identify these stimulus circumstances. Since the theory does not propose an answer to this question (leaving it open to the other fields of emotion research perhaps), it seems that it is possible to identify moral emotion types without implying moral judgments. Still, paradigmatic behaviorism leaves the nature of such stimulus quite ambiguous. I take it, then, that the theory meets this condition although not in a particularly clear way. Other theories of emotion take up the nature of these emotion elicitors more directly, and we can then consider whether specific stimulus circumstances work better or worse for the norm-expressivist. Some of them may even be compatible with norm-expressivist, and in this way other theories may amplify paradigmatic behaviorism in ways that more explicitly meet the norm-expressivist’s needs. For now, I will just note that paradigmatic behaviorism is consistent with C3 even if the details are not explicitly indicated by this theory. This ambiguity is perhaps problematic for the norm-expressivist, in that it leaves C3 hanging. Still, there is no inconsistency between paradigmatic behaviorist theories of emotion and C3.
Unlike philosophical behaviorism and radical behaviorism, the paradigmatic behaviorist theory of emotion claims to offer a definition of emotion (Staats and Eifert 541), and a framework (or heuristic) for identifying and explaining emotion types (Staats and Eifert 542), in a way that is consistent with the norm-expressivist’s conditions.

3.4.1.1 The Limitation of Paradigmatic Behaviorism

Though the paradigmatic theory of emotion is theoretically consistent with the norm-expressivist’s conditions, it may not be the most ideal option for the norm-expressivist insofar as accounting for the nature of the narrow moral emotions (guilt and anger). While the theory provides a heuristic in order to systematically integrate multiple sciences in the study of emotion into their research program, this seems to make their theory only loosely “behavioristic” in comparison with the origins of behaviorism anyhow. This broad analysis of emotion may well allow for an approach that is more compatible (than is Ryle’s behaviorism or radical behaviorism) with norm-expressivism, but the account of guilt and anger in particular will need to draw from other (less behaviorist) theories of emotion. The identification of emotion types will not turn out to be described in strictly behaviorist terms, being that the “cause” of an emotion type (which may define/identify the emotion type) is left quite open on this theory: The cause of an emotion type might be a certain sort of judgment, or perhaps a perception, or perhaps even a combination of the two. It is unclear, upon my read, what type of internal process (or external circumstance) counts as the defining “cause” of a specific emotion type. The theory leaves the identification (and definition) of emotion types up to the other sciences to demonstrate. Its intent is to offer a unified framework for all emotion studies, such that the theory is then consistent with multiple answers to this question. In doing so, though, the theory by itself cannot offer the norm-expressivist a clear account of emotion typology. It offers a general answer, by saying that we can distinguish between emotion types in terms of their causes, but this theory does not purport to tell us the nature of such causes. The nature of the causes will be demonstrated through empirical sciences, and in such a way is behavioristic. Nonetheless, we will have to look to other theories of emotion to then account for emotion typology in greater detail, recognizing that some of these theories do not count themselves among “behaviorist” theories.

Because this theory incorporates so many fields of emotion research into its heuristic, and because some of these fields seem to deal with more internal processes (such as cognition), we might wonder whether the theory is clearly “behaviorist.” We need not squabble over a
name, but perhaps it would encourage other cognitive scientists or philosophers of mind to join in this effort if the theory was rather called an “empirical paradigmatic theory.” The structure of the paradigmatic theory of emotion is, in my view, still a rather loose one although (perhaps as a virtue of this) it can incorporate other empirical theories within its own framework.

If other theories of emotion meet the conditions for the norm-expressivist, and if they offer a cleaner account of emotion typology (one in which the nature of the cause which identifies emotion types is more consistently understood), then it may seem that this theory would be preferable to paradigmatic behaviorism in terms of its independent merits (its simplicity). On the other hand, we should note that Staats and Eifert’s goal in proposing paradigmatic behaviorism’s is largely to unify all emotion studies. Thus, if there are other theories of emotion which can explain the nature of the causes of distinct moral emotions, then such theories may actually compliment paradigmatic behaviorism. Either way, then, it is important that we look at other theories of emotion in hopes of a clearer account, not just of how we might study emotion, but as to what the distinct causes of moral emotions might turn out to be. We should survey more theories of emotion in order to recommend (to the norm-expressivist) a theory which provides the most definitive means for identifying and explaining moral emotions in a fashion that is consistent with the norm-expressivist’s conditions.

3.5 Conclusion

Philosophical behaviorism and radical behaviorism are unable to offer theories of emotion that meet the norm-expressivist’s conditions for an acceptable theory of emotion. Both theories, though sometimes for different reasons, fail to meet C1, C2, C4, C5, C6 and C7. The paradigmatic behaviorist theory of emotion fares well in meeting the norm-expressivist’s conditions, offering a framework that is consistent with all seven conditions. I suggest that the inclusive methodology promoted by Staats and Eifert suggests a positive direction for the norm-expressivist in her search for a theory of emotion. In that norm-expressivism is largely grounded upon Gibbard’s naturalism, and given Gibbard’s belief that the moral life includes an account of normative governance in relation to moral emotions, it would be beneficial to find an empirically supported theory of emotion in which both normative judgments and behavioral responses are accounted for in an integrated manner. The paradigmatic theory of emotions offers much in this way. Nonetheless, while the
norm-expressivist may be consistent with paradigmatic behaviorist theories of emotion, this theory of emotion would work only if it can be amplified through other theories of emotion that more specifically account for the cause of moral emotions. We will find, I believe, that somatic theories of emotion may particularly offer the norm-expressivist the same virtues as do paradigmatic behaviorist theories, though they offer a more specific account of emotion typology. Cognitivist theories of emotions, too though, may offer some options in accounting for such emotion typology. I will now consider how cognitivist theories of emotion respond to this task.
4. Cognitivist Theories of Emotion

4.1 Introduction

Cognitivist theories of emotion purport that cognition is necessary for emotion. Cognitivists support this view with claims that emotions are about something (their objects), and they maintain that emotions must then entail a way of interpreting (or evaluating) their objects’ relevance to the experiencers’ well-being. As Martha Nussbaum puts it, “emotions are suffused with intelligence and discernment, and … they contain in themselves an awareness of value or importance” (Upheavals of Thought 1). Cognitivist theorists hold that cognition is necessary for interpreting (or evaluating) the object of an emotion. It is noted that emotions are not only about something, but they are often caused and altered by beliefs. For example, if one finds out that a misdeed by another toward oneself did not in fact occur, then one’s anger tends to dissipate. If emotions are taken to be judgments that respond to reason of some sort, then these theories might seem to explain the frequent response of emotions to judgments of their rationality. In this way, emotions may seem to be particularly subject to normative governance. Since the normative governance of emotion is critical to any theory’s compatibility with norm-expressivism, we should consider whether cognitivist theories meet all of the norm-expressivist’s conditions for a theory of emotion.

There are several ways in which emotions may be accounted for in terms of cognitive evaluation. Some cognitivists argue that an emotion is identical to a judgment. That is, they maintain that anger just is a judgment of a specific sort (such as the judgment that “an offence has taken place”). Whereas some cognitivists, as we will see, argue that emotion is a form of judgment with propositional content, others maintain that these judgments may be described in propositional terms even though they deny that emotions have a propositional content. (That is,

---

17 See Nussbaum, Upheavals of Thought, 38; Otorny et al. 4; Lazarus “Cognition and Motivation” 353.
even between philosophers who maintain that emotion is identical to judgment, we will find some disagreement among them regarding the role of propositional thought in emotion.) Alternatively, other cognitivists argue that, although an emotion is not identical with a judgment, emotion is in part constituted by cognitive evaluations among other things. For example, one might say that anger is a bodily response (and possibly an action tendency as well) in addition to the judgment that an offence has taken place. Still again, other theorists argue that emotion itself is some sort of physiological or phenomenological response that is necessarily *caused* by a cognitive evaluation. That is, an emotion is neither identical to nor constituted by cognition, but emotion is caused by cognition. Under this view, one might argue that a judgment of a specific sort (such as the judgment that “an offence has taken place”) is necessary in order to cause the emotion as a bodily response. Cognitivists will vary in the way they explain the relationship between the cognition and emotion in these ways, though they all maintain that it is the form of cognition that distinguishes emotion types from one another.

In this chapter, I will consider how different cognitivist theories of emotion meet the norm-expressivist’s conditions for an acceptable theory of emotion. I will survey several cognitivist theories of emotion, from both philosophy and psychology, each of which proposes that we understand the form of cognition that distinguishes emotion types in a different way. Martha Nussbaum argues that emotions are judgments of propositional content. Robert Solomon argues that emotions are phenomenological judgments which constitute engagement with the world. Dimensional appraisal theories of emotion, in psychology, propose that we understand emotion taxonomy in terms of the logical structure of emotion evaluations. Attributional theories (a kind of psychological theory considered by Gibbard) purport that a person is in an emotional state whenever one attributes culturally learned concepts to one’s affective state. And finally, Patricia Greenspan argues that emotions are feelings with propositional content that respond to a distinct form of rationality. Since these theories represent a spectrum of major cognitivist theories of emotion, I will introduce each theory and then consider in detail whether each meets the norm-expressivist’s conditions. In spite of Gibbard’s resistance to cognitivist theories, I find that all of these cognitivist theories meet the norm-expressivist’s conditions for a compatible theory of emotion *so long as* we either embellish or modify their characterization of specific moral emotions. In having to do this, though, we should admit that these theories as they stand are *less than ideal* for norm-expressivism even if they need not discredit norm-expressivism.
4.2 Gibbard’s Circularity Concern

As we survey cognitivist theories of emotion, we should recall Gibbard’s concern with circularity. A theory of emotion cannot entail that an emotion is a moral judgment, in the sense of being a judgment of wrongness or blameworthiness anyhow (Wise 51), without making norm-expressivism a circular theory. Norm-expressivism, as we recall, is ultimately a metaethical theory proposing a meaning for the moral claim. The moral judgment that “X is wrong” is taken to express a judgment of rationality with respect to anger and guilt. The judgment that “X is wrong” is taken to mean that “it makes sense to feel (i.e. one ought to feel) guilty if one does S, and it makes sense for others to feel (i.e. others ought to feel) angry at one who does X.” If guilt and anger were taken to be moral judgments (i.e. judgments of “wrongness” or “blameworthiness”) with respect to oneself and others respectively, then norm-expressivism would entail that a moral judgment that “X is wrong” just means that “it makes sense to judge that it is wrong for one to do X, and it makes sense to judge that it is wrong for others to do X.” The definition of the moral judgment would then be circular in its explication of “wrong.” Should anger and guilt be another form of a moral judgment (perhaps of moral relevance in some way, though not a judgment of blameworthiness or wrongdoing), then perhaps the moral claim could just be analyzed in terms that are inter-defined without strict circularity.\(^\text{18}\)

If, though, anger and guilt turn out to be judgments of wrongdoing, or blameworthiness, then norm-expressivism would imply a circular explication of the term “wrongness.”

Gibbard’s concern with this circularity challenge is noted in C3 (per Chapter Two): “The theory of emotion cannot imply that anger and guilt are, nor that they are necessarily responses to, moral judgments nor can moral emotions be defined in terms of moral concepts.” As noted by Gibbard (as well as Sinnott-Armstrong and Hill), to analyze the meaning of my moral judgment with reference to a moral judgment of the same sort is but circular and uninformative. If the cognitive judgment constituting a moral emotion is not a moral judgment, but rather a judgment of another sort, then this may seem to avoid the circularity challenge for norm-expressivism. However, if the moral emotion required the use of moral concepts (concepts of moral wrongdoing or blameworthiness), then norm-expressivism would still have problems. Norm-expressivism would then be offering at least a circular definition of “morally wrong,” for “morally wrong” is taken by the norm-expressivist to just mean that it

\(^{18}\) Thank you to Björn Eriksson for noting this specific need for clarification.
makes sense to feel angry at those who do X and it makes sense to feel guilty for doing X oneself. If anger and guilt are constituted in part by concepts of wrongdoing, or blameworthiness, then the meaning of “morally wrong” would be explained in part by the concept of wrongness to which these emotions employ. Additionally, if one can only conceive of guilt and anger concepts in terms of narrow moral concepts (i.e. wrongdoing, blameworthiness, or morally wrong), then one could not explicate a moral judgment of wrongdoing (in terms of moral emotions) without employing these very same moral concepts that one seeks to explicate. If a cognitivist theory of emotion can account for the evaluations made in the experiences of guilt and anger without employing moral judgments or moral concepts of wrongdoing, then such cognitivist theories would address C3.

The circularity challenge does not just pose a problem to the norm-expressivist with respect to the definition of a moral judgment of wrong-doing. As seen in Chapter One, Gibbard’s account of the meaning of a moral judgment (a judgment of wrongdoing anyhow) was derived largely from his naturalistic claims about how moral judgments do, in fact, operate in relation to the moral emotions of anger and guilt. His theory presupposes factual claims such as that moral judgments (of wrongdoing) express one’s judgment regarding the rationality of guilt and anger and that moral judgments exert normative influences upon guilt and anger. Thus, if his theory is well grounded, then anger and guilt cannot in fact be moral judgments of wrongdoing, even though such moral judgments must be among their influences in order for normative governance to hold. If moral emotions (in the narrow sense, of being anger and guilt) are identical to or constituted by moral judgments (of wrongdoing), or if moral emotions even require such judgments as their triggering cause, then Gibbard’s naturalistic argument must have things wrong. The same could be true if moral emotions imply the use of moral concepts. If anger and guilt require judgments of wrongdoing in any way, then moral judgments of wrongdoing cannot be judgments about the rationality of guilt and anger. In this case, our moral lives would not work in the way so described by norm-expressivism. If these moral emotions just were moral judgments, then moral emotions would not need normative governance to guide them, because all cases of moral emotions would be cases of moral judgments already. It must be possible, that is, for the moral emotions to be irrational – going against the moral judgments about them. If they could never be irrational (in Gibbard’s sense), then there is no purpose for their normative governance. Feeling norms would have no purpose.
Cognitivist theories of emotion may be able to address this circularity challenge if they can distinguish between emotive judgments and judgments of rationality. If emotive judgments are always different sorts of judgments than are judgments of wrongness, then the cognitivist theory may not imply that NE is a circular theory. The challenge here, though, will be to then clearly explain the distinct nature of emotion cognition. Gibbard is skeptical of attempts to describe emotions in terms of “‘cognitive’ judgments.” He expresses this concern as follows:

We have no sharp notion of cognitive judgment to apply. At one extreme, mental arithmetic is clearly a matter of cognitive judgment, and at another extreme, a tired feeling is clearly non-cognitive and seems fairly non-judgmental. Emotions, though, are not much like either; they have some of the features of cognitive judgments in their clearest instances and lack others. (Wise 131)

If cognitivist theories of emotion are to work for the norm-expressivist, then we will need a clear account of what this cognition is and in what sense (upon the judgmentalist theories) they are judgments. At minimum, the cognitivist can address C3 only if he can clearly depict the cognition constituting the emotion. Ultimately, this must be done without implying moral judgments or moral concepts in the narrow moral emotions (guilt and anger).

Should cognitivist theories explicate the emotive evaluation so that anger and guilt somehow appraise their objects without either consisting of moral judgments (in Gibbard’s sense) or implying moral concepts, then the norm-expressivist may find that these cognitivist theories of emotion pose no circularity challenge after all. We will now consider different ways in which cognitivist theories of emotion account for the nature of cognition and the circularity challenge.

4.3 Attributional Theories of Emotion in Psychology (Emotion as Attribution of Emotion Concepts to Inner Experience)

Gibbard considers what he calls “attributional theories” of emotion (Wise 141), and he takes these theories to be promising for the norm-expressivist. He explains that, according to attributional theories, “the workings of emotion are crucially guided by attributions of some kind – to oneself, to circumstances, or to both” (Wise 144). According to these
theories, an emotion occurs whenever one attributes an emotion concept (through applying a “label” of an emotion type) to one’s own bodily state in response to circumstances (Wise 143). According to this theory, we attribute emotion concepts to ourselves by labeling our own internal states according to rules that we have learned to use in identifying specific emotions. Thus, for example, if Dean feels an unavoidable smile on his face and excitement in his body after winning a debate competition, then he may attribute the label “happiness” to his state. Attributing the label of “happiness” (or “pride” or even “relief”) to oneself, seeing oneself in such a way, is part of what makes one happy (or proud or even relieved) on this view. Thus, Gibbard explains that, “Seeing oneself as angry plays a role in producing the syndrome we label anger – so crucial a role that a person cannot be angry unless he can think himself angry” (Wise 143).

Studies by Stanley Schachter and Jerome Singer are often taken to demonstrate that one cannot experience an emotion due to a change in one’s physiological state alone. They argue that, whereas one may experience an injection of adrenaline, as was the case in Schachter and Singer’s study, it is the interpretation of this bodily state in response to one’s surroundings (and in light of one’s society rules for attributing specific emotions) that makes the experience one of humor or anger. Attributional theories of emotion take such studies to indicate that my interpretation of the feeling, in light of my cultural repertoire, and in response to my other circumstances, is what constitutes the emotion. If I am around a silly stooge and on adrenaline (per Schachter and Singer’s studies), I’ll interpret (or “label”) my experience as being humorous and, thus, this is my emotion (384-386). If, instead, I happen to be taking an offensive questionnaire (asking questions that, in my culture, are regarded as insulting), then my response may show as anger. These theorists argue that emotions ultimately consist in one’s own interpretation of one’s state in light of other circumstances. Thus, I am angry if I see myself as angry, and I feel guilty if I would describe myself as so.

Gibbard notes that attributional theories may allow for the emotion to consist in the attribution of an emotion-label “to oneself, to circumstances, or to both” (Wise 144). The specific labeling of one’s own arousal in response to certain circumstances may cause, and distinguish, the specific type of emotion. The promising idea here, for the norm-expressivist, is that the labeling of an emotion concept need not imply a judgment about moral properties. The idea is that we can have the emotional response (the labeling) in direct response to the circumstances that elicit the emotions in keeping with the rules of our society. Gibbard notes that attributional theories allow for one’s attribution of an emotion type
to be (at least largely) determined by the societal rules defining a culture’s repertoire for emotions. Perhaps having one’s childhood teacher announce that one performed best in the class would, in some cultures, constitute a circumstance that calls for “pride,” and yet, in other cultures, such an event may call for “embarrassment” (or even a sort of “shame”) at standing out. The attribution of these specific emotion concepts (in response to specific circumstances, or cues so to speak), takes place per the rules learned in one’s culture. Thus, different emotions may exist in different cultures, whenever these cultures have disparate emotion concepts which respond to distinct stimuli. Attributional theories hold that emotions are attributions of emotion concepts to one’s internal state in light of culturally defined circumstances, whereas self-attributional theories (as a variety of attributional theories) purport that the attributed concepts which identify emotions are those attributed by oneself to one’s own state. I would only feel guilty, under the self-attributional theory, if I attribute the emotion guilt to my own state in response to specific circumstances.

Gibbard argues that this theory would allow for emotions to be not only culturally relevant attributions, but attributions that include judgments of rationale as well. He argues that a “self-attribution view could be combined with … judgmentalism … to give a broader attributional theory of emotion” (143). In such cases, he depicts this process as follows:

The workings of anger, we might claim, are guided not only by a view of oneself as angry, but by a view of one’s situation as somehow calling for anger. Both the ways one can see oneself and the ways one can see one’s circumstances are largely set by one’s culture, and so in both these ways anger will be peculiar to a culture or a range of cultures. (Wise 143)

Anger and guilt could, on this view, entail the judgment that one’s circumstances meet the rules by which one’s society identifies these emotions, and the judgment that the circumstances actually “call for” the emotion. To say that circumstances call for an emotion is not to say that they necessarily “warrant” it in a reasoned way. One may find that an emotion is “called for” in the sense that the emotion is taken to be a quite understandable response to the circumstances; the emotion is justified in the sense that it is found to be a reasonable response, and yet this need not be a judgment that the emotion is “consistent, coherent or plausible” in light of all other accepted norms. If someone is experiencing great hardship in life, then we may find it quite understandable that she feels
angry over a friend’s success, even if we don’t think that she “should” feel angry. She may even find the anger to be quite inconsistent with her other feeling norms and yet find that it feels understandable nonetheless. In this case, one finds that “it is understandable to feel angry at one who does X” even though one does not find that “it makes sense (in the way of being “rational”) to feel angry at one who does X.” Later, we will review Patricia Greenspan’s proposal about the special rationality that warrants emotions; emotions, she argues, require only enough to recommend them and thus their rationale does not require a recommendation of them based upon the balance of reasons. Perhaps Gibbard’s proposal that such emotions (on the attributional account) could entail a judgment of their warrant could refer to a special sort of rationale for emotions in this way.

In recommending such theories of emotion to the norm-expressivist, Gibbard finds that if anger and guilt are taken to be the self-attribution of these emotions (in response to one’s societal rules) combined with a finding that the emotions are understandable called for in such circumstances, then there need be no circularity for norm-expressivism. He alleges that this may carry the advantage of avoiding moral judgments, for it may be understandable to feel something according to cultural rules even if we would not say that one “should” (nor that is “rational” to) feel it. So long as the emotions are identified according to circumstances rather than moral judgments, there is no circularity. Thus, he supposes that the emotion may, on the attributional view, be a (combination of) judgment(s) while at the same time addressing his circularity concern.

4.3.1 Would Attributional Theories of Emotion Suffice for Norm-Expressivism?

Walter Sinnott-Armstrong criticizes Gibbard’s argument that attributional theories of emotion are compatible with norm-expressivism, and his criticisms highlight some interesting challenges for attributional theories in meeting both C3 (the circularity challenge) and C1 (the requirement that the theory of emotion provide a framework to distinguish between distinct types of moral emotions). Gibbard, though, refutes these criticisms. Below, I will summarize the debate between Gibbard and Sinnott-Armstrong, and I will argue in the end that Gibbard is correct in maintaining that attributional theories of emotion are a good fit for norm-expressivism. I will then move on to consider the compatibility of attributional theories with the other conditions.
4.3.1.1 Sinnott-Armstrong and Gibbard Debate (C3 and C1)

Walter Sinnott-Armstrong argues that the narrow moral emotions (of guilt and anger) cannot be identified unless one employs a judgment of wrongdoing. He rightly points out that, “If we cannot tell when an emotion counts as anger or guilt, we cannot tell when a judgment is a moral judgment” (308-309). He then argues, unconvincingly in my view, that we cannot tell when an emotion counts as anger or guilt unless we employ judgments of blameworthiness/wrongdoing. If either anger or guilt must consist of or be caused by moral judgments, or if either of these emotions is to be defined in terms of moral concepts, this would make norm-expressivism circular.

Sinnott-Armstrong’s critique is not presuming that these moral emotions need be identical to a moral judgment of oneself, but he does argue that (at least) guilt requires a moral judgment as its cause. He argues that, because physiological and phenomenological experiences lack specificity to guilt and anger, we are left relying upon an understanding of moral judgments (or at least moral concepts) in order to identify the moral emotion in terms of their causes (309). If this critique is correct, then it would render norm-expressivism circular in its metaethical analysis of the “moral judgment.” After all, if moral emotions required moral judgments as their cause, then the norm-expressivist would have to maintain that the moral claim “X is wrong” means “It makes sense to have an experience Y in response to the judgment that X is wrong.” Again, the judgment about the wrongness of X would be used to explain the meaning of the judgment that X is wrong. The implication of Sinnott-Armstrong’s criticism of norm-expressivism would be that, in meeting C1 (providing a framework for distinguishing the moral emotions from one another), the attributional theory of emotion would fail to meet C3 (the circularity concern). I will here summarize the debate between Gibbard and Sinnott-Armstrong, arguing in the end that Sinnott-Armstrong is incorrect in his critique. That is, I maintain that attributional theories of emotion may address both C3 and C1 after all.

4.3.1.1.1 The Nature of the (Alleged) Problem

Certainly, the recognition of circumstances in accordance with societal rules need not constitute a moral judgment in the norm-expressivist’s sense. As seen in Chapter Two, the norm-expressivist takes moral judgments to be judgments of rationality in a specific sense; they are judged to be rational in light of their consistency, grounding reasons, and plausibility. When we accept that a norm is rational in this sense, we would be willing to defend it in normative discussion. According to the attributional theory of emotion, one may label one’s state as being one of
“anger” or “guilt” without implying a judgment that the response is rationally consistent, grounded, and plausible. The labeling of a state in terms of an emotion concept does not imply that one would defend the attribution of this concept in normative discussion if only uncensored. If the labeling of a state in terms of an emotion concept does imply some sort of “judgment of rationality,” it is certainly of a different sort than is a moral judgment in the norm-expressivist’s sense. Thus, the moral emotion labels (of “guilt” and “anger”) can be attributed independent of any moral judgment in this sense anyhow. Sinnott-Armstrong’s critique does not suggest that the guilt and anger would imply moral judgments in the norm-expressivist’s strict sense, although he does argue that guilt and anger must be distinguished from other emotions (and thus recognized) in light of some sort of moral judgments. Of course, if people do make different sorts of moral judgments, some of which are not characterized by the norm-expressivist’s analysis, this would in itself weaken the norm-expressivist’s metaethical analysis. Nonetheless, I will focus on Sinnott-Armstrong’s critique of the norm-expressivist’s claims with respect to the relationship between moral judgments and the narrow moral emotions (guilt and anger). They may not be the same sort of moral judgment in all of the ways in which Gibbard stipulates such a judgment, but Sinnott-Armstrong maintains that one’s recognition of (at least) guilt would imply that the moral judgment refers to an emotion that can only be conceived of in terms of a moral concept. In this lies the circularity.

We can recall that C3 states that a compatible “theory of emotion cannot imply that anger and guilt are, nor that they are necessarily responses to, moral judgments, nor can moral emotions be defined in terms of moral concepts.” Sinnott-Armstrong argues that we can only identify guilt by employing moral judgments. As we will see, he argues that a moral judgment must be the cause of guilt. Even if guilt and anger are neither caused by nor constituted by moral judgments, though, if these emotions are nonetheless defined (or conceived of) in terms of moral concepts, then norm-expressivism is still circular. Thus, even if Sinnott-Armstrong fails to prove that the identification of these emotions requires moral judgments, we should still consider whether moral concepts are nonetheless required. If the recognition of these narrow moral emotions (guilt and anger) requires the use of moral concepts, this too would fail C3. If guilt and/or anger must be defined in terms of blameworthiness, then these emotions could hardly be references used to explain the concept of wrongness without making the meaning of “morally wrong” a circular one. This is so, Sinnott-Armstrong reminds us, because, “On Gibbard’s account, an act is morally wrong just when the agent would be to blame if he were to be responsible (44) so the basic issue is when an
agent is to blame” (308). Thus, if the norm-expressivist is to be compatible with attributional theories of emotion, then we will need a way to conceive of guilt and anger (within the terms of an attributional theory) without using a moral judgment or the concept of blameworthiness.

This problem need not pertain to the recognition of anger as it does to the guilt. Anger is not a problem in this regard, since we can (seemingly) recognize it in other animals without presuming that moral judgments are the cause of anger. Anger, on the attributional theory for example, might be attributed to one’s angry feeling – which is quite different than, say, the feeling of sadness – whenever this state is accompanied with other states found in anger such as perhaps the “urge to hurt (or even if intense enough, to destroy) someone” or perhaps “the urge to change a situation which threatens to obstruct a goal” (Berkowitz 268).

Sinnott-Armstrong aims his critique at the norm-expressivist’s (alleged) inability to characterize guilt short of circularity. His critique implies that in meeting C1 (providing a framework by which to identify distinct emotion types), the attributional theory could not meet C3 (the circularity challenge).

4.3.1.1.2 An Attributional Account of Guilt, Independent of Moral Concepts

Gibbard finds that attributional theories of emotion could satisfy this circularity concern, for (on this account) moral emotions are taken to be attributions of moral emotion “labels” which are derived from cultural rules that refer directly to (the perception of) circumstances rather than moral judgments (Wise 143). He does not find that the circumstances which call for guilt, according to the rules of one’s culture, need be understood through the recognition of such circumstances in terms of blameworthiness. Gibbard anticipates this particular circularity concern, recognizing that such a tempting identification of anger and guilt in terms of moral concepts would render his theory a circular one in this way. He provides his response as follows:

At first glance that looks plausible enough: to feel guilt is to feel as if one were at fault [or blameworthy]. Nothing, though, forces the observer to characterize guilt in this way. He need not invoke a notion of fault itself. Instead, he can appeal directly to the central kinds of circumstances in which, in my culture, one is thought to be at fault…The observer can then go on to use this characterization of guilt to say what it is to think a person at fault. (Wise 149)
Gibbard maintains that we can conceive of anger and guilt as the attribution of labels in response to the perception of circumstances themselves.

In recommending attributional theories of emotion, and in finding that they address the circularity concern, Gibbard suggests that it is quite plausible that emotions (and, thus, moral emotions) are attributions caused by the recognition of distinct feelings in direct response to circumstances. That is, we can respond quite directly to similar objects that are categorized in terms of relevant emotion concepts without “judging” these objects in any ordinary way. Gibbard proposes that, “by evolutionary design, different felt bodily reactions will go with different mechanisms for dealing with social cues, and for prompting the kinds of motivation that, with our ancestors, tended to promote reproduction when such cues were received” (“Reply to Sinnott-Armstrong” 322). Other theories (somatic theories of emotion, Chapter Five) develop arguments to further explain the possibility of such psychic mechanisms in more detail than do attributional theories of emotion. Sinnott-Armstrong is correct to note that, given the norm-expressivist’s account, we can only recognize moral judgments if we can recognize moral emotions. The emotions need not be caused by ordinary (deliberated, conscious, reportable) beliefs, but the emotion types (of anger and guilt) must somehow be recognizable to us if we are to recognize (and make) moral judgments. Still, whatever mechanisms may exist allowing us to respond directly to social cues, we cannot recognize moral judgments (on the norm-expressivist’s account) unless we can comprehend the specific moral emotion concepts pertaining to guilt and anger.

Thus, Sinnott-Armstrong would no doubt ask, “What type of circumstances could explain the attribution of guilt to oneself in a way that is truly independent of any moral concepts?” To avoid this circularity, the circumstances would need to be definable in recognizably non-moral terms in order for us to then recognize the moral judgments that reference these emotions. If the recognition of the circumstances which call for guilt required an understanding of blameworthiness, then we have not avoided the circularity that Gibbard seeks to avoid. Gibbard proposes that one can attribute guilt and anger based upon the “central kinds of circumstances” that elicit these emotions, as well as the “expressive and action tendencies” found with such circumstances (Wise 149). The important question according to Sinnott-Armstrong’s critique is how we can describe or recognize these “central kinds of circumstances” to which Gibbard refers. To recognize an emotion type, we would need to understand what these circumstances have in common in order to then recognize additional circumstances that elicit the same emotion type. We
need a way to describe and recognize such paradigm cases in non-moral terms.

Sinnott-Armstrong argues that it is implausible that someone could recognize the circumstances that call for “guilt,” in contrast to other emotions, if one does not already have a concept of “morality” (310). One’s conception of such a paradigm case of guilt, anyhow, just would seem to entail a moral concept in his view. He argues that moral concepts are required in order to recognize the circumstances that count as particularly relevant to moral emotions, for without moral concepts we can’t tell the difference between these circumstances and other socially rejected circumstances. We cannot, he maintains, grasp moral emotion concepts (in order to then make and comprehend moral judgments) unless we then do so via moral judgments. He explains, for example, that one could not recognize the difference between mistakes in “etiquette” (such as putting the fork on the wrong side of the plate) and mistakes in “morality” (such as theft) if one did not already understand what it was to be morally at fault (310). In identifying one’s experience as guilt rather than embarrassment, I have then (if Sinnott-Armstrong is right) referenced a moral concept. I have identified these circumstances as calling for guilt rather than embarrassment because the circumstances that call for guilt are mistakes in morality rather than etiquette. It is only by recognizing circumstances which constitute moral failures that I tend toward feeling guilt instead of embarrassment or some other emotion type. Thus, Sinnott-Armstrong argues that moral emotions can only be recognized in terms of their causes, which just are “moral judgments” (310). He concludes that, “Gibbard’s theory eventually relies on moral judgments to pick out moral judgments, and he falls back into the kind of circularity that he was trying to avoid” (310). If even moral concepts were required in order to recognize the circumstances that call for moral emotions, norm-expressivism would still be a circular theory. If norm-expressivism is to avoid such circularity, then the concepts of anger and guilt must both be understandable without utilizing the concept of blameworthiness or any other concept equivalent with blameworthiness (such as morally wrong or fault). While anger is not such a challenge in this regard, Sinnott-Armstrong finds guilt to be quite problematic.

I see no reason as to why the circumstances that elicit (the attribution of) guilt could not work in this way, though. There may be ways to describe such defining circumstances in non-moral terms. Gibbard responds to Sinnott-Armstrong by noting, first of all, that these emotions do in fact have recognizably phenomenological qualities which contribute to our capacity to recognize them. The attributional theory, after all, proposes that emotion types are identified by the labeling of body states
according to the rules learned in one’s society. Gibbard disputes Sinnott-Armstrong’s claim that “we cannot identify anger or guilt phenomenologically or physiologically, because they feel different and have different physical causes and effects in different people at different times” (309). To this, Gibbard responds that, “If things were that bad, I’m baffled how we would ever have learned to speak about emotions” (“Reply to Sinnott-Armstrong” 322). He proposes that we must in fact be able to recognize the phenomenological experiences that indicate particular emotion types, for we would need this ability in order to learn how to label these particular states as this or that emotion. He argues that, “Suppose … I have cause both for guilt and for fear, and I have a bad feeling. I may be able to sort out whether this feels like guilt or like fear without establishing the etiology of my feeling, and without finding a proposition embedded in it” (“Reply to Sinnott-Armstrong” 322). If one feels a sense of alarm (freezing up, heart racing), then this state is unlikely to be labeled as sadness and more likely to be labeled as fear or anger. Thus, Gibbard reasons that recognizable physiological or phenomenological states, particular to the individual moral emotions, seemingly exist. He maintains that the label of a particular moral emotion may be attributed (at least in part) to one’s state in response to phenomenological or physiological experiences.

Along these lines, Gibbard considers that an “adaptive syndrome” theory of emotion, a theory which maintains that emotions are identical to the physiological syndromes that distinguish them, is also compatible with norm-expressivism (“Reply to Sinnott-Armstrong” 323; Wise 138-141). In Chapter Two, I discussed evidence that may support the claim that anger and guilt do have their own distinct brain patterns. On the attributional view, it would be the labeling of these experiences that constitutes the emotion. Nonetheless, while Gibbard finds that it is possible that the attribution of a particular type of emotion is caused by the recognition of even distinct physiological (or phenomenological) experiences, he notes that a purely physiological (or phenomenological) account of emotion is unlikely given anthropological data (Wise 142; “Reply to Sinnott-Armstrong” 323). As discussed in Chapter Two, it seems wise indeed to admit that we may not have sufficient evidence (at this time anyhow) that anger and guilt are distinct (and entirely unique) bodily perturbations. Gibbard notes that “adaptive syndrome” theories of emotion, which purport that emotion types can be understood to be universal bodily responses to stimuli, are quite questionable in terms of their own merits (Wise 138-141). He rather suspects that the same physiological experiences may be shared across different emotions and, thus, distinct emotion types are more likely to exist upon those states being la-
beled differently in response to different rules about the objects of such states. Thus, while Gibbard finds that the phenomenological quality of a state may help to explain our capacity to recognize emotion types on the attributional view, it is still necessary to find a way to describe those other distinct, defining circumstances in non-moral terms.

Gibbard does propose some criteria by which we might recognize paradigm cases of guilt. He states that, “When I think of myself as guilty, I see myself as being in a state that I conceive as follows: it is typically caused by my own acts of certain kinds, it is expressed by a guilty mien, and it typically moves me toward apology and amends” (Wise 148). I have to agree with Sinnott-Armstrong, though, that this description (though couched in non-moral terms) is an insufficient description of guilt. Unfortunately, Gibbard does not explicate these “acts of a certain kind,” and if we are to grasp emotion concepts (in order to make moral judgments), then we need to understand what it is about these acts that explains particular types of emotions in contrast to other emotion types of emotion. If the best definition of guilt I get is that it is the attribution of a label in response to any number of disjunctive circumstances (direct circumstance, such as theft, pushing another off a swing, kicking a dog, etc.), then it seems I have no understanding of the underlying shared features of such circumstances. In this case, I have no basis for identifying these circumstances under the same cluster concept of guilt. This is a problem in meeting C1, for I have no real framework by which to distinguish emotion types from one another. We can hardly just depend upon a movement toward apology and a guilty mien to identify guilt. It is not hard to imagine a case in which one feels guilt over an action that one is not inclined to apologize for, perhaps because one believes that such an apology or effort at amends would only worsen things. Similarly, it is quite conceivable that one may feel guilt that one does not show in a guilty mien. Thus, Gibbard’s proposal here has not addressed Sinnott-Armstrong’s inquiry into the non-moral circumstances which might allow us to truly distinguish guilt from other negative self-regarding emotions.

Still, I think the norm-expressivist can address this criticism/inquiry. Gibbard’s general idea is that we could describe the kinds of acts that comprise the cluster of circumstances that call for the attribution of guilt, and that we can do so without employing moral concepts. This is quite plausible in my view. First off, we should note that in describing the “kind” of circumstances which distinguish guilt, we are providing an account of the emotion concept. To explain the emotion concept need not imply that the emotion is caused by such explicit thoughts, for (as we have seen) Gibbard denies this in his proposal that
we have psychic mechanisms that can respond to social cues independent of such explicit thought. To explain the emotion concept, through identifying the shared nature behind these defining circumstances for guilt, we may seem to undermine our ability to identify guilt as a direct reference to the circumstances that elicit it. The payoff for the norm-expressivist, though, is that we get a clearer framework for distinguishing the moral emotions from one another and we can still accomplish this without employing moral concepts.

It seems quite plausible that the norm-expressivist could find an attributional, and yet non-moral, description of the kind of circumstances which particularly elicit guilt. An attributional theory could do so, for example, by proposing that one attributes the feeling of guilt to one’s bodily state (perhaps when it feels like sadness19) whenever one perceives that one’s own actions are of the sort to meet a damaged interpersonal relationship. One may be concerned, here, that the reference to a “damaged” relationship implies a normative judgment. After all, to find that a relationship is harmed (or damaged) implies a standard, by which the relationship is found to fall short of the normal or good relationship. Still, so long as the moral emotions can be characterized independent of both moral norms and moral concepts, there need be no circularity for norm-expressivism.

To illustrate such a possibility, we could embellish Gibbard’s discussion of the attributional theory of emotion with Baumeister et al.’s analysis of guilt. Based upon their review of literature in psychology, Baumeister et al. propose that we understand guilt in light of its role in interpersonal transactions. They explain guilt as follows:

> guilt is an interpersonal phenomenon that is functionally and causally linked to communal relationships between people. The origins, functions, and processes of guilt all have important interpersonal aspects. Guilt can be understood in relationship contexts as a factor that strengthens social bonds by eliciting symbolic affirmation of caring and commitment. (243)

Guilt, on their view, is elicited by the perception that one’s own actions of the sort to damage “communal relationships.” By a “communal relationship,” they mean a relationship with “social bonds and caring” (243). Conceiving of guilt in such communal terms, Baumeister et al. then explain that typical types of behaviors to meet damaged communal relationships with others are taken to be “the infliction of harm, loss, or dis-

19 Prinz proposes that guilt may be an elaboration of the basic emotion (and thus basic bodily response) sadness. This will be discussed more in Chapter Five.
tress on a relationship partner” (245). Thus, the attributional theorist could maintain that whenever one has this bodily state in response to one’s own actions which are of the sort that tend to meet a damaged (communal) relationship partner, one then sees the bodily state as being the emotion “guilt.” Gibbard focuses on the possibility that we can attribute such emotions to ourselves in direct response to the circumstances that we have learned to associate with one another. In this way, one comes to feel guilty about “theft,” even if one does not at the time consider “I have damaged a communal relationship.” In applying this learned skill to specific individual instances (like stealing from a friend) one would recognize one’s sad feeling to be the feeling of “guilt” because one has learned to see this theft as the sort of action by oneself that tends to find a damaged interpersonal relationship. In such a way, the norm-expressivist might explain the circumstances that characterize the attribution of the emotion guilt to oneself without employing moral judgments or even moral concepts.

None of this is to dispute that people can also feel guilty in response to behaviors that affect distant people. Baumeister et al. explain such cases as follows:

Although guilt may begin with close relationships, it is not confined to them; guilt proneness may become generalized to other relationships, including even minimal intergroup phenomena (Brewer, 1979) and "fellow feeling" based on community spirit or collective membership or commonality. (245)

When one does feel guilty in such cases (e.g. for not sending relief money to the needy in a foreign land), one is conceiving of these suffering people as members of one’s community. Though Baumeister et al. note that there is evidence that one can attribute a feeling of guilt to oneself in such cases, they also note that “An interpersonal approach would predict, however, that guilt reactions would be stronger, more common, and more influential in close relationships than in weak or distant ones” (245). Thus, while one may well feel guilty for stealing from a stranger, we

---

20Baumeister et al. note that, whereas guilt typically responds to actions by oneself, it can also result from other situations – such as one’s profitable state – whenever this state as seen as a cue of a damaged communal relationship. (They discuss “survivor’s guilt” as but one example of this; 251-252) Perhaps we could amend the circumstances which elicit guilt to include “states” in some way, or perhaps we could conceive of such cases (as genuinely feeling “guilty” for having survived) as an instance in which seeing one’s omission to act as the circumstance to meet the damaged relationship.
might say that it is likely (on this account) that an agent would feel even *more guilty* about stealing from a friend.

Here, though, Sinnott-Armstrong would likely point out that people do not always feel guilty in response to behaviors that are known to damage communal relationships. This may seem to suggest that it is not the “circumstances” (of merely doing things that one recognizes as the sort of behavior that tends to meet damaged communal relationships) that calls for the label of guilt, but rather the judgment that such things are indeed *wrong*. Even if one generally would feel guilty about stealing from a friend, for example, there may be circumstances where one does not feel guilty about it. Sinnott-Armstrong would likely insist that this is because the perception of circumstances, alone, is insufficient to explain the attribution of a guilty feeling.

To this, though, the obvious response (available to the norm-expressivist) would be that in such cases as this (when one does not feel guilty at the theft) it is because one does not truly find one’s action to accord with the circumstances of meeting a damaged communal relationship. There are many influences upon one’s perception of circumstances, and these will no doubt impact one’s attribution of an emotion to oneself. One may not perceive the “communal” nature of the relationship at all for the time being. (One may not truly care about the relationship in the moment, and thus it is not a communal one.) One may care about the relationship and yet not anticipate the harm to the relationship. (Maybe the agent does value the communal relationship, but she just does not find her theft to really be the sort of action to meet a damaged relationship for whatever reason – distractibility, poor anticipation of others’ expectations, etc.) None of this refutes the proposal that guilt could be identifiable as the attribution of such a circumstance, though. The unconscionable thief example only illustrates why it is that we sometimes fail to attributed guilt to ourselves in cases when many others would do so. The norm-expressivist could argue that, in such cases, there is a disagreement of sort about whether the circumstances pertain, but not a disagreement about what characterizes the circumstances of guilt.

We should also note, of course, that even when people *act* in ways that (per their cultural norms) they know tend to meet damaged communal relationships, this does not show that they lack feelings of guilt. We can hardly conclude that one who steals (even with full knowledge about the consequence to his valued relationships) feels no guilt about it. Even when one does see one’s actions as the sort to meet a damaged interpersonal relationship, there are cases in which no harmless choice exists. In these cases, I suspect that we often just choose the action which mitigates harms even though we still feel guilty. I realize that
in reporting students for cheating, for example, I will meet a damaged relationship with the reported student, and yet I still report them. In some way, I regard my relationship with students as a communal one. When I report students for cheating, I often have a response that I would label as “guilt.” I sometimes do feel guilty (or so I think) when reporting students, whenever I perceive that my actions are of the sort to meet a damaged relationship with a student (by throwing the student under the bus, so to speak). At these times I also realize, though, that I would feel guiltier about failing my relationship to the college and other students in the case of my not reporting it, and thus I make the choice to report it. It seems that it is only in those cases where I do not experience much of a communal relationship with the student, or where I most fully perceive that it is not my action (but the student’s action after all) that has been the one to meet such a damaged relationship in my culture, where I feel no guilt. Nothing in the mere fact that I report it shows that I do not attribute the emotion guilt to my state. There just may be other beliefs, and even competing emotions, that more strongly influence my choice to act at the time.

It seems quite plausible that attributional theories can allow for anger and guilt to be identified, and thus referenced in the norm-expressivist’s moral judgment, without circularity. Therefore I take it that attribution theories (or at least some version of them) can meet C3. Still, we should directly answer Sinnott-Armstrong’s concern with respect to distinguishing emotion types from one another. He was particularly interested in how we would distinguish guilt from embarrassment. This presses us to consider C1 even further.

This above approach (the attributional embellishment on Baumeister et al.’s definition of guilt) would seem capable of distinguishing guilt from embarrassment, in that guilt is attributed to oneself in response to the kinds of actions that particularly meet a damaged interpersonal relationship whereas embarrassment would be attributed to different circumstances. It is quite plausible that embarrassment is attributed to oneself in response to a different feeling (such as that which accompanies blushing rather than becoming tearful), as well as in response to differently perceived actions. Embarrassment, for example, may be attributed to one’s state in response to one’s own actions that meet a “negative judgment by others” and a “desire to hide/flee” (Ho 70). It is quite possible, I believe, for one to experience guilt over mistakes in etiquette after all – even when one realizes that the mistake was one of etiquette. Imagine a child who makes a faux pas at an important dinner for his parents. He may just feel guilty (rather than embarrassed) about the faux pas if he finds it to have caused great distress to his par-
ents and if he also cares little about the judgment of others in the room (thus having no inclination to flee). Because the defining circumstances could be conceived of without the use of moral concepts, we can distin-
guish between these emotions (thus meeting C1) without rendering norm-expressivism circular – thus meeting C3 as well.

Clearly, these are not conclusive arguments in evidencing such non-moral characterizations of guilt, though I hope to have shown that it is quite plausible that there are non-moral terms by which we can identify moral emotions within the framework of an attributional theory of emotion.

4.3.1.2 Sinnott-Armstrong and Gibbard Debate (Moral Rebels)
Sinnott-Armstrong also critiques attributional theories in terms of the theories’ own independent merits with respect to “moral rebels” (309). Although such a criticism may be more of a problem with attributional theories themselves, rather than a problem with the relationship between norm-expressivism and attributional theories, it is in the interest of norm-expressivism to consider putting this concern to rest. Though Sinnott-Armstrong himself does not mention it, the particularly unfortunate im-
plication of this criticism upon norm-expressivism would seem to be that (if Sinnott-Armstrong’s critique of attributional theories with respect to moral rebels is correct), if norm-expressivism accepts attributional theo-
ries of emotion, no one could actually make a counter-cultural moral judgment.

Sinnott-Armstrong argues that attributional theories would lead to intuitively absurd conclusions with respect to “moral rebels” (309). He insists that, on the attributional view, we would not be able to recognize guilt in moral rebels, such as in vegetarians within a meat eating society. It is, rather, in spite of cultural rules that they claim to feel guilt. This vegetarian may well claim to feel guilty for eating meat, even though this vegetarian probably well realizes that the rules in her culture do not treat this action as one calling for the attribution of guilt (310). The attributional theory of emotion implies, though, that emotion types are defined according to the (culturally) learned circumstances which explain distinct emotion concepts. Thus, he takes the attributional theory of em-
o
tion to imply that, in such cases, we would have to say that this vegetari-
an is incorrect to describe her feeling as one of guilt (310).

Norm-expressivism would clearly offer an odd metaethical theo-
ry if, in aligning itself with attributional theories of emotion, this implied that moral rebels lack genuine moral emotions with respect to their own counter-cultural commitments. If this was the case, then (given the norm-expressivist’s analysis of a moral judgment) one should wonder how
anyone is to make a moral judgment that opposes the cultural norms for how we currently feel. Any counter-cultural moral judgment would then, on the norm-expressivist’s account, be a judgment that it makes sense to feel the way that the members of one’s society currently do feel in response to said circumstances (since the society’s rules for emotional responses just do define the emotion types) – which is hardly a counter-cultural moral judgment of course. If guilt and anger just are attributions of emotion labels in response to circumstances that are determined by my society, then it seems impossible to make a judgment that it makes sense to feel other than society does feel in response to such circumstances. If Sinnott-Armstrong’s argument about the moral rebel in relation to attributional theories is correct, then this implication would certainly undermine the credibility of norm-expressivism. Granted, this is not a problem in addressing C1 or C3, but it is hard to imagine any serious metaethical theory that rejects the possibility of counter-cultural moral judgments.

Clearly, Gibbard could not have overlooked such a pitfall to norm-expressivism when recommending the attributional theory of emotion. Rather, I believe that Sinnott-Armstrong overstates the shared nature of these culturally defined rules. The moral rebel criticism does not seem to be so dooming. The attributional theorist could account for the vegetarian’s guilt through finding that the vegetarian does appeal to circumstances for guilt as learned through her own (vegetarian) sub-culture. It could also be that Gibbard’s characterization of the circumstances for guilt (the guilty mien and movement toward apology) is just not quite the right one. It is certainly true that vegetarians need not express a guilty mien or try to move toward amends to the animals they eat. If guilt, though, is taken to be a response to circumstances in which one finds that one’s actions are of the sort to meet a damaged communal relationship, then the vegetarian may well see eating meat in this way. She may well see her eating meat as the type to damage her commune with animals. I find that the attributional theorist could respond to the moral-rebel critique in one of these ways, and thus I suspect that attributional theories of emotion pose no absurd consequence to the norm-expressivist’s analysis of the moral judgment. Moral judgments may certainly express critiques of societal norms.

In the norm-expressivist’s search for compatible theories of emotion, I find that much will ride upon the ability of the norm-expressivist to respond to such critiques by Sinnott-Armstrong. Attributional theories of emotion can meet both C1 and C3 due to the possibility (under such theories) that emotions can be caused by direct responses to perceived circumstances which require no moral judgment. Likewise, the reference
to guilt and anger in a moral judgment need not imply the use of moral concepts. The recognition, or identification, of these emotion concepts need not require the use of moral concepts, for it is quite possible that we can draw upon non-moral descriptions of the circumstances that elicit guilt and anger. I find that the norm-expressivist can respond to Sinnott-Armstrong’s critiques of attributional theories of emotion, and in doing so the norm-expressivist can provide a plausible account of the ways in which anger and guilt are caused and identifiable in ways non-circular to norm-expressivism. Similar arguments would then be available to other theories of emotion. So long as the emotions are identifiable according to (non-moral) defining circumstances, the attributional theory of emotion will address C1 and C3.

4.3.1.2. Attributional Theories & the Other Conditions

There is nothing inconsistent between attributional theories and the norm-expressivist’s other conditions. Recall, for example, that C4 requires a theory of emotion to explain how specific types of emotions motivate action and that C2 requires a theory of emotion to provide a framework capable of explaining how some emotions (such as anger and guilt) tend to be stronger motivators of action than are other moral emotions. If emotions are taken to be attributions of specific concepts to oneself, then such thoughts (attributing certain labels to one’s own state, thus seeing oneself in a certain way) would seem to bring with them their own action tendencies. It is quite conceivable that some attributions (such as attributions of anger and guilt) may carry stronger action tendencies than do other emotion concepts and, if Gibbard is correct, then the emotion concepts of anger and guilt would carry stronger action tendencies than would the emotion concepts of the other broad moral emotions. Attributional theories of emotion provide a framework of emotion to account for the motivational force of moral emotions. Additionally, though, attributional theories could include behaviors in the circumstances which call for attributing certain emotion-labels to one’s state. Attributional-judgmentalist theories would allow for guilt and anger to just be (or at least include) the attribution of strong action-tendencies; there is nothing to stop the attribution of action tendencies more so in some emotions than others. Of course, one might wonder whether the attribution of an action tendency is enough to actually motivate action; it need not be sufficient, but it seems reasonable that the attribution of action tendencies promotes such actions. (This is more of an empirical question, though.)

There is no real problem in the theories’ ability to meet C6 and C7. C6, as we know, requires that the theory of emotion provide a
framework that explains how broadly moral sentiments shape guilt and anger. This could be accounted for by the attributional theories in that many of our belief systems have structures in our mental lives; thus, it is quite conceivable that the attribution of broad moral emotions (determining that one feels compassionately toward a certain cause, for example) could then influence one’s tendencies to experience narrow moral emotions (feeling guilty, for example, at then acting ambivalently toward such a cause). C7 requires that the theory of emotion explain how guilt and anger respond to considerations regarding their rationality (their consistency with other norms, grounding reasons and plausibility). Attributional theories do not claim that emotions are judgments made by reason, and yet the theories (as Gibbard considers them) may include the judgment that the inner states are “called for” in some reasonable way. While the emotions would not need to consist of attributions that are consistent, grounded and plausible in the way in which we come to endorse moral judgments, there is nothing to prohibit judgments of rationality from influencing the way in which we attribute emotions. If emotions require attributions that we take to be “called for” by the circumstances that elicit them, and thus include judgments that they are understandable, this would not make them determined by judgments of rationality; nonetheless, there is nothing to prevent judgments of their rationality from influencing them. After all, it seems quite understandable to attribute concepts to oneself which one finds to be rational to attribute.

The attributional-judgmentalist theory can meet C5 as well, though this may be at first glance unexpected. C5, as we recall, requires that the theory of emotion explain how emotions can be caused by the acceptance of norms that we are more or less aware of at the time. If one must attribute an emotion to oneself (by labeling it in response to specific circumstances that call for the emotion-concept), then, in one sense, I cannot have the emotion without identifying the norms that call for the emotion. Of course, one might respond by saying that one can attribute emotions to oneself that one is not really aware of. This, though, is then a puzzling notion of self-attribution. It is more plausible to say that there are subconscious norms for circumstances that I perceive (perhaps even circumstances that I perceive subconsciously) which cause me to attribute in a more conscious way the emotions that I experience. It is quite conceivable, after all, that people have subconscious norms that bias them toward experiencing certain emotions in response to subconsciously perceived circumstances. If someone fears another, for example, one may consciously attribute the state of “fear” to one’s state and consciously find that this state results from limited circumstances: the mere fact
that one is walking alone at night and that a crowd of people is approaching. One need not realize (or admit to oneself) that it is, perhaps, because it is a crowd of young males that this situation results in fear. On a more conscious level, one may only recognize one’s state as fear in response to “a crowd approaching me, alone, at night,” and thus fear exists in that one has attributed the label to one’s situation. This does not preclude norms that we are less aware of from contributing to and causing in a sort of higher/more removed way one’s attribution of an emotion concept. Thus, attributional theories would seem to address C5.

Unlike Gibbard, I find that attributional theories are only compatible with norm-expressivism assuming that we find a non-moral way of describing the emotion concepts guilt and anger. It is not enough to just say that the emotions can result from a label in direct response to a cluster of circumstances. We need a way of adequately describing these circumstances and we need a good argument for such a depiction. Gibbard never proposes a description of the moral emotion concepts in any terms that specify their eliciting circumstances, and yet this is necessary if attributional theories are to work for norm-expressivism as he thinks they do. Without that, as Sinnott-Armstrong correctly notes, we cannot recognize moral judgments on the norm-expressivist’s view in any way that avoids circularity. I find that it is quite plausible to describe these moral emotion concepts in terms of non-moral claims pertaining to the perceived impact of actions upon interpersonal relationships. Other cognitivist theories of emotion may certainly borrow from this possibility as well. Whenever this is so, though, I find that these theories will meet the needs of norm-expressivism only through a revision (or at least an embellishment) of their own depictions of these moral emotion concepts. Thus, I take it that attributional theories of emotion can then meet all seven conditions for a compatible theory of emotion for norm-expressivism in a conditional way. We can now consider whether such plausibility transfers to other cognitivist theories as well.

4.4 Nussbaum’s (Pure) Cognitivism: Emotion as Cognitive Judgment Only

Gibbard would surely refer to Nussbaum’s theory of emotion as a form of “judgmentalism” (Wise 129). Within the study of emotion, judgmentalism is taken to purport that an emotion is a form of judgment. Nussbaum argues that “emotions are forms of evaluative judgment that ascribe to certain things and persons outside a person’s own control great importance for the person’s own flourishing” (Upheavals of Thoughts...
These judgments, she argues, are of a specific sort. She claims to be endorsing the Stoics’ view that “a judgment is an assent to an appearance. In other words, it is a process that has two stages. First it occurs to me or strikes me that such and such is the case” (36). I may have a quite immediate impression, for example, that the bear before me poses a threat. An emotion, in her view, ultimately consists in the second stage, at which point I “accept or embrace the way things look, take it into me as the way things are: in this case the appearance has become my judgment, and that act of acceptance is what judging is” (37). At this stage, I judge that the bear is indeed a threat upon my well-being – and this assent constitutes fear. She maintains that, in judging a situation as so, we assent to the appearance that these objects bear upon our well-being (41). Nussbaum argues that emotions are, thus, judgments that constitute “eudaimonistic” evaluations (147). Nussbaum’s theory offers a form of pure cognitivism. Thus, an emotion, in her view, is identical to an agreement of sort with an appearance that is relevant to one’s well-being. In other words, an emotion is identical to such judgment.

Nussbaum claims that, in assenting to an initial impression (or way of seeing something), one makes a cognitive judgment and she disputes that feelings are even a necessary part of emotion (61). The necessary and sufficient component of emotion, she argues, is rather the cognitive judgment. Nussbaum explains the cognitive role in emotion as follows:

Assenting to or embracing a way of seeing something in the world, acknowledging it as true, seems to be a job that requires the discriminating power of cognition …. In this case, it is reason itself that reaches out and takes that appearance to itself, saying, so to speak, “Yes, that’s the one I’ll have. That’s the way things really are.” (38)

Nussbaum provides the example of her grief at the death of her mother. She argues that her grief was identical to the judgment that “someone tremendously beloved is forever lost to me” (40). Her grief just is reason’s assent to the claim that someone tremendously beloved is lost to her. It is reason, then, that assents to a particular way of regarding the object of one’s emotion in light of one’s own well-being.

Though Nussbaum’s judgmentalism is a cognitivist theory of emotion, she adopts a broad view of what these judgments (i.e. cogni-

---

21 In sections 4.4 and 4.4.1, all page citations for Nussbaum are from Upheavals in Thought: The Intelligence of Emotions (Cambridge University Press, 2001) unless otherwise noted.
tion) consists in. She explains that “by ‘cognitive’ I mean nothing more than ‘concerned with receiving and processing information.’ I do not mean to imply the presence of elaborate calculation, or computation, or even of reflexive self-awareness” (23). In defining emotion cognition in such broad terms, Nussbaum’s position on this point is not uncommon among cognitivist theories of emotion today. As we will see, cognitivist theories of emotion often argue that emotion cognition can process information and result in responses to its interpretation of stimuli in spite of even conscious rejections. Even if we know that a haunted house is just an illusion, for example, we may still feel frightened at the images and the thoughts we engage while walking through it. In that such thoughts evaluate our perceptions, the thoughts are the mental states that suffice for cognition. In Nussbaum’s language, if I feel scared at these illusions, then I have “assented” to their appearances. I need not have deliberate, controlled thoughts, but if I am frightened then I do judge in some way that my initial impressions are correct. Assenting to these impressions is an act of reason. She maintains that, through reason, I choose these impressions as correct.

The term “cognition” is clearly used in a broad sense within Nussbaum’s theory. It refers to both faster (primitive) and slower (more deliberative) forms of judgment. It is important to note that Nussbaum’s theory includes these lower forms of cognition in emotive judgment in order to consider the broad way in which she denotes emotion. Tamar Szabo Gendler argues for a new term in order to specify these distinct, faster, more primitive states of cognition. She coins this term “alief,” and by this she means “a more primitive state than either belief or imagination: it directly activates behavioral response patterns (as opposed to motivating in conjunction with desire or pretended desire)” (642, 634). While Nussbaum may not agree that such primitive or unconscious cognition need activate behavioral (or affective) response patterns, she would seem to agree that cognition includes mental states that occur in a much more automatic way independent of ordinary conscious beliefs and desires. Gendler makes the point, too, that cognition includes representations that may be quite distinct from these more ordinary forms of conscious belief. While we may “believe” (in a more controlled, consciously deliberated way anyhow) that a threat in a haunted house, for example, is an illusion, we may respond with the “alief” that threats are present indeed. Gendler proposes, too, that both modular and enduring

22 The cognitivist Mikko Salmela argues that “a cognitive theory need [only claim that emotions] necessarily involve cognitions, whether or not these cognitions are under our organismic control during the emotional experiences” (“Can Emotions be Modelled on Perception?” 8).
emotive judgments would qualify as alief, under this description (642). I may be quite consciously aware that I fear the haunted house, even though I know it is an illusion. I do not believe in any ordinary way that the house is a threat, and yet my fear of the house (under such a view) would be constituted by this cognition. The concept of alief highlights attempts (such as those by Nussbaum) to include lower (faster) forms of mental representation in the cognition associated with emotional response. Though Nussbaum does not discuss “alief,” Nussbaum’s use of “cognition” would seem to include such fast and unconscious states as will many cognitivist theories of emotion.

While Nussbaum denies that emotions need to be ordinary calculated or self-aware judgments, she emphasizes that there are specific ways in which emotions are cognitive: she argues that emotions entail intentionality, belief, and evaluations (24-31). While emotions are not necessarily conscious, controlled computed beliefs and evaluations, emotions entail intentionality in that they are “about” something (their objects) (27). If one is angry, one is angry about something (such as a deception) and, likewise, if one feels guilty then one feels guilty about something (such as an insult). Nussbaum explains that, with respect to emotions, “Their aboutness is … internal, and embodies a way of seeing” (27). If one feels guilty for stealing, then one sees one’s own theft in a certain way and one accepts that this appearance is correct. She maintains that emotions are ultimately about their “propositional content,” and specific thoughts constitute this propositional content (39). She explains, again with the example of her grief over her mother’s death, that “The appearance … has propositional content or at least combination: it combines that thought of importance with the thought of loss, its content is that this importance is lost” (39).

Nussbaum argues that the identity of an emotion type depends upon its propositional content. She explains the significance of propositional content as follows:

it is of crucial importance to get clear about the precise content of thought we ascribe to the person. For if we were to make the salient thought one with no evaluative content, say, “Betty Craven is dead” (my mother’s proper name), we would be right to think that the acceptance of that thought could be at most a cause of grief, not identical with it. Again, if we put value in without the reference to the self, saying that the content of the thought was, “Betty Craven, a most valuable person, is dead,” again – we would not have a thought that we could plausibly identify with grief. (41)
Grief is the assent to an appearance, and this appearance (or thought) has the propositional content that “someone tremendously beloved is forever lost to me” (40). Nussbaum maintains that the propositional content which identifies an emotion type tells of the personally significant evaluative claim which is assented to in all experiences of emotion of that same type. The propositional content of an emotion is “both evaluative and eudaimonistic, that is, concerned with one or more of the person’s important goals and ends” (41). In keeping with this analysis, then, anger and guilt are judgments of their propositional content. These emotion types would be defined according to the distinct propositional content which reflects these eudaimonistic evaluations. On Nussbaum’s theory, emotion types are taken to be identical to judgments with distinct propositional content.

This does not imply that all emotion types entail conscious, articulate thought (talking to oneself about such evaluations, or some such thing), for emotions may include judgments in an even broader sense (such as, I would suggest, judgments akin to something like Gendler’s “alief”). Still, the identity of an emotion type is determined by its propositional content on Nussbaum’s view. We should note that, while Nussbaum takes emotive judgments to be evaluations made by reason, she does not employ these “judgments of reason” in the way in which Gibbard uses the phrase “judgment of rationality.” Nussbaum’s emotive judgments are not identical to judgments of rationality, in Gibbard’s sense, because Nussbaum includes broad sorts of cognition that are not consistent with Gibbard’s description of rationality (per Chapter One). An emotive judgment, for Nussbaum, need not be a judgment about the “consistency, grounding reasons, and plausibility” of propositional content. There may be emotion types whose cognition is more akin to alief. An emotive judgment, on Nussbaum’s view, does not imply a willingness to defend this propositional claim if only uncensored, as does a judgment of rationality on Gibbard’s view (which implies the acceptance of norms in response to normative discussion). As Nussbaum explains these judgments, there need not be any reflexive self-awareness or elaborate calculation. We might wonder, then, if the moral emotions (of guilt and anger) could not, on Nussbaum’s view, be identified according to propositional content that is quite distinct from genuine moral judgments. If so, though, then the propositional content of these emotions must (of course) be of a sort that does not entail the moral concept of wrongness (nor any concept equivalent to this).

Nussbaum acknowledges that non-human animals and infants also have emotions, and she explains that, “even in the [adult] human case language is far from being the only medium in which an emotion’s con-
tent can register” (91, 128). She accepts that emotions can be both caused and registered without our full awareness and through means which are, themselves, not translatable into linguistic form. She cautions that with regard to animal emotions anyhow, “the distortion [that occurs when trying to articulate all emotional content] involves not only shifting to a new medium, but also ascribing to the emotion a level of articulateness and definiteness that it has not attained, and perhaps could not attain” (128). Nonetheless, while Nussbaum admits that this may also be the case with the self-ascription of some adult human emotions, her theory maintains that we do (and should) describe adult human emotions in terms of the propositional content that defines them (127). Adult human emotions are, for the most part, taken to be identical with judgments of their distinguishing propositional content. Animal and infant emotions aside, Nussbaum holds that the identity of adult human emotions is realized in an accurate articulation of their eudaimonistic evaluations.

Given Nussbaum’s regard for adult human emotions as generally consisting of propositionally describable judgments without any loss of clarity upon translation, she then proposes propositional content for anger and guilt. She proposes that guilt is a “keen identification of a past wrongdoer with the agent’s own present self” (146). Anger also seems to imply a judgment of a wrong (by another in this case). She claims that, in cases of her own anger anyhow, “a background attachment to one’s own worth and self-respect (seen as vulnerable) is combined with a judgment that certain pervasive wrongs were taking place” (74). The (narrowly defined) moral emotions, for the norm-expressivist, are identified as judgments of wrongdoing on Nussbaum’s view.

Nussbaum does not try to give non-moral descriptions of the propositional content for moral emotions. She is not concerned with the circularity problem, since she is not a norm-expressivist of course, and thus this is not a flaw in her view, but she finds it necessary to include the individual’s normative judgment in our understanding of the propositional content of moral emotions for other reasons. She emphasizes the “evaluative” (and not merely descriptive) nature of the emotion’s propositional content. Thus, she rejects “social constructivist theories of emotion,” which tend to provide (attempts at) descriptive (rather than normative) meanings for emotion. Nussbaum claims that social constructivist theories purport that societal “practices, in their specificity, make a difference to a society’s emotional repertory” (151).23 (The target of her criticism here may well include attributional theories of emotion.) Emotions, on this view, are judgments that are essentially dictated by rules

---

23 Nussbaum directs us to the work of James Averill (1980, 1982).
constructed within one’s society. She argues, well I think, that while it is true that our emotional repertoires are certainly impacted by our cultural norms and use of language, emotions are at times very individualistic judgments which are determined by features other than societal rules (150-169). She argues that social constructionists are mistaken in their view that “cultural forces leave no room for individual variety and freedom; that they make the details of a personal history aetiologically unimportant; that they create mutually inaccessible worlds” (169). The individual person’s emotions are not likely to be identified by the person’s assent to cultural norms alone (as an assent to merely descriptive claims which recognize societal rules), for such descriptive claims are not sufficient nor perhaps even necessary for individual emotion. She takes it that societal rules could not account for the personal and eudaimonistic nature of emotive judgments. Anger and guilt, on her view, are evaluative judgments of wrongdoing.

4.4.1 Would Nussbaum’s Pure Cognitivism Suffice for Norm-Expressivism?

If Nussbaum’s judgmentalism is compatible with norm-expressivism, then it will need to meet the norm-expressivist’s conditions per Chapter Two. The most obvious challenge for cognitivist theories of emotion, as noted above (Section 4.2) is the circularity challenge. Nussbaum’s theory would seem to fall within the target of Gibbard’s circularity concern with judgmentalist theories of emotion. In addressing this circularity concern, the norm-expressivist’s third condition (C3) stipulates that moral emotions cannot be moral judgments, nor can they be either caused by moral judgments or even defined in terms of moral concepts. Since Nussbaum holds that ordinary adult human emotions (like moral anger and guilt) are identified by their propositional content, and since this propositional content is taken to be of the sort that can be articulated, we ought to consider whether there is any conceivable description of the propositional content for guilt and anger which meets C3 for the norm-expressivist. In order to consider norm-expressivism’s compatibility with Nussbaum’s theory, we should explore whether there is a way to articulate the propositional content of anger and guilt in non-moral terms.

Nussbaum’s theory would seem to meet C1, which requires that the theory of emotion provide a framework capable of distinguishing between guilt, anger, shame and disdain (and other broadly moral emotions). Her theory does provide a framework for distinguishing these moral emotions from one another, for these emotions would be distinguished from one another according to their specific propositional con-
tent. As seen above, Nussbaum proposes that guilt is a “keen identification of a past wrongdoer with the agent’s own present self” (146). Such a definition, though, would violate C3 by implicating a moral judgment in the definition of guilt. Anger, as she describes it, would also violate C3, for she (in part) defines anger in terms of “wrongs” that have occurred. A moral concept is then also employed in the definition of anger. Emotions are taken to be evaluative judgments, and in the case of anger and guilt these judgments are conceived of (in part at least) in terms of moral concepts on her view. Though her theory can meet C1, the way in which it does this undermines her theory’s ability to then meet C3 because anger and guilt (as she describes them anyway) are thus taken to imply moral judgments of wrongdoing.

As with the attributional theory of emotion, though, we should consider whether a modified characterization of guilt and anger (still within Nussbaum’s general framework explaining emotion) might better allow her judgmentalism to address both C1 (distinguishing between guilt, anger, shame and disdain) and C3. If we modify our description of the propositional content needed to define the emotion concepts of anger and guilt, then we might better avoid moral concepts. To do so, I must be cautious of new definitions which are either insufficient as accurate depictions of this propositional content (failing to provide both necessary and sufficient properties for this thought content) or definitions which somehow smuggle in moral concepts.

Upon consideration, several seemingly alternate definitions of anger and guilt will show to imply moral concepts. Obviously, if I say that one experiences anger whenever one assents to the claim that “another has disrespected me,” that is insufficient. If one is not at all invested in the respect of another person, or if one finds the disrespect to be called for, then one may be quite unlikely to be angered by the disrespect. Anger, then, is not sufficiently characterized by such a claim. If we seek a more sufficient definition, and we consider that anger is assenting to the propositional content that “a person whom I desire respect from has disrespected me in a way that I think was unwarranted,” then my use of “unwarranted” now seems to stand for “blameworthy.” This would clearly make norm-expressivism a circular theory. Likewise, if I say that guilt is assenting to the view that “I have done something to harm another” or that “I have done something that unreasonably harmed another,” we run into the same problems. If one tosses a baby out a window in an irrational rage, and someone walking by happens to catch the baby, it is quite conceivable that the agent would (in time at least) feel guilty for the act – regardless of whether things turned out okay. Even if the baby, herself, was unaffected by the experience and even if the person who caught the
baby experienced no distress (which is hard to imagine of course), one can still well imagine the agent feeling guilty about the act. To explain that guilt is assenting to the claim that “My actions have brought about unwarranted dangers” (such as tossing the baby) would seem to but bring in moral concepts again – thus violating C3.

We will need to make the definitions of guilt and anger more specific, denoting a shared feature in each emotion type other than mere “harm,” “unwarranted harm,” or such “unwarranted risks.” Since guilt is an emotion that (as discussed earlier) is especially relevant to interpersonal/communal relationships, and since trust is a critical value in interpersonal relationships, one might try to state the propositional content of guilt as “This action by me will cause others to lose trust in me.” Here again, though (as with the harms), we can feel guilty even when we know that no actual loss of trust occurred. Likewise, we may not feel guilty merely because one stopped trusting us, so long as we find the trust to be inappropriate or undesirable at least. If a person decides to confess to the neighborhood that she’s abusing neighborhood children in her home and she “trusts” the neighbors not to report it, then it is quite conceivable that the neighbors will report it and that the neighbors will not feel guilty about reporting it. (If someone trusts us to do the “wrong” thing, then it seems we frequently don’t feel guilty in these cases.) Thus, realizing that others have lost trust in oneself is neither necessary nor sufficient for depicting cases in which guilt occurs. This would not, then, be the propositional content of guilt. Likewise, if we say that the propositional content of guilt is “My acts have wrongly broken another’s trust,” then this definition clearly employs the concept of wrongdoing. Whereas guilt seems to bear a special relevance to interpersonal relationships, anger does not seem to be as closely correlated to these interpersonal relationships. People certainly can become angry in interpersonal relationships, of course, but they can also become angry at the weather or another driver on the road without even focusing on this other driver as a relationship partner. Thus, the concept of trust is unlikely to be a central tenet in the propositional content of either guilt or anger.

One might try to frame anger and guilt as judgments about something like “non-moral goals.” After all, Nussbaum maintains that the propositional content of each emotion type is “concerned with one or more of the person’s important goals and ends” (41). Perhaps it is the case that one feels guilty for failing to meet one’s own goals which can be described in non-moral terms, and one feels angry at another (or even oneself) whenever this person interferes with one’s goal. This approach is on the right track toward meeting the norm-expressivist’s needs, although (as we will now see) an explicit reference to goals works better for
anger than it does for guilt. In the case of both anger and guilt, it seems that the propositional content of each emotion type would need to be quite modified from Nussbaum’s characterization if it is to work for the norm-expressivist.

Guilt might seem to have a negative relationship with self-image and/or self-esteem, and thus perhaps we should wonder if guilt is the judgment that “Through a specific act (or failure to act), I have failed to live up to my own character-goal.” One might argue that such character goals are chosen out of desire, rather than a moral judgment. If guilt is the judgment that one has failed to live up to one’s own character-goal, then, for example, one would feel guilty about failing to practice piano enough whenever this lack of practice is judged to be a failure to satisfy one’s own goal of being a good pianist. One could feel guilty about ignoring one’s child’s request for attention, whenever this behavior is judged to be a failure to satisfy one’s goal of being an attentive parent. One could feel guilty for saying something that hurts another, whenever this is action is judged to be a failure to satisfy one’s goal of being a kind person (one who causes others to feel better, instead of worse that is). We could try arguing that such judgments about “failures” are not moral judgments, but rather descriptive evaluations – the mere recognition that this conduct is incongruent with one’s own goal. So long as the goals are taken to be caused by some state of mind (such as by desire) other than moral judgments themselves, and so long as the “failures” to maintain one’s goals are also interpreted to be judgments about congruency rather than moral judgments, then it seems that no circularity need exist for the norm expressivist. If this explanation of the propositional content of guilt is to work, then norm-expressivism would entail that the moral claim that “X is wrong” means “It makes sense to judge that I have failed to meet the character-goals I desire whenever I do X.” (Anger need not have the same propositional content as guilt, and thus we cannot know at this point in our discussion how the moral judgment’s reference to anger should translate. We can only see, though, that this reference to guilt would not make norm-expressivism circular anyhow.)

Although this approach would meet C3, it does not reasonably address C1. To characterize the propositional content of guilt in terms of failed character goals will not sufficiently characterize cases of guilt. The problem with trying to stipulate the propositional content of guilt in terms of non-moral character-goals driven by desire (or some other mental cause other than moral judgments) is that the types of character goals which seem to be most obviously based (solely) on desire rather than moral judgments also seem to be unlikely candidates for guilt. One may desire to be the sort of person who is highly effective at work. One may
desire to be a humorous person. These goals may seem to be (non-moral) character goals, in that one aspires to be this kind of person. Still, if for whatever reason a person with such goals is quite unproductive at work and dull, then the emotional response need not be one of “guilt.” He may be quite angry at himself (or the world), or he may even be very saddened about these missed goals, but we have no reason to assume he would feel guilty about it. It is not the case that assenting to a missed/failed character goal always elicits guilt, and thus we cannot explain guilt in terms of character goals, even if it is possible to explain character goals without moral concepts. The type of character goals that would seem to elicit feelings of guilt (upon acts that fail to meet such a goal) would be those goals that more obviously employ moral concepts. One may have the goal of being “kind” or “honest,” but if one does feel guilty about failing to act in ways congruent with these goals then it would seem to be not simply that the person desired to be kind and honest, but rather that the person thinks he should be kind and honest. Thus, this attempt to define guilt in terms of failed character goals appears to be insufficient. The norm-expressivist, then, will not be able to address the circularity challenge (C3) in Nussbaum’s judgmentalism by reframing guilt in terms of non-moral character goals because such are either insufficient at best, or they imply moral judgments at worst (for the norm-expressivist).

Thus, I find that the most plausible approach for framing the propositional content of guilt in non-moral terms is, as discussed earlier (Section 4.3.1.1.2), to depict this content in terms of “the type of acts (or state) to meet a damaged communal relationship.” It is quite plausible to depict the propositional content of guilt (and thus define guilt) in terms as proposed by Baumeister, et al. Guilt is not the assent to having caused harm, or unwarranted harm, or breaking trust. It is just the assent to having done the kind of thing that tends to meet a damaged interpersonal relationship. If one considers “interpersonal relationships” to be a goal, then perhaps we could conceive of guilt in terms of this goal. Surely, Nussbaum would agree that interpersonal, communal, relationships have a eudaimonistic role in our lives. Preserving them is an important goal. Thus, the propositional content of guilt would seem to be describable, on Nussbaum’s theory of emotion, as the claim that “I have acted (or am in a state of the sort) to meet a damaged communal relationship.” The propositional claim is that the action (or state) is of the kind that tends to meet such a damaged relationship. Should Nussbaum’s theory be revised in its depiction of the propositional content of guilt in this way, then the norm-expressivist would find no problem in terms of circularity with respect to our concept of guilt. Since Nussbaum does not suggest that
emotional experiences are deliberate, self-reflexive judgments, her theory would not imply that experiences of guilt are caused by moral judgments – *even if* the definition of guilt required moral concepts. Still, it would be important that we can recognize the *concept of guilt* (as the reference of a moral judgment) without employing moral concepts. If we cannot, then her theory would still fail to meet C3. I hope to have shown that we can identify the emotion concept guilt, in keeping with her judgmentalism, by but modifying her description of the propositional content of guilt in (roughly) the same way we did with the attributional theories in meeting Sinnott-Armstrong’s critique.

With respect to defining the propositional content of anger, we can return to our consideration of depicting the emotion’s propositional content in direct relation to goals. In their studies on emotion typology, Tanja Wrankik and Klaus Scherer admit that “goals” are of use when describing only certain emotion types (248). Anger, though, is one of these emotions. Wrankik and Scherer state that, “Despite the nuances … anger experiences often share some general features …. Anger is elicited when an individual evaluates an *important goal as obstructed*” (248). Leonard Berkowitz argued that a necessary component to anger is the perception of an obstructed goal, as well as the perception that it is within the person’s power to overcome the obstruction (through, for example, lashing out) (271). For our purposes here, it seems we could infer from such findings and consider the propositional content of anger to be something like, “Another has obstructed my important goal,” or even “Another has obstructed an important goal and it is within my power to overcome this obstruction.” (After all, we may feel sad or frustrated, rather than angry, at an obstructed goal that we cannot overcome.) It need not be a “character” goal, nor a goal of any specific kind – just a genuine goal. Wrankik and Scherer illustrate, for example, that one can get angry at one’s boss for failing to secure funding or one can even get angry at oneself for failing to do so. Typically, as they explain it, one tends to become angry at another (rather than oneself), and typically one tends to be angry about things that were (seemingly) intentional or at least easily preventable, but even these stipulations need not always pertain (248). The important overarching criterion, to defining anger, is that the person who experiences it is responding to the judgment that an important goal was obstructed and that it is within one’s power to change this. After all, if one’s goal was obstructed and one did not find that it is within one’s power to change things (such as getting caught by a police officer in the middle of a crime, or knowing that an opportunity is lost with nothing to do about it), then the response is more likely to be one of fear or sadness rather than anger. We need not pinpoint a specific type of goal whose
obstruction elicits anger, for Wranik and Scherer note that “Recent research has shown that high trait and low trait anger individuals have different appraisals in reaction to the same situation” (257.) If one is not at all angry, though, then this definition would of course imply that either the harm did not affect a genuine goal (being something that one was aiming at and working toward) or the impact of the harm was not perceived as a real obstruction.

So long as the concept of “goals” does not imply moral concepts, the norm-expressivist could seemingly reframe the propositional content of anger into “Another has obstructed my important goal and it is within my power to overcome this obstruction.” Indeed, this depiction of the propositional content would recognize the eudaimonistic nature of the emotion (being that anger is about important goals, about highly desired/values goals in other words) as well as the object of the emotion (being but the general sense of a goal itself). If this explanation of the propositional content of anger is to work, and if we are to adopt the propositional content for guilt as suggested above, then norm-expressivism would entail that the moral claim that “X is wrong” means “It makes sense to judge that I have failed to meet the character-goals I desire whenever I do X, and it makes sense to judge that an important goal is obstructed and that I can do something about it whenever another does X.” In this case, we seem to meet both C1 and C3. We have identified the propositional content by which we can define both anger and guilt, and we have done so without employing moral concepts.

No one would deny that persons can get angry when another interferes with our ability to reach a goal, and this is perhaps especially true when one has previously maintained the goal or made progress toward it. Still, it may be retorted that the judgment that another has interfered with one’s attainment of a goal is not a sufficient condition for anger. Imagine that Jane’s parents own a restaurant and she runs the restaurant, and the rule is that she is supposed to pay for her own meals (at some discount) just as do the other employees. If an employee catches Jane taking food without paying and the employee tells her that she plans to report this behavior (seemingly obstructing Jane’s goal, since the employee just may report it), it is quite conceivable (to me anyhow) that Jane may feel some emotion other than anger. She may feel more ashamed than angry. She may even only feel ashamed and not the least bit angry, even if it was in her power to overcome this obstruction. (Perhaps Jane could just fire her since she manages the place, and perhaps she could even obstruct the report.) Even though the employee intentionally obstructed Jane’s attainment of a goal and even if it is in Jane’s power to overcome this obstruction, she need not be angry. We might
even explain that, in such a case, Jane’s lack of anger is likely *because* she knows that the employee is quite within her rights to call her out on this to begin with. One might argue that all such instances, in which one graciously accepts defeat, can only prove that anger is not the assent to merely the appearance of an obstructed goal. One might even go so far as to argue that anger is likely to be one’s assent to the propositional content that, “A just, or morally permissible goal, was obstructed.” Clearly, though, this would violate C3.

The response to such a criticism, though, is that one’s lack of anger in such cases is explainable in terms of higher order goals. If Jane is not angry at being caught stealing food in this case, then this may (for example) be due to the fact that she ultimately values being a rule-abiding citizen (or a respectful child to her parents anyhow), even if she doesn’t always make choices (in the moment) that reflect this higher order value. Of course, the critic may argue that higher order goals must then be implied by all goals, and one may argue that these higher order goals themselves imply moral concepts about the “good” at some point. To this, though, it seems possible to respond (as Gibbard likely would) that higher order goals need not entail moral concepts, for these higher order goals are grounded in higher order norms which themselves need only be grounded upon what Gibbard calls “loosely intuitionistic” and “pragmatic” considerations (*Wise* 282-284). Gibbard certainly makes the case that we can endorse claims and higher order norms without the use of moral concepts, for norm-expressivism holds that we are taken to accept higher norms which then inform the way in which we choose the moral norms we accept. The norm-expressivist would seem unlikely, then, to be concerned that all goals imply moral judgments via higher order goals.

Thus, I take it that the norm-expressivist may find friendly amendments to Nussbaum’s theory of emotion with respect to the propositional content of anger as well as guilt. In modifying the propositional content of anger to “Another has obstructed my important goal and it is within my power to overcome this obstruction,” and in modifying the propositional content of guilt to “I have acted (or am in a state) of the sort to meet a damaged communal relationship,” the norm-expressivist may find that Nussbaum’s judgmentalism can meet both C1 (in distinguishing these moral relationships from one another) and C3 (by addressing the circularity challenge). Thus, while Nussbaum defines the propositional content of guilt and anger in a way that would render norm-expressivism circular, there are other plausible ways to define the propositional content of these emotions in a way which quite meets C3 (the circularity challenge) and C1 (identifying the emotion types). One may
take issue with Nussbaum’s claim that a judgment alone is identical to an emotion. One may argue that it is possible for someone to agree with a judgment and yet not experience an emotional response to the content. To this, Nussbaum responds that, in these cases, no real agreement has in fact occurred. An emotive judgment, in her theory, just is a caring one, and thus assenting to the judgment will always result in the experience of an emotion (whatever this may involve).

There is nothing inconsistent between Nussbaum’s theory and the norm-expressivist’s other conditions either. Recall that C4 requires a theory of emotion to explain how specific types of emotions motivate action and C2 requires a theory of emotion to provide a framework capable of explaining how some emotions (such as anger and guilt) tend to be stronger motivators of action than are other moral emotions. Since thoughts or judgments can motivate actions and some perhaps more than others, it is conceivable that moral emotions could work in this way under her theory. Thus, her theory is not inconsistent with C4 or C2. Nonetheless, we should recognize that Nussbaum’s theory provides more of a humanistic analysis of emotion typology (in terms of the emotion’s intentionality) than an account of any empirically realized causes and their impact upon action tendencies. Thus, her theory does not aim to explain emotions’ ability to motivate action. Her point is not to provide a scientific account of the emotion in relation to the body and its actions, but rather a humanistic understanding of the way in which emotions evaluate objects with regard to human flourishing. She largely rejects reductionist attempts to define emotion in empirical terms, for she finds such attempts to neglect the intentionality of emotion (119). There is nothing inconsistent about her theory and C4; emotions, as judgments, can motivate actions (or action tendencies anyhow) in her view. While it is true that we act on many judgments, and perhaps especially upon evaluative judgments, her theory leaves it unclear as to what causes the action, though. (We can, after all, have evaluative thought without acting on it.) While her theory is not inconsistent with these conditions, I do not find her theory to be particularly strong in accounting for the special action-guidingness of anger and guilt per C2. Nor does her theory offer a particularly strong address of C4. Thus, while her theory is consistent with C2 and C4, it provides a weak account of them.

On other hand, her theory would seem to address well C5, C6 and C7. C5 requires that the theory of emotion explain how emotions can be caused by the acceptance of norms that we are more or less aware of at the time. Just as we may be more or less aware of the beliefs that motivate many of our judgments, emotions as judgments may also be caused by the acceptance of norms which are more or less recognizable. C6 re-
quires that the theory of emotion provide a framework that explains how broadly moral sentiments shape guilt and anger. Under her theory, this could be accounted for in that broad moral emotions may help cultivate the narrow moral emotions in the same way that general beliefs (or judgments) help to cultivate more specific ones. C7 requires that the theory of emotion explain how guilt and anger respond to considerations regarding their rationality (their consistency with other norms, grounding reasons, and plausibility); on Nussbaum’s account, adult human emotions (such as moral anger and guilt) are judgments that respond to such rational considerations, just as do other types of judgments after all.

The norm-expressivist is able to, then, utilize Nussbaum’s judgmentalist theory of emotion so long as the propositional content of guilt and anger are amended in non-moral terms. Interestingly, the ability of her theory to be consistent with norm-expressivism will be equal to the ability of attributional theories, which Gibbard himself recommended, even though Gibbard dismissed the ability of judgmentalist views to work due to (what he took to be) ambiguity around the nature of the emotive judgment and circularity concerns that such views would seem to pose for norm-expressivism (Wise 128-131). Since norm-expressivism would require attributional theories to provide for a non-moral description of moral emotion concepts (in order for us to recognize genuine moral judgments, as Sinnott-Armstrong points out), the propositional content of these emotions under Nussbaum’s view could adopt a similar depiction of moral emotion concepts as well. Nonetheless, while Nussbaum’s theory is consistent with norm-expressivism, her theory does carry weaknesses in addressing C2 and C4. As we continue on, it may turn out that other theories of emotion offer the norm-expressivist even stronger accounts of the action-guidingness of anger and guilt.

4.5 Dimensional Appraisal Theories of Emotion in Psychology (Emotion as Bodily Response to Cognition)

Psychology offers cognitivist theories of emotion which are, in some ways, quite similar to Nussbaum’s judgmentalism above. Dimensional appraisal theories of emotion propose structures to distinguish between distinct types of emotions based upon their cognitive evaluations. A difference between Nussbaum’s judgmentalism and dimensional appraisal theories is that, while she claims that emotions are these cognitive evaluations (or assents to propositional content), dimensional appraisal theorists argue that emotions are experiences (and even action tendencies as
well), which are caused by these cognitive evaluations (Lazarus *Emotion and Appraisal* 121). Dimensional appraisal theories aim to identify the organizational structure of emotive evaluations, thereby providing a framework by which to understand the nature of the cognitive content that distinguish emotion types from one another. Although Lazarus argues that emotions actually include their cognitive causes in the emotions themselves, as we will see, he shares the view (with other dimensional appraisal theorists as we will see) that emotions are caused by their distinguishing cognitions. Nussbaum and dimensional appraisal theories share the view, though, that we can distinguish emotion types from one another based upon the distinct content of their cognitive evaluations. Sometimes referring to this emotion-cognition as “appraisals” and sometimes referring to emotion-cognition as “construals,” dimensional appraisal theories of emotion propose a typology of emotion based upon the structure of these emotive judgments (Lazarus 822; Ortony et al. 4, 34).

Andrew Ortony, Gerald Clore and Allan Collins argue that emotions “issue from cognitive interpretations imposed on external reality” (4). It is the way that an object is construed, or appraised, that determines the emotion. They explain, for example, that the response of the winning team and the losing team to a major game is a response to the same event, even though their emotional responses are likely to differ. It is the interpretations of the event, “their construals of the event that are different” and, thus, it is the construals that determine their differing emotions in response to the game (4). Lazarus also argues that emotion requires a form of cognitive evaluation, claiming that emotion “involves an appreciation of a particular harm or benefit in the relationship with the environment, with its manifold implications for well-being, action, and coping” (*Emotion and Adaptation* 121). These theories propose that it is the nature of cognitive appraisals that properly define particular types of emotion, distinguishing them from one another. Thus, we can consider whether these theories can explicate moral emotions as forms of cognition that are compatible with norm-expressivism.

While Nussbaum appreciates the dimensional appraisal theorist’s proposed structure of an emotion typology in accordance with identifiable evaluations (107), Ortony et al. and Lazarus propose that emotions are actually bodily responses to cognitive appraisals; emotions, then, include affective change. In this way, then, Ortony et al. and Lazarus propose hybrid cognitivist theories of emotion. As we will see, Ortony et al. and Lazarus propose dimensional appraisal theories that include bodily feelings in the constitution of emotion and Lazarus includes behavioral tendencies in the emotion as well. Still, while these theories propose
structures to identify the typology of emotion, they maintain (unlike Nussbaum) that emotion types can be identified according to specific dimensions of appraisal.

4.5.1 Dimensional Appraisal Theory: Ortony, Clore and Collins

Ortony et al. aim to present a typology of emotion derived from the cognitive structure of emotive evaluations. Emotions, they claim, “arise from cognition” (4). Emotions are here taken to be “valenced reactions to events, agents, or objects with their particular nature being determined by the way in which the eliciting situation is construed” (Ortony, et al. 13). They do not dispute that emotions are physiological reactions to their stimuli. Still, they maintain that even if emotion types are (eventually) shown to be distinct physiological responses, cognitive construals are nonetheless necessary and sufficient causes for them. As they put it:

we do not feel obliged to take sides in the debate about whether there is a unique pattern of physiological activity (of the sympathetic nervous system) associated with each specific emotion, because patterns of physiological activity are not directly relevant to the cognitive antecedents of emotion …. To believe … that the importance of the cognitive determinations of emotion is in any way contingent on the final resolution of this issue … is to misunderstand the nature of the cognitive claim. (12)

Though they admit that emotions include experience, they maintain that construals are the necessary and sufficient “eliciting conditions for a particular emotion” (3). They explain that, “To say that emotions arise from cognition [or “construals”] is to say that they are determined by the structure, content, and organization of knowledge representations and the processes that operate on them” (4). On their view, then, there is a structure to the content of one’s emotive construals that serves to explain these evaluations that elicit specific emotion types. Though they do not deny that we can also explicate emotion types in light of cultural worldviews, they look for a way in which to identify emotion types in a more universal way – in a way that is not “local to a specific time or cultural group” (ix). Thus, distinct from attributional theories (which depict emotion as the labeling of one’s state in keeping with cultural rules) and distinct from Nussbaum’s judgmentalism (which depicts emotion as an individualistic assent to propositional content), Ortony et al. aim to specify the “psycho-logical” structure of emotion types (x).
Their goal, then, is to provide an analysis of the logical structure behind specific emotion types. They explain their interest as follows:

Our approach will be concerned more or less exclusively with trying to characterize the differences between emotions in terms of the different kinds of cognitions we take to be responsible for them. Taking the perspective of empirical psychology and cognitive science, we start with the assumption that emotions arise as a result of the way in which the situations that initiate them are construed by the experiencer. (1)

They propose that empirical science and cognitive science suggest that emotions have a logical structure to be discovered, and (in their view) this structure is what properly characterizes an emotion in terms of its typology. Thus, while Clore argues that emotion itself (as a bodily experience) will always be a conscious experience (“Never Unconscious” 286-287), he explains that we can be wrong about identifying our own emotions whenever we fail to realize the cognitive evaluations that elicited the experience. As an example, he notes that a professor’s tightening chest may be understood to be “fear” (instead of heart attacks) only when one later realizes that it subsided upon her making tenure (Clore “Never Unconscious” 286-287). One cannot always recognize the cognition which causes a particular emotional experience, but dimensional appraisal theorists suggest that the proper identification of an emotion (whenever it is possible to identify it) would require an understanding of the cognitive cause of the experience. This cognition might be inferred from one’s changing response to stimuli, for these changing responses are taken to imply distinct construals of the situation. The typology of an emotion, then, is determined by the recognition of its logically implied construals. Thus, Ortony et al.’s dimensional appraisal theory proposes that emotion concepts can be depicted in terms of the cognitive appraisals that cause them.

As with Nussbaum’s judgmentalism, this theory conceives of “cognition” in a broad sense. Nothing in this theory implies that the cognitive cause (the construal) consists of talking to oneself, or consciously recognizing in a reportable way the nature of an object. Ortony et al. do not presume that the cognitive construals which elicit emotion (and thus define emotion types) need take place through psychological mechanisms such that the person (experiencing the emotion) is able to identify one’s emotion at the time. The appraisals may take place below the threshold of such awareness. As they put it, “the cognitive claim is that emotions are reactions to (or representations of) the personal meaning
and significance of situations, not that emotions originate in the cerebral cortex” (Clore and Ortony 42). Mikko Salmela draws upon Fred Dretske’s argument for semantic structures in animals in order to justify his claim that cognitive evaluations (and even semantic structures) are necessary to an organism’s appraisal of an object even when this appraisal takes place through sub-cortical (pre-linguistic) pathways in the brain (9). Salmela argues that, per Dretske, the “specialized response patterns compel us to assign them internal states with a specific semantic content. After all, when an internal state functions in a cognitive role, it qualifies as a real cognition” (9). Ortony et al.’s position is not that emotion types are caused by conscious or linguistically formulated appraisals. It is rather that we can realize an articulate structure to the appraisals which cause distinct emotion types. Specific types of emotions are caused by these describable appraisals regardless of the psychic mechanisms by which these appraisals are represented in the emotion itself.

We ought not take the articulate structure of such cognitive causes to undermine the fact that, in Ortony et al.’s view, these emotions are caused by cognitive acts. Ortony et al. are not proposing that these cognitive construals are superimposed, or attributed, to emotions in light of the way in which the emotions function. They emphasize that these cognitive interpretations must actually take place and thus cause distinct emotion types. Clore takes emotion proper to require a sort of cognition unavailable to other animals and even humans in some cases that would otherwise appear to be an emotion. He recommends that we not confuse those more immediate sorts of experiences (which may seem like emotions) for the types of genuine emotions addressed in his theory. He acknowledges that some “situations may require very little processing,” though he claims that “the resulting physiological, experiential, or behavioral reactions [in these cases] often do not constitute emotions as such” (Clore “Require Cognition” 190). Clore proposes that immediate “emotion-like reactions” which result from one’s physiology (such as facial expressions and freezing in response, which he calls a “startle reflex,” in response to someone jumping out and yelling “Booh!” and even more specific innate reactions which draw upon complex recognitions in the world (like the chicken’s fierce avoidance of chicken hawks are better understood as instincts or some phenomenon other than emotion (“Require Cognition”182, 183). Whatever information processing is required for such responses, it is not the kind of cognition sufficient for emotion proper. According to this line of thought, then it would seem that the more immediate emotional responses (such as Gendler’s “alief” response which could occur in a rather modular way) would not qualify for emo-
tion proper (genuine fear) at all. It is not enough to identify an emotion that the body has in response to certain stimuli, for without a certain sort of appraisal the emotion does not exist.

Given that Ortony et al. admit (above) that appraisals exist whenever an organism represents the information, it is unclear why such responses discussed here fall short of a true emotion in their view. While they seem to admit (above) that emotive appraisals need not include higher level thought processes, it appears here that a functional response to certain stimuli would not constitute the relevant cognition in Clore’s view. Emotion proper would seem to require a higher-level, slower thought process distinct from the modular sort at least, even if they admit that it occurs below the threshold of one’s full awareness. Thus, while moral emotions need not be caused by conscious judgments, they are taken to be caused by fairly sophisticated cognitive acts along specific logical dimensions. The typology of distinct moral emotions would then be determined by these cognitive antecedents.

Ortony and Clore explain that their dimensional appraisal theory “proposes three broad classes of emotions distinguished in terms of the cognitive focus involved” (“Require Cognition” 186). What all emotions have in common is that they are either positive or negative appraisals of their objects, and what distinguishes them is how the positive/negative affect manifests and what the objects of the affective response turn out to be. Emotions are taken to fall under cognitive structures of either 1) pleasure/displeasure with events, 2) approval/disapproval of agents, or 3) like/dislike of things (“Require Cognition” 186).

According to Ortony et al.’s typology of emotions, “shame” and “reproach” are emotions caused by the “approval/disapproval of [an] agent”. In cases of shame, one disapproves of actions caused by oneself and in cases of reproach one disapproves of actions caused by another. Thus, shame and reproach are said to be “Agent-based emotions.” Ortony et al. explain the nature of Agent-based emotions and Attribution emotions as follows:

Agent-based emotions are Attribution-of responsibility or, simply, the Attribution emotions. When the agent is judged to have done something praiseworthy, the person experiencing the emotion is inclined to approve of the agent’s action, and when the

---

24 Since startle responses to “Booh!” would be quite analogous to the rat’s freezing at the initial sound of a buzzer, and since Clore takes such startle responses to “Booh!” (and even more complicated responses to compound stimuli) not to be a genuine case of emotion, I take it that modular responses are non-emotions in Clore’s view.
Thus, shame and reproach are taken to be attributions of blameworthiness (of one’s own blameworthiness in the case of shame, and of another’s blameworthiness in the case of reproach). These judgments of blameworthiness are taken to result from the judgment that the agent has failed to meet the experiencer’s “standards.” Ortony et al. explain that “in the case of the Attribution emotions the central variable is praiseworthiness (which, in its technical sense, we take as including blameworthiness) computed in terms of the standards that are invoked in evaluating the agent’s action” (134). Given that Ortony et al. claim both that “Standards concern the states of affairs that one believes ought to obtain” and that “one might call these moral or quasi-moral standards,” it is clear that the cognitive construals that cause both shame and reproach require the use of moral concepts (even moral judgments) on this view (45). Because reproach is treated by Ortony et al. as the counter-part to shame, we might suppose that reproach stands for “disdain” under Gibbard’s view (noted in Chapter Two) (69).

In this case, then, Ortony et al.’s dimensional appraisal theory would hold that both shame and disdain require the use of at least moral concepts (if not moral judgments) in their causal antecedents. The implication on Ortony et al.’s view, here, is not just that these broadly moral emotions (as Gibbard regards them) can be caused by moral judgments (or at least by the use of moral concepts), but that they must be caused by such judgments. As Clore says, “A common assumption is that the essence of emotion lies in the physiology, phenomenology, or behavior of emotion. We suggest that these are constituents of emotion only when they are tied to (possibly unconscious) cognitive appraisals of situations …” (Clore “Where Does Anger Dwell?” 59).

Ortony et al. discuss anger, though they do not specifically analyze guilt. In that “remorse” is taken to be the counter-part to anger in their view though, then perhaps we should take the norm-expressivist’s use of “guilt” to be captured in Ortony et al.’s treatment of “remorse” (69). In this case, Ortony et al. would propose that anger and guilt (or “remorse” as they call it) also require moral judgments. Anger and remorse are taken by them to be “compound emotions” which combine “Attribution emotions with Well-being emotions” (146). Whereas Attribution emotions (as judgments of an agent’s action) are appraisals of blameworthiness in their view, well-being emotions (as judgments about the consequences of an event) are taken to be appraisals of the desirability of events. Remorse, they claim, is the combination of “shame” (the
judgment of one’s own blameworthiness) and “distress” (distress being “displeased about an undesirable event”) (146). Likewise, anger is taken to be reproach and distress (146). Thus, Ortony et al.’s dimensional appraisal theory would hold that anger and guilt require moral concepts to be used in their causal antecedent conditions (their cognitive appraisals), just as do shame and disdain. Certainly, then, our ability to understand the narrow emotion concepts of guilt and anger on the norm-expressivist’s view (in order to make and recognize moral judgments) would, if we adopt Ortony et al.’s theory of emotion in its entirety, require our use of moral concepts.

As it is, then, Ortony et al.’s dimensional appraisal theory would entail that the use of moral concepts (even blameworthiness in particular) is necessary to the causal antecedents of moral emotions, and our ability to conceive of anger and guilt concepts would also require the use of such moral concepts.

4.5.2 Dimensional Appraisal Theory: Lazarus

Like Ortony and Clore, Lazarus includes bodily change in emotion. Lazarus makes “physiological change a hallmark of emotion” in order to “to distinguish emotional processes from non-emotional ones such as cold cognitions, homeostatic processes, and reflex or automatized adaptations” (“Progress” 822). Unlike Ortony and Clore, though, Lazarus adds that “each emotion involves its own action tendency which can be concealed or overridden by the process of coping” (“Progress” 822). Like Ortony et al., Lazarus argues that the specific types of emotion are caused by distinguishing cognitive appraisals, though (unlike Ortony et al.) Lazarus maintains that the emotion itself is then a combination of its cognitive cause, the physiological state that accompanies it, and the behavioral tendencies that the emotion carries (“Cognition and Moral Emotions” 353-354). Unlike Ortony and Clore, Lazarus argues that emotions are not but caused by cognition, for emotions themselves include their cognitive causes; as he put it, “cognition, which is causal, also carries into the response state” (“Progress” 823). Like Ortony and Clore, though, Lazarus argues that the typology of emotion is provided for through an analysis of the cognitive appraisals that differentiate distinct types of emotion.

Though emotions include physiological states and action tendencies in Lazarus’s view, he proposes that each emotion type is “defined by its core relational theme and pattern of appraisal.” The “core relational theme” of an emotion is, what he calls, the “molar appraisal” (“Progress” 822). He explains that “A core relational theme is simply the central
(hence core) relational harm or benefit in adaptational encounters that underlies each specific kind of emotion” (*Emotion and Adaptation* 121). Emotion types also have distinct appraisal patterns, along both “primary” and “secondary” dimensions in Lazarus’s view. Lazarus explains that primary appraisal “concerns whether something of relevance to the person’s well-being has occurred”; more specifically, primary appraisals concern the “goal relevance” of the stimuli, the “goal congruence or incongruence” of the stimuli, and “goal content, or type of ego involvement” (“Progress” 827; *Emotion and Adaptation* 133). Secondary appraisal “concerns coping options – that is, whether any given action might prevent harm, ameliorate it, or produce additional harm or benefit” (*Emotion and Adaptation* 133). Lazarus explains that, to fully identify individual types of emotion, though, we need to identify these “molecular appraisals,” which include both the primary appraisals about goal relevance and congruence and the secondary appraisals which specifically concern “blame or credit,” “coping potential,” and “future expectations” (“Progress” 827). Lazarus argues that “to fully understand the emotion process requires knowledge of the molecular appraisals by means of which the individual arrives at the core relational theme that underlies [(defines)] each emotion” (“Universals” 164).

Like Ortony et al., Lazarus defines moral emotions in light of his dimensional appraisal theory. Lazarus proposes that the core relational theme for guilt is “having transgressed a moral imperative” (“Universals” 164; “Progress” 826). Thus, guilt consists of the cognitive appraisal that one has transgressed a moral imperative. Such an appraisal need not be a moral judgment in precisely the way in which a norm-expressivist regards a moral judgment (per Chapter Two). That is, one needn’t find that a judgment of blameworthiness is something that is supported by grounding reasons, checks for consistency and plausibility, nor that one is able to defend such judgments of blameworthiness if only uncensored; this, after all (per Chapter Two), is implied by a moral judgment in the norm-expressivist’s view. Still, such appraisals would surely include moral *concepts*. Lazarus does consider “several alternative emphases” for the core relational theme of guilt, noting the following depictions of guilt:

(1) In the psychoanalytic view guilt feelings are conceived of as a reaction to unacceptable impulses in oneself and traced to hostile and sexual impulses in childhood that are repressed as a result of fear of punishment and loss of parental love; (2) guilt feelings are said to be based on prosocial feelings and empathic concern over the distress of others, which evolve developmental-
ly into wanting to do well by others (see Hoffman, 1982a\textsuperscript{25}); and (3) guilt feelings are primarily cognitive and not present in very early childhood but are aroused when the children learn to perceive and understand the social significance of violations of standards of conduct (see Wicklund, 1975\textsuperscript{26}). \textit{(Emotion and Adaptation 241)}

Lazarus finds his proposed core relational theme for guilt (being a personal transgression) to be consistent with all of these alternative descriptions of guilt, claiming that “these views of guilt [are] overlapping.” Though such depictions of guilt highlight (or propose) different “dynam- 
ics and origins” of guilt, Lazarus claims that they all always involve “transgressing a moral imperative” \textit{(Emotion and Adaptation 241)}. Ultimately, then, one must grasp such concepts as “transgression” and “moral imperative” in order to experience guilt on his view. Whether the behavior by oneself which elicits guilt is appraised to be unacceptable, or contrary to doing well by others, or a socially significant violation of conduct, Lazarus finds that it always amounts to an appraisal of transgressing a moral imperative. Thus, there must be a semantic structure allowing one to recognize such violations even if one need not formulate such linguistically describable recognitions in the form of “talking to oneself” in the course of recognizing them.

Lazarus proposes that the core relational theme for anger is “A demeaning offence against me and mine” or an “unwarranted offence” (“Universals” 164; “Progress” 829). He explains anger, in light of these primary appraisals, as follows:

Anger…depends on an appraisal that one’s ego identity, the active goal content, is at stake, which also implies goal relevance. When this identity has been threatened or harmed, there is goal incongruence in what is deemed to be an unfair slight or insult (c.f. Aristotle). In anger, blame is also necessary, and it depends on the attribution that someone is accountable and has full control over the demeaning action” (“Progress” 828).

Anger, too, entails that one attributes “blame” to another. Anger, like guilt, employs moral concepts on Lazarus’s view. Guilt would, then,


depend upon an appraisal that one’s ego identity has been threatened by one’s own moral transgression and that oneself is to blame (“Progress” 827). Like Ortony and Clore, then, Lazarus’ dimensional appraisal theory would seem to purport that emotions are caused by (and even constituted by) cognitive evaluations and that moral emotions are caused by moral judgments. Though Lazarus does not deny that judgments can change to varying degrees in the course of the same emotion, he maintains that emotions are caused by and constituted by their defining judgments.

As with Ortony and Clore, Lazarus admits that the cognition that causes a distinct emotion type may occur in subconscious ways. He does not maintain that this is the case, but that it may turn out to be so. Lazarus broadens cognition even more, though. He states that, “Although it conflicts with traditional usage among cognitive psychologists, a view of the evaluative process of appraisal as often nonvolitional and unconscious may be emerging” (“From Psychological Stress” 15). Unlike Ortony and Clore, Lazarus does not reject the possibility that the cognitive causal antecedents of emotion proper are at times even *automatic* (modular) responses to stimuli. If moral transgressions (and even harms to one’s ego ideal) are the types of things that can somehow be recognized through such unconscious and modular means, then Lazarus’ theory may allow for moral emotions to represent the semantic content of a moral judgment without being a moral judgment. Still, so long as the moral emotion concept is identified (i.e. *conceived of*) in terms of moral concepts (like transgression), one could not make a moral judgment as purported by norm-expressivism without employing moral concepts.

Lazarus himself though does not argue that moral emotion appraisals can or cannot be made in such unconscious and particularly modular ways. When he speaks of the moral emotions in terms of moral transgressions, it would seem that he intends this to mean a recognition that takes place through cognitive mechanisms that allow for such sophisticated judgments as that of blameworthiness, given that he never specifically proposes that such recognitions can take place in automatic or low-level forms of cognition. Rather, he tends to avoid such implications. Lazarus’ dimensional appraisal theory was intended by him to propose a typology of adult human emotions. His reason for doing so highlights his general commitment to the view that dimensional appraisals are forms of higher level cognition. Animals and young children do not have the same cognitive capacities as do adult humans, and thus he notes that his taxonomy of emotion is not suited for an understanding of emotion-like states in other animals or young children. He notes, for example, that while young children may have responses that appear to be *anger* in response to certain stimuli (such as having their heads re-
strained), the emotion process is not the same for human infants or young children who seemingly have a different (if even existent) sense of warrant in their own experiences of anger ( “Progress” 829). This being so, his dimensional appraisal theory would then not be intended for identifying other non-human emotions. Likewise, it would seem questionable whether he considered emotive appraisals to explicate guilt and anger through such pre-linguistic, and unconscious, forms of reasoning. Others, as we will see in Chapter Five, modify his theory to support an understanding of emotion taxonomy for non-human animals and infants as well, but that was not his intent. Lazarus takes the appraisals, which cause specific types of emotions, to be actual cognitive evaluations of stimuli and this would imply an actual ability to conceive of the concepts which compose the dimensional appraisals.

4.5.3 Would Dimensional Appraisal Theories of Emotion Suffice for Norm-Expressivism?

Though the dimensional appraisal theories of emotion considered here explain the moral emotions as appraisals of blameworthiness, we might wonder whether there is another way in which moral emotions could be explained so that they do not require moral judgments. Because of Gibbard’s circularity concern (noted in C3), it is important for the norm-expressivist to find a theory of emotion that does not account for anger and guilt in terms of wrongdoing (or blameworthiness). If a cognitive appraisal employs the concept of wrongness in some form, then this appraisal is a judgment of wrongdoing (or blameworthiness of sort). If I make the appraisal that “I am blameworthy for X,” then I have made a moral judgment of some sort even if this judgment lacks some features of a moral judgment as depicted by Gibbard. (In this case, we may have cause to question whether Gibbard has the features of a moral judgment totally right.) If moral emotions must be caused by either moral judgments or appraisals that at least employ such moral concepts as both Ortony et al. and Lazarus hold, then the norm-expressivist cannot explain the meaning of moral judgments in terms of moral emotions. If the dimensional appraisal theorist is right, and moral emotions are experiences which must be caused by (or even co-occur with) moral judgments, then norm-expressivism would hold that the moral claim “X is wrong” means “It makes sense to have an experience Y (and action tendencies Z, for Lazarus) in response to (or with) the judgment that X is wrong.” Here, the judgment about the wrongness of X is being used to explain the meaning of the judgment that X is wrong. As with Nussbaum’s theory, then these dimensional appraisal theories would clearly violate C3 (by
making norm-expressivism circular) so long as we accept the theory’s definition.

Both of these dimensional appraisals theories hold that the person experiencing an emotion must cognize about specific concepts in order to make the distinct appraisal. If we could find a way of describing the appraisals which define moral emotions without implying that these concepts are equivalent to the concept of wrongdoing, then dimensional appraisal theories need not violate C3. Reframing a dimensional appraisal theory in any conclusive way so as to accomplish this is no small task, though it does seem plausible nonetheless.

One might try to solve the norm-expressivist’s circularity problem with dimensional appraisal theories by reframing the molecular (secondary) appraisals that make up the core relational themes. Clearly, this would be a departure from Lazarus’ own view (since, as we saw above, he finds alternative descriptions to imply an appraisal of one’s moral transgression), but we might consider it further of course. In his review of the different ways in which appraisal theories depict the features of anger, Berkowitz reports that only two of the four theories he reviews (those by Ortony et al. and Lazarus) take anger to imply a judgment of blameworthiness (271). He reports that other theories (as proposed by Roseman and Scherer) agree that anger implies the judgment of an interference with an important goal and that another is “responsible” for this obstruction of the goal, though these theories do not depict anger as resulting from the necessary judgment that “someone or some thing must be blamed” (271). Berkowitz explains that there is some debate around the necessity for “blame” in anger appraisals (and thus, conceivably in my view guilt as well) due to the “problem of causal direction” (275). He notes that, “We cannot unequivocally determine the causal direction of the connection between anger and blame in many of the relevant studies. The blaming may have been, in some cases at least … a consequent rather than an antecedent of anger arousal” (275). If this is so, then that would certainly be encouraging news to the norm-expressivist. In this case, appraisal theories must offer a definition and explanation of anger independent of this moral concept (of blame) and this will help to addresses C3.

To achieve this, we would need to explain what such a judgment of “responsibility” is about if not blameworthiness. For example, one

might suggest that the dimensional appraisal theorist could try saying that it is not “blameworthiness” that comprises secondary appraisal (since this molecular appraisal seems to particularly bring the moral judgment into the emotion), but rather this secondary appraisal is about the “directness of the causal locus.” Perhaps this appraisal of blameworthiness could be reframed as an appraisal about non-moral concepts—such as the cause of the event, including the intentionality of the agent’s action that caused the event and the agent’s effort that went into avoiding the undesirable event. Such criteria may explain one’s appraisal of responsibility without implying a judgment of blameworthiness. While Ortony et al.’s and Lazarus’s theories may be among the better known appraisal theories, there are certainly other appraisal theories for the norm-expressivist to consider.\footnote{Berkowitz notes theories of emotion that include appraisal features as put forth by Ira J. Roseman and Klaus R. Scherer. See Berkowitz’s bibliography for an extensive list of references for each of these theories.} I will not offer a comprehensive or conclusive analysis of competing appraisal theories of emotion, though I hope to establish the possibility that such theories may in fact offer definitions of anger and guilt without the use of moral concepts. If they can describe anger in terms of another’s “responsibility” instead of “blameworthiness,” then this may be fruitful for the norm-expressivist. As I see it, it seems quite possible to explain anger in such a way. If, for example, I find that the person whose action interferes with my relevant goal intended such interference, or that the interference with my goal was caused by the agent’s thoughtlessness, then I find myself “angry.”

For example, suppose that I am angry at my neighbor for driving on my lawn. Must my anger result from my finding him “blameworthy”? We might try to describe these antecedent evaluations in another way. I find that his driving on my lawn is relevant to my goal of having a nice lawn, that his driving on it is incongruent with my goal (since it damaged the lawn), and that he is the one responsible for it (in that he is the one who did it and that he did so either intentionally or with little concern). The norm-expressivist might hope that in re-framing Lazarus’s secondary appraisals in this way, one could avoid the circularity challenge. Instead of saying that secondary appraisals include judgments of “blameworthiness,” perhaps the norm-expressivist could just say that the dimensional appraisal translates more properly (on the topic of “blameworthiness”) into appraisals of “directness of the causal locus as determined by the agent’s intention and effort.”

Likewise, we could try reframing Ortony et al.’s appraisals of moral emotions (guilt/remorse, anger, shame and disdain/reproach) in terms that pertain to the “directness of the causal locus as determined by...
the agent’s intention and effort” as well. If the person’s driving on my lawn is found to negatively impact my well-being (thus distress is present), and if the person is found to have intentionally or thoughtlessly done so (thus a modified sort of reproach is present), then we could find that this constitutes appraisals which cause anger. If the person’s driving on my lawn is found to negatively impact my well-being, but this person is thought to be incapable of controlling it, then perhaps these appraisals would cause disdain/reproach. If I am the one who intentionally or thoughtlessly drove on another’s lawn, and this interferes with my goal to preserve communal/interpersonal relationships, then these appraisals could be said to cause my feelings of guilt. If I drove on the lawn, committing this act that is incongruent with my goals for myself, but I did so because I couldn’t help it, then these appraisals would cause my shame.

The dimensional appraisal theories as proposed by Ortony et al. and Lazarus claim that there is a distinct cognitive appraisal that defines each type of emotion, and moral emotions are defined by moral judgments. These theories do not claim to superimpose such judgments upon these emotions, but they rather claim to discover the actual cognitive structure of the emotions. They claim that such interpretations of the stimuli take place. Thus, moral emotions would need to (if they are correct) actually be constituted by such moral judgments or at least evaluations that employ moral concepts. Thus, these dimensional appraisal theories, taken in their entirety, would render norm-expressivism a circular theory (violating C3) through their depiction of the moral emotions. Should such modified characterizations to these appraisals work, though, then dimensional appraisal theories may satisfy C3 after all. I maintain, then, that dimensional appraisal theories are consistent with C3 in this conditional way.

As with Nussbaum’s judgmentalism, Ortony et al.’s dimensional appraisal theory struggles with C2 and C4. Recall that C4 requires a theory of emotion to explain how specific types of emotions motivate action and that C2 requires a theory of emotion to provide a framework capable of explaining how some emotions (such as anger and guilt) tend to be stronger motivators of action than are other moral emotions. Ortony et al.’s theory is consistent with these conditions, for it is quite possible for their theory to be embellished by another theory that more explicitly explains the relationship between specific emotions and outward behavior. Like Nussbaum’s theory, though, Ortony et al.’s theory by itself is quite uncommitted to any view on the relationship between emotion and action tendencies, and thus it provides no framework for which to explain this relationship. Ortony et al. admit that, “Certainly, these aspects [the physiological, behavioral, and expressive aspects of emotion] would
be crucial to a complete answer to the question of what an emotion is” (2). Nonetheless, the intent of their theory is only to account for emotion typology, and they find that cognitive structures are sufficient for this. Thus, explaining these other features is not their focus, for “they are less central to answering the question of where emotions come from” (2). Furthermore, they “remain unconvinced that [action tendency] is a characteristic of all emotions,” proposing that “action tendencies … turn out merely to be concomitants” (11). Thus, I conclude that Ortony et al.’s theory is consistent with C2 and C4 although it offers only a weak framework to address these conditions.

Lazarus does include action tendencies in emotion, and thus (according to his theory) every emotion would carry the motivation to act. His theory, then, meets C4. In that Lazarus’s theory takes cognition and action tendencies to be co-occurring in the experience of emotion (in that cognition carries into the emotion as he sees it), there may well be some way for his theory to then explain the tendency of some emotions to result in stronger motivation via the thought processes that occur alongside these action tendencies. Thus, I would find that, whereas Ortony et al.’s theory is also consistent in its address of C4 and C2, Lazarus’s theory offers a stronger address of both conditions.

These theories could meet C1, C5, C6 and C7 just as does Nussbaum’s theory. C1, we can recall, requires a theory of emotion to provide a framework for distinguishing between shame, guilt, disdain, and anger. Leaving aside the empirical question (of whether emotions are correctly identified as dimensional appraisals), both dimensional appraisal theories provide frameworks to distinguish these emotions. Lazarus proposes distinct core relational themes for “guilt,” “shame,” and “anger,” and it is quite conceivable how one could add disdain as a counter-part to shame (with the exception that it is an appraisal about another instead of oneself) (Emotion and Adaptation 122). Ortony et al.’s theory does not explicitly define guilt and disdain, but perhaps we could find remorse and reproach to be identical to these emotions. This could be questionable, in that there may be empirical challenges with equating remorse with guilt and disdain with reproach, though it is quite conceivable that these emotions are synonymous with one another. Even if they are not, the framework for distinguishing them according to other dimensions is provided in this theory; if need be, there would just be additional branches on the tree defining these distinct Attribution and Attribution/Well-Being emotions. C5 requires that the theory of emotion explain how emotions can be caused through the acceptance of norms that we are more or less aware of; appraisals, being cognitive evaluations according to dimensional appraisal theories, would be able to open to the influence of epis-
temic norms just as are other forms of judgment. Just as other types of judgments can be influenced by norms that we are more or less aware of, so can an appraisal. C6 requires that a theory of emotion provide a framework to explain how broadly moral sentiments shape guilt and anger; it is conceivable that broadly moral sentiments shape guilt and anger through the distinct causal appraisals of the broadly moral emotions. The judgments that form the causes (and thus define) the broadly moral emotions can surely impact the judgments that then cause (and define) the narrow moral emotions. (If one has specific sorts of thoughts that cause one’s sympathy for another, it is quite conceivable that these thoughts would then also – in some related way – cause one to feel guilty for harming the person.) C7 requires that a theory of emotion explain how guilt and anger respond to considerations about their consistency with other norms, their grounding reason, and their plausibility; because of the dimensional appraisals that are taken to cause (and, for Lazarus, constitute) emotions, there is a proposed logical structure to them which could explain their response to judgments of their consistency, reason, and even plausibility.

The dimensional appraisal theories offered by Ortony et al. and Lazarus are consistent with the norm-expressivist’s conditions if we successfully modify their depiction of the appraisals for guilt and anger and if these theories adequately incorporate the action-guiding nature of emotions. Both of these conditions are quite plausible, though. Given Berkowitz’s review of the different depictions of anger found in appraisal theories, I find such non-moral characterizations of guilt and anger (within dimensional appraisal theories) to be quite plausible even if these debates are ultimately left to the scientists. Lazarus’s theory already includes action tendencies in the emotion itself, even if it is not the focus of his analysis. It is entirely plausible that an appraisal theory of emotion may provide an account of emotion that is open to some greater specified relationship between action and emotion.

The dimensional appraisal theories of emotion put forth by Lazarus, and Ortony and Clore, may be consistent with norm-expressivism upon our embellishing these theories with other theories that more explicitly explain the relationship between emotions and action guidingness and upon our modifying these theories by way of arguments proposed in other dimensional appraisal theories. There are friendly variations of these theories which may help avoid the characterization of anger in guilt in terms of moral concepts. Still, the norm-expressivist is well served to continue inquiring into other theories of emotion that may offer more than plausible modifications and embellishments by which they meet the needed conditions.
4.6 Solomon’s Engaged Judgmentalism: Emotion as Judgments of the Body

While we have now considered Nussbaum’s version of “judgmentalism,” Gibbard makes the point to consider Solomon’s form of it (Wise 129). In his discussion, though, Gibbard considers Solomon’s older version of judgmentalism in which Solomon speaks of emotion as belief. As we now know, judgmentalism purports that emotions are judgments. Gibbard is concerned that the nature of judgments, on Solomon’s view, would render norm-expressivism circular. He quotes Solomon’s claim from 1976 that, “A change in my beliefs (for example, the refutation of my belief that John stole my car) causes and constitutes a change in my emotion (my being angry that John stole my car). I cannot be angry if I do not believe that someone has wronged or offended me” (qtd. in Wise 129). This view of the emotion, Gibbard notes, would cause problems for norm-expressivism in terms of its circularity concern (C3) since anger would entail a judgment of wrongdoing. Solomon’s judgmentalism, though, has more recently broadened to describe the emotion’s judgment in terms that are clearly distinct from ordinary belief. Justin D’Arms and Daniel Jacobson argue that Solomon is now better considered as a “quasijudgmentalist,” since Solomon more recently maintains that emotions are “‘hasty and dogmatic’ judgments, which can conflict with one’s considered belief or evaluation” (“Significance of Recalcitrant Emotion” 130). 29 In Solomon’s more recent writings, he no longer proposes that emotions are either caused by or constituted by ordinary, belief-centered sorts of judgment. His most recent, and thus final, position on the topic is that emotions are broadly construed judgments. In this section, I will explain Solomon’s quasijudgmentalist theory of emotion, and we will consider how this theory fares in light of the norm-expressivist’s conditions. In the end, I find that Solomon’s theory of emotion is consistent with Gibbard’s norm-expressivism after all, although only if certain assumptions are made within the framework of Solomon’s theory.

In his more recent writings, Solomon rejects the suggestion that emotions are sufficiently understood as “propositional attitudes.” Unlike Nussbaum, he denies that emotions have propositional “contents” at all. Rather, emotions are taken to be judgments which express our “subjective engagements in the world” (“Emotions, Thoughts” 77). 30 Solomon

30 In this section, all page citations are from Solomon’s “Emotions, Thoughts and Feelings: Emotions as Engagements with the World,” from his edited collection of essays.
maintains, as we will see, that emotions are phenomenological experiences that serve to evaluate the world. Emotions are taken to express our evaluation of the world itself, which (as Solomon sees it) is not to say that they express our evaluation of propositional-like content. He maintains that, because emotions are most directly judgments of the world itself, the emotion is not a propositional attitude at all. As he explains it, “If Fred loves Mary and hates spinach, the objects of his emotions are Mary and spinach, respectively, not propositions” (77). Solomon claims that “Emotions as judgments must accept as their “objects” both propositions and ordinary objects of perception (imagination, memory, etc.).” The objects of our judgments may include propositional content (like “I won the lottery”), though Solomon points out that we can also judge mere “perception (imagination, memory, etc.)” (82). One could be frightened by the thought that “my account is overdrawn,” by the vivid memory of frightening night, or by the surprising perception of a growling dog. I need not entertain propositions at all, and thus (on his view) it is misleading to claim that emotion is identical with a response (of any sort) to an evaluation of propositional content. Emotions may be very immediate responses to their objects, which require no linguistic analysis or conscious deliberation as their cause, thus he finds it improper to speak of emotions as if they were propositional attitudes or judgments with propositional content. Therefore, Solomon distinguishes between thought and judgment as follows:

while I find the language of “thought” just too intellectual, too sophisticated, and too demanding in terms of linguistic ability, articulation, and reflection to apply to all emotions, “judgment” seems to me to have the range and flexibility to apply to everything from animal and infant emotions to the most sophisticated and complex human emotions such as jealousy, resentment, and moral indignation. (82)

Thus, Solomon argues that we need not judge propositional content, in an emotion, but that we rather engage with the stimulus itself (which may or may not be of a propositional sort).

In maintaining that emotions are evaluations of the world, Solomon emphasizes the distinctly experiential aspect of these evaluations. Like the dimensional appraisal theory, he finds that emotion is constituted by experience rather than thought. Unlike dimensional appraisal theories, though, Solomon emphasizes the phenomenological experience of

these appraisals in his explanation of the emotive evaluation. In having an emotion, Solomon claims that we find ourselves to be “cognitively grappling with the world” (77). Perhaps Solomon would agree that the cognitive evaluation that takes place in an emotion involves a sort of cognitive grappling of its own sort; after all, Solomon explains that the emotion-judgment “has at its very basis and as background a complex set of aspirations, expectations, evaluations (“appraisals”), needs, demands, and desires.” Still, he does not take this cognitive grappling to be independent of bodily response. He would seem to reject the attributional theory of emotion, though; emotions are not but attributions of emotion-concepts to one’s inner state (as if the thoughts and the inner state are independent of one another), for emotive judgments are phenomenological experiences on his view. Solomon explains that the “traditional notion of intentionality – and, I now suspect, the concept of judgment, too – lack the keen sense of engagement that I see as essential to emotions” (77). He maintains that the phenomenological experience of emotion is realized through bodily changes, claiming that “arousal and action readiness should be subsumed under the more general phenomenological rubric of getting engaged in the world” (86). Given that all emotions are ultimately “engagements in the world,” and thus moral emotions are engagements in the world, then (on Solomon’s view) the judgments which constitute moral emotions would (under this theory) include arousal.

In this way, Solomon most recently claims that emotions are “judgments of the body.” Solomon maintains that, not only can emotions include affective change as part of their judgment, but all emotions include such affective/bodily change. As he puts it, “a great deal of what is unhelpfully called ‘affect’ and ‘affectivity’ and is supposedly missing from cognitive accounts can be identified with the body, or what I will call (no doubt to howls of indignation) the judgments of the body” (87). This is not a denial of his cognitivism, for he argues that, “a cognitivist theory can include affect [i.e., feelings], or much of what is intended by that misleading term” (81). Like the dimensional appraisal theories, Solomon’s cognitivism is a hybrid theory. Emotion is cognition and bodily experience; in fact, this cognition takes place (in part anyhow) via bodily response. As for the “bodily feelings in emotion,” Solomon proposes that they include “the autonomic nervous system (quickened pulse, galvanic skin responses, release of hormones, sweating)” and “the whole range of bodily preparations and postures” – all of which might be subsumed, he suggests, under the general category of “action readiness” (86). Unlike Nussbaum, then, Solomon is ready to concede that emotions are phenomenological engagements, in that they entail a bodily response and
even action readiness. He claims that, traditionally and in his own past writings, “bodily feelings have been ‘left out’ of the cognitive account, but I also believe that ‘cognition’ or ‘judgment’ properly construed captures the missing ingredient” (85-86). Still, while emotions (under Solomon’s view) include even the most immediate affect, they are responses to objects in the world and in this way he maintains that they are ultimately judgments. Emotive judgments, then, are broadly conceived and not “necessarily involving (or excluding) reflective appraisals and evaluations” (82-83). In describing such broad judgments, he now claims that the experience of the body plays an integral role.

Solomon’s theory highlights the distinction between an analysis of emotion in general and an analysis of emotion typology. Solomon’s more recent writings offer an analysis of emotion in general. The thing that all emotions share in common, on his view, is that they are broadly construed (and subjective) judgments about the world. (Emotions, then, do not have propositional content.) To analyze what all emotions have in common, we must do so by construing their judgments in a very broad, flexible sense.

Nonetheless, our ability to distinguish between different types of emotion may require us to analyze their distinct processes and, in some cases even, their distinct objects. With regard to emotion typology, Solomon notes the following:

   it is left open whether some emotions might be better analyzed in terms of perception, others in terms of thoughts or judgments, others in terms of construals, still others in yet more dynamic terms. The real work will continue to be with regard to particular emotions, and often with specific regard for the particular instance of a particular sort of emotion. (84)

Solomon’s theory is left loose with respect to the manner in which we distinguish the various emotions from one another. His theory does not propose or imply a specific structure for the typology of emotions. He suggests, rather, that our framework for identifying the typology of emotions may depend upon the distinct means of evaluations employed in each type of emotion (whether the evaluation is derived from ordinary thought, perception, some sort of construal, etc.). Like the paradigmatic behaviorist theory which purports that emotion types will be identified according to their distinct changes in the central nervous system as well as by their distinct elicitors (Chapter Three), Solomon’s theory holds that emotion types may be best determined by whatever researchers conclude about the “particular instances of a particular sort of emotion” (84). If
specific types of emotions are best analyzed in terms of their “thoughts or judgments,” then specific types of thoughts and judgments will distinguish them from one another (84). Solomon does propose, for example, that “moral indignation” (above) is distinct from animal emotions in that it is more sophisticated and requires complex cognition (82); if this is right, then moral indignation (and perhaps other moral emotions too) might seem to require specific types of judgment or thought as its cause. Even in this case, though, the sophisticated judgments that identify moral indignation might turn out to be describable in non-moral terms, just as proposed in the case of appraisal theories of emotion and Nussbaum’s judgmentalism. That being said, anger should not be confused for moral indignation, for anger is a more generalized response to offences. (The norm-expressivist needs a characterization of anger, not moral anger.)

On Solomon’s view, emotions are judgments which can respond to immediate perceptions or to propositional beliefs. The fact that moral emotions must be capable of responding to propositional claims (e.g., a feeling of guilt as a response to the belief that “I just cheated on my taxes”) does not necessarily negate the possibility that moral emotions can also respond to non-propositional objects. I see no reason to preclude a mother feeling immediately angry at someone harming her child, without thinking that “this is a harm”; after all, mother bears become enraged at such perceptions, and Solomon would surely claim that this anger is about the other creature (harming her cub) rather than a proposition. Given that humans have evolved from other animals, and given that human mothers may well report having had immediate emotional responses at the mere perception of harms to their own children, it seems quite plausible that humans can also experience anger in direct response to perceptions as do other animals. (While others may surely argue that these perceptual responses are in fact responses to propositional content, Solomon seems to consider “propositional thought” in a more literal sense – implying an ability to think in words. Bears don’t think in words, and neither do humans at all times that they experience anger. Thus, not all thoughts or judgments imply propositional thought in this sense.) It is also quite conceivable, I think, that guilt can respond, at times, to direct perceptions (e.g., a feeling of guilt at the perception of a dejected child).

Interestingly, Solomon notes that anger, like other sophisticated emotions, entail “judgments of responsibility” (83). We need not, though, equate this judgment of responsibility for a moral judgment. Other animals would seem (on many accounts anyhow) to get angry, after all, and they cannot make moral judgments in the norm-expressivist’s sense anyhow. A judgment of responsibility seems to, here, be but a way of identifying the source of a threat/harm and it would
seem possible to do this without employing moral concepts as evidenced by the immediate response to danger in some cases. While Solomon sometimes discusses anger as “a set of judgments of accusation, an indictment for a personal offence,” and while this language certainly has a moralistic tone to it, we should note that anger is nonetheless distinct (in his view) from moral indignation (“Logic of Emotion” 47). Either these discussions are intended to discuss moral indignation (or a specifically moral anger), or we can interpret these statements in distinctly non-moral terms. Thus, if such descriptions pertain to anger, then we need not characterize this indictment and accusation in terms of moral concepts. His theory would not dictate that basic, and especially primitive, anger need be characterized as a moral judgment. Solomon’s theory does not imply that anger and guilt respond “only” to propositional thoughts. Anger and guilt, like many other emotions, may turn out to respond to beliefs and perceptions. Clearly, then, anger and guilt need not (on his view) be caused by moral judgments nor even by judgments which (somehow) necessarily employ concepts of blameworthiness.

Solomon suggests that we might also explain the typology of distinct emotions in terms of propositional descriptions, even though we ought to bear in mind that emotions are not propositional judgments at all. He admits that “Belief may be perfectly appropriate in explaining emotion but it is inappropriate in the analysis of emotion” (80). In order to analyze distinct types of emotion, as noted above, we can look to their distinct processes – whether they arise from belief, perception, or some dynamic combination. Solomon suggests that to explain an emotion in terms of belief does not imply that belief constitutes emotion. He seems to regard the explanation of an emotion, in terms of belief, as but a way of describing the meaning of the emotion. Describing anger and guilt in terms of beliefs may allow us to conceive of guilt and anger concepts. Given that Solomon admits that other animals have emotions as well, there is nothing to preclude us explaining an animal’s emotion in terms of belief. For example, we might find the punishing behavior of bats and ravens toward cheaters in food-gathering\textsuperscript{31} to indicate that the bat is “angry,” and we might explain the bats’ anger as “their belief that the other bats acted wrongly.” It may be helpful to describe the bats’ anger as an “offence at the wrongs of others,” or some such thing, in order to understand the way in which these responses function in their own social lives, but none of this implies that bats entertain actual propositions in Solomon’s view. Thus, while Solomon finds that it is incorrect to analyze emotion as if it required an actual “propositional content,” he maintains

\textsuperscript{31} Damasio, \textit{Looking for Spinoza}, 160.
that we may be able to explain the meaning of an emotion (in terms of its functional representation) by describing it in terms of propositional belief. Explaining moral emotions in terms of belief, on Solomon’s view, does not imply that they are caused by belief (80). It only provides a way to identify these emotion concepts.

Interestingly, Solomon does not propose that the typology of emotion may be explained in light of distinct phenomenological experiences (or distinct bodily processes). It might be tempting to think that his theory would allow for emotion types (in some cases anyhow) to be distinguished from one another in this way, since the “cognition” which constitutes the emotion is so loosely construed and since emotions are judgments of the body. A distinct neurological or physiological manifestation of each emotion, if it could be found, might seem to provide clearer means of distinguishing between the emotions, though. Solomon does not propose a bodily distinction of any sort (whether physiological, neurological or phenomenological for example) between the emotions. Perhaps this is because he focuses upon emotions as subjective evaluations of the world. Here, we see Solomon emphasize the evaluative role in any analysis of emotion:

[Emotions] may not be analyzable in the mode of propositional analysis, but neither are they simply manifestations of the biological substratum…There are feelings, “affects” if you like, critical to emotion, but they are not distinct from cognition or judgment and they are not mere “read-outs” of processes going on in the body. They are judgments of the body, and this is the ‘missing’ element in the cognitivist theory of emotions. (88)

Emotions as judgments are not analyzable in terms of biological mechanism alone on his view. A distinction between emotions in terms of distinct bodily changes (or phenomenological experiences), alone, would not illustrate the nature of these distinct evaluations he finds so central to emotion.

As judgments, individual types of emotions must be distinct evaluations of the world. We would need to explain the typology of emotions in light of their distinct judgments, according to their means of judgment (perception, belief, etc.) and perhaps even the objects of these judgments. We may even explain an emotion’s judgment of a defining object in terms of belief. None of this need imply that the objects which define emotion types are conceived of in such terms by one who experiences the emotion. The bodily response can evaluate objects without such propositional thought on Solomon’s view. We may, though, explain
emotion types according to the means of judgment and the belief-like description of the way in which such judgments function.

4.6.1 Would Solomon’s Engaged Judgmentalism Suffice for Norm-Expressivism?

According to Solomon, it is incorrect to say that emotions have propositional content. Thus, guilt and anger would not be taken to be judgments of propositional content. This meets part of the circularity challenge, in that anger and guilt cannot be moral judgments according to C3. Clearly, if anger and guilt are not judgments about propositional content, then they are not moral judgments. C3 also dictates that moral judgments cannot be necessary causes for anger and guilt and that the concepts of anger and guilt must be defined without the concept of wrongdoing. According to Solomon, it is incorrect to claim that emotions have propositional content because this implies that emotions are always about propositions (which they are not in his view, as seen above). His point seems to be that it is incorrect to claim that emotions are always caused by propositional claims. Thus, there is no reason (as dictated by Solomon’s theory anyhow) to suppose that the moral emotions must be caused by moral judgments. It may even be the case, on Solomon’s theory, that anger and guilt could be caused by either percepts or judgments. In identifying these emotion types, it is important to consider how anger and guilt would be identified (per C1), though I find that in doing so we again find that nothing in Solomon’s theory implies that our understanding of guilt and anger concepts requires the use of moral concepts. Thus, Solomon’s judgmentalism would be consistent with C3 so long as guilt and anger concepts are definable without the concept of wrongdoing (i.e. blameworthiness).

Recall that C1 requires a theory of emotion to provide a clear framework for distinguishing between anger, guilt, shame and disdain (and other broadly moral emotions). Solomon’s theory is consistent with this condition, in that it provides a number of ways in which one could explain emotion types according to the distinct processes (and perhaps even objects) involved in the emotion. We could also distinguish between emotion types according to beliefs that they appear to functionally represent. Like the paradigmatic behaviorists, though, Solomon’s theory does not propose a specific nature for the cause of all emotions, and thus the defining criteria for moral emotions are left quite open. Solomon’s theory would allow for emotion types to be defined and identified according to the distinct processes that cause them. This may entail that certain emotion types are then identified and defined according to the
objects that elicit their distinct processes – whether these objects are thoughts, perceptions, or some combination of them. It may be that moral emotions are caused by specific moral judgments, as Solomon originally proposed (in his suggestion that anger was a judgment of wrongdoing), but his theory of emotion need not imply this. Thus, I take Solomon’s theory to be consistent with C1, for it offers a number of ways in which we can distinguish moral emotions from one another.

If anger and guilt can be caused by both perceptions and judgments, then we could distinguish them according to the ways in which they function. For example, it may be that sadness functions to appraise loss and, thus, sadness could be explicated as a belief (in Solomon’s view) that a loss of personal significance has occurred even though (as he notes) no ordinary “belief” need take place at all. Though Solomon does not specifically mention explaining emotion typology in terms of such functions, nothing in his theory’s framework would prevent this approach in order to explain emotion types in terms of such distinguishing appraisals. Emotions may be responses to stimulus circumstances (as attributional theories propose), which can be described as if they were beliefs of sort (given the way they function), even if such emotional responses can occur also quite automatically in response to a cluster of stimuli. If we try to explain emotion typology in terms of belief, then we will obviously need to say what each type of belief is about. It seems we could adopt the same friendly option for norm-expressivism as we did in the case of attributional theories and Nussbaum’s judgmentalism. The concept of anger could quite plausibly be identified in terms of the belief, perhaps as the belief that “An important goal has been obstructed and it is in my power to change this obstruction” or even that “An important goal has been obstructed and I desire to lash out.” The concept of guilt could (also quite plausibly) be identified in terms of the belief that, “I have acted (or am in a state) that tends to meet a damaged communal relationship.” In this way, Solomon’s theory could satisfy both C1 and C3.

A drawback to Solomon’s broad approach toward identifying emotion types is that (as with the paradigmatic behaviorism) this framework is not very informative by itself with respect to explaining emotion types. The loose options, by which we might distinguish moral emotions, carry an ambiguity that is not highly desirable for the norm-expressivist in seeking specific answers on the nature of moral emotions. Thus, we find another cognitivist theory of emotion that would meet C1 and C3 so long as sufficient assumptions are made in accounting for guilt and anger. The assumptions are both consistent and quite plausible ones, but they do purport greater specificity than the theory of emotion itself put
forth. Given that Solomon’s theory provides a framework to define (and distinguish) these moral emotions in a way that may, but need not, avoid moral concepts, I would say that his theory meets C1 although (as with the other cognitivist theories thus far) it does so in a weakened way. His framework for defining moral emotions is too loose to really *ground* a strong argument for the nature of moral emotions (with respect to moral concepts) one way or the other. The flexibility of his theory in this regard may certainly be a merit in the independent merits of his theory (in that it may work with various empirical findings), though it leaves the norm-expressivist wanting more reassurance with respect to C3. Unless we define moral emotion concepts (particularly *anger* and *guilt* emotion concepts) with specific and supportive arguments showing that these emotion concepts employ no moral concepts in their definition, which Solomon’s theory does not provide, then his theory will remain only conditionally consistent with C3. I accept that such arguments are plausible, and that they are consistent with Solomon’s judgmentalism, though we should recognize this as the embellishment that it is to his theory.

The norm-expressivist’s other conditions are also met by Solomon’s theory. Because the emotion is taken to subsume action readiness as well as judgment, the theory would seem to provide an account of the emotion’s action-guidingness (meeting C4). Inasmuch as bodily changes can differ with different emotions (prompting different sorts of action readiness) and inasmuch as emotive judgments are evaluations of their objects which engage us with the world, it is quite consistent within his theory that some emotions (such as, potentially, anger and guilt) tend to be stronger motivators of action than are other moral emotions. Thus, the theory meets C2. In that emotions are phenomena that we share with other animals and human infants, the theory would allow for (though it need not require) moral emotions to be derived from norms that we are unaware of in any conscious, reportable way. (After all, though dogs do not “accept” norms, they do have norms governing their emotional responses and they need not be aware of the norms.) At the same time, the theory allows for beliefs to be the object of emotion; thus, moral emotions can be derived from norms that we are well aware of at times. The theory, then, would meet C5 in that the theory would explain how emotions can be caused by the acceptance of norms that we are more or less aware of at the time. Because emotions can respond to beliefs (and because this is especially so with regard to the more sophisticated human emotions), the theory would explain how guilt and anger respond to considerations of their consistency with other norms, their grounding reasons, and their plausibility. Thus, in accounting for ways in which emotions respond to norms that we are more or less aware of, the theory
meets C7. Solomon’s theory does not prohibit emotions from shaping other emotions; some judgments, after all, help shape other judgments. Thus, the theory is not inconsistent with C6. While the theory meets C6, it does provide a weak explanation as to how certain emotion-judgments impact other emotion-judgments and norms for them; because we lack a clear proposal, from this theory, on the specific processes involved in distinct emotions, it is unclear how such influences occur between the emotions. While his theory does not explain how such influences occur, it does not seem to prohibit the possible explanation.

Solomon’s theory meets all of the norm-expressivist’s conditions, though it does so in a conditional way. It does so with embellishments (in that the compatibility between his theory and norm-expressivism would require that we explain guilt and anger in terms of non-moral concepts), though (in spite of Gibbard’s concerns) we need not modify any of Solomon’s arguments with respect to his more recent theory of emotion. Judgmentalism (again) turns out to be as compatible with norm-expressivism as are the attributional theories of emotion that Gibbard favors.

4.7 Greenspan’s Perspectivism

Like others, Greenspan proposes a hybrid cognitivist theory of emotion as well. Greenspan’s theory shares features with Solomon and Nussbaum’s theories. She explains that, “I take the affective element of emotion as a propositional attitude, an attitude with a propositional thought as its content” (“Emotions, Rationality, and Mind/Body” 130). As with Nussbaum, dimensional appraisal theorists, and attributional theorists as well, Greenspan finds that we can accurately identify the nature of emotions through propositional descriptions of the emotions’ judgments. Like Nussbaum, Greenspan argues that we can identify the emotions of other animals and infants (as well as adult humans) by describing their emotive judgments in propositional form. Like Solomon, though, Greenspan emphasizes that emotions include affective changes in the body. Like Solomon, she aims to bridge together affect and cognitive judgment in one theory of emotion. She argues that emotive judgment results from a special rationale, a form of reasoning that she takes to be distinct to our emotional lives. In this section, we will consider how this rationale is taken to explain moral emotions on her view. Again, we will then consider if her account of moral emotions is consistent with the norm-expressivist’s conditions.
An emotion, in Greenspan’s view, occurs when one has an affective response to propositional content. Emotions, she argues, are “propositional feelings” (Emotions and Reasons 4). As she puts it, “Emotional affect or feeling is itself evaluative – and the result can be summed up in a proposition” (“Emotions, Rationality” 132). She distinguishes propositional content from belief or judgment. She explains that an emotion need not be caused by the belief in propositional content, but emotion does require that we entertain this propositional content (perhaps imagining it). She argues that emotion is a combination of the affective response and the propositional content. Thus, she proposes the following analysis of emotion:

we look at emotions as compounds of two elements: affective states of comfort or discomfort and evaluative propositions spelling out their intentional content. Fear, for instance, may be viewed as involving discomfort at the fact – or the presumed or imagined fact (I shall say “the thought”) – that danger looms. (Emotions and Reasons 4)

Greenspan argues that, when we have an affective response to the propositional content of an emotion (such as a fearful affective response to one’s imagining that “danger looms”), then we experience “feeling as if” the propositional content were true. Perhaps this “feeling as if” is analogous to Gendler’s “alief” (discussed above, in Section 4.3); the idea is that we need a concept to identify the way in which people can respond to impressions short of full-scale, ordinary belief or higher level thinking. Greenspan maintains that her theory is distinct from judgmentalism in that she does not purport that an emotion is an actual judgment of the propositional content (Practical Guilt 151). The evaluations that constitute emotions are not “beliefs” because emotions can exist in spite of disbelief (such as is cases of recalcitrant emotions, when we have lingering emotions in the face of changed beliefs) (Emotions and Reasons 84). It is not distinct beliefs that distinguish emotion types, but rather the propositional content to which we entertain and respond which distinguishes the emotion. Additionally, though, Greenspan notes that the “affective element” can contribute to our study of emotion typology. While she admits the “inadequacy of affect as a basis for distinguishing different emotions types,” she maintains that “it does not follow that affect adds nothing relevant” (“Emotions, Rationality” 131). Thus, emotions on her view are affective responses to propositional content, and we can

---

32 Others, like D’Arms and Jacobson, argue that her view is (like that of Solomon) a form of “quasijudgmentalism” (130).
distinguish between distinct types of emotions by identifying their
distinct affective patterns and propositional content.

Greenspan is perhaps most noted for proposing a distinct emo-
tion-rationality. (She is not proposing that our emotional lives “should”
operate according to this rationality, but rather that our emotional lives
“do” operate according to this distinct rationality.) Greenspan analyzes
the rational structure of emotion in contrast to ordinary judgments. She
admits that “the appropriateness [of emotion] may be explained in terms
of belief warrant” (Emotions and Reasons 4). Nonetheless, though, she
proposes that the “rational warrant for an emotional response varies with
evaluative perspective – in a way not recorded in qualifications to the
content of emotions, unlike what is supposed to be the case for judg-
ment” (“Emotions, Rationality”128).

Greenspan calls her theory a “perspectival” theory, because it
emphasizes the subjective way in which people consider the warrant of
their emotions. She explains the perspectival account as follows:

On the perspectival account, what emotions register, when the
mechanism is working properly, is not necessarily the “all things
considered” view of things by which we assess our beliefs. To
say that an emotion is reasonable, or rationally appropriate, is to
say that a certain evaluative belief that represents the content of
the emotion (for anger, for instance, that someone has done me a
wrong) would be warranted by a significant subset of the evi-
dence – significant in the sense of “worth holding in mind.”
(“Emotions, Rationality” 128-129)

Emotions are regarded as “adequately warranted by a partial subset
of the total body of evidence bearing on its corresponding judgment – a
perceptual “slice” of the evidence” (Practical Guilt 151). Greenspan
argues that emotions are rational, rather than irrational or arational, in
that they respond to reasons and they express judgments of warrant.

Greenspan fleshes out (what she takes to be) the distinct nature
of emotion warrant. She makes the point that reasoning about the warrant
of one’s emotions (which can be very fast, and even pre-linguistic as
noted above) is of two sorts: it concerns “evidential or representational
rationality” and “strategic or instrumental rationality … adaptiveness”
(“Emotions, Rationality” 129). We may find evidence of offence, and
decide it is not strategic to take offence (be angry). If it’s my boss instead
of my child who said something a bit harsh, it could be best to let it go
lest I seem to adopt an air of authority with my superior. If my child uses
a harsh tone with me, perhaps it is best to take some offence in order to
promote good manners. Likewise, we may find barely the right circumstances to warrant excitement, but find it very strategic to adopt the emotion; I may become quite enthused for someone who desperately needed any bit of good news. It is not always conducive to our goals to feel guilty or angry with others, and this pragmatic realization can impact our tendency to actually have such emotions in that such realizations affect the way we perceive the warrant for the emotions. Greenspan suggests that our strategic analysis of an emotion’s warrant quite naturally modifies our evidential requirements. (My boss would, perhaps, have to use a much harsher tone than my child in order to warrant my taking offence.) The rationality of an emotion consists in evaluating such propositional content in light of both evidential and strategic reasons and the interplay between such reasons.

It is not the preponderance of reasons that calls for an emotion, but an emotion is naturally taken to be called for whenever there is enough reason to call for it; we need not, or do not, necessarily check emotions against counter-evidence in order to find them warranted. As she says, an emotion is treated as rational whenever it is “something like ‘rationally acceptable,’” or adequately grounded in the situational evidence, rather than “rationally required,” or mandated by the evidence, as on the usual standard for assessing belief” (129). It only needs a “slice” of the whole picture to warrant an emotion (*Practical Guilt* 151). It is quite rational to feel sad about moving, even when one is excited for the next opportunity; the loss of whatever good things are being left behind is what elicits the sadness. Focusing on the exciting new prospects is the rationale for one’s joy.

In her book *Practical Guilt*, Greenspan has much to say about the nature of guilt. I suggest that her methodology, for defining guilt, could then readily be applied to a description of anger. Greenspan claims to defend “a nonjudgmentalist view of guilt that sees the subjectively guilty agent as ‘feeling as if’ he were morally responsible” (151). When one feels guilt, one feels discomfort with the “thought” that one has done something wrong. One need not judge that one has done something wrong, but (as discussed earlier) one feels “as if” one has done so. She maintains that guilt can be analyzed as follows:

guilt amounts to discomfort with a certain evaluative propositional object and hence may be said to correspond to a judgment, though one can undergo the feeling without holding the judgment … My view will allow us to interpret survivor’s guilt and other disputed cases as having the same sort of content as moral instances of guilt…”(151)
Because emotions can be warranted by a sub-set of evidence, in contrast to ordinary judgments, Greenspan maintains that “guilt is sometimes appropriate, in contrast to blame, when we do not have adequate warrant for the corresponding judgment” (152). Though guilt would necessarily (on her account) entail an affective response to this propositional content, she does not utilize the affective element in her account of this emotion’s typology. Perhaps the affective response could account for the strength, or various degrees of guilt, but we can identify guilt without such references in Greenspan’s view. Guilt is a negative affective response to the thought of one’s own moral responsibility in some regard. Guilt is feeling as if one is morally responsible for a wrong.

Likewise, anger is then feeling as if some propositional content holds true. Anger, though, is shared with other animals and infants. Thus, the propositional content of anger need not refer to moral responsibility. Greenspan notes that anger may also be defined by its own distinct affective changes in addition to its propositional content, for “feeling quality needs to be brought in when we attempt to make finer distinctions between rage and resentment, say, which may have the same internal object” (Emotions and Reasons 54). Still, anger will also be defined (at least in part) by its propositional content. Greenspan notes that anger comes in many forms – as illustrated in animals (such as the cat who becomes angry at a perceived threat), infants (who may lash out when restrained), and adult humans (who can become angry at insults and broken down cars) (Emotions and Reasons 48-54). She proposes that, in keeping with Aristotle, “anger intrinsically involves discomfort at a thought of revenge as something the agent should arrange for” (Emotions and Reasons 52).

While anger can be impersonal (such as when we become angry at the weather), Greenspan notes that anger at moral violations is of the personal sort; we are, in these cases, angry at the agent. Greenspan describes personal anger as follows:

If I am angry at X in response to his insult, that is- not simply hurt, or angry about his role in something unintended – I must view his insult as a wrong that I ought to repay. I must “feel as though” I ought to repay it at least partly because he is to blame for it – and because I am uncomfortable at that thought. (Emotions and Reasons 54)

While anger is always a negative affective response at (the thought of) a desire to take revenge (even if one does not ultimately desire, or choose,
to take revenge), personal anger is a response about the desire to take revenge against another who is found to be blameworthy. On this view, then, moral anger requires that one feels as if another is blameworthy. It may be tempting to suppose that, if a moral judgment is to call for anger, then it would only be reasonable to think that it calls for moral anger. This, though, is not the norm-expressivist’s claim, for this would surely make the moral judgment a circular one. Rather than looking to Solomon’s “moral indignation” or Greenspan’s “moral anger,” we need only consider a more general (even primitive) anger. It is most likely to be the most basic, non-moral anger, to which a moral judgment need refer in the norm-expressivist’s view for this type of anger is most likely to be elicited by non-moral objects as well as moral judgments.

Greenspan’s perspectivism implies that both anger and guilt are affective responses to propositional content. Only guilt, though, is clearly identified in terms of moral concepts. We can now consider whether her theory holds promise for norm-expressivism.

4.7.1 Would Greenspan’s Perspectivism Suffice for Norm-Expressivism?

Greenspan’s theory of emotion would meet C1 in that her theory provides that we can distinguish emotion types from one another according to their propositional content and perhaps even their affective elements as well. Emotion types cannot, on her view, be identified solely on the basis of one’s affective state, and thus one’s physiological changes alone will not suffice to define an emotion type. To distinguish emotion types from one another, and thus understand emotion (type) concepts, we need a description of the emotion types’ propositional content. As she puts it, “Emotions can be viewed as having a content expressible propositionally, or in terms of what they ‘say’ about their objects.” Each emotion type “registers” particular information about its object(s) (e.g., fear registers a thought of danger), and this information is articulable in a propositional claim which defines the emotion type (“Emotions, Rationality”116).

In order to address the circularity concern (C3), the distinction between moral emotions would need to be accomplished in terms of articulated propositional content that specifies their distinct evaluative content without employing moral concepts. Greenspan provides a description of anger (though not moral anger) which perhaps already employs no moral concepts, even if her description of guilt clearly entails the concept of moral responsibility. If the concept of guilt were (as she
proposes) defined in such a way, this would violate C3. Greenspan’s characterization of anger may also violate C3.

Greenspan describes anger as “discomfort at a thought of revenge as something the agent should arrange for.” Though I do not take this description to necessarily require moral concepts, this is arguable. This description may, quite arguably, imply a moral concept in that revenge would seem to imply a perception that the harm sought is in response to wrongdoing (which is, after all, what makes the harm to be done to another an act of revenge). Although this argument has some merit, we are then back at the possibility that the norm-expressivist may revise this thought content into something less moralistic. The thought content for anger could just be that “one desires to lash out (act aggressively) in response to an obstructed goal.” Thus, anger would be defined as “discomfort at the thought that one desires to lash out (act aggressively) in response to an obstructed goal.”

If her theory is to satisfy C3, then we would also need to modify the propositional content of guilt. This would surely be a new twist on Greenspan’s theory, though, and she would likely resist such proposals being that she has developed her own careful arguments for taking the propositional content of guilt to be as she does. She does find that the most plausible propositional description we get of guilt seems to entail moral concepts. Nonetheless, if the norm-expressivist is to accept that other cognitivist theories can avail themselves of this option, then there need be no more difficulty with Greenspan’s theory doing so. Whether we aim to couch these moral concepts as propositional content, objects of the emotion, or dimensions of an appraisal, the challenge is the same. Distinct moral emotion concepts must be conceivable without moral concepts, or else norm-expressivism is a circular theory. Greenspan’s theory, like Nussbaum’s theory, would emphasize that we can only distinguish between emotion types by articulating precise propositional content. There is no reason, though, why someone might not adopt her framework for explaining emotion types and yet describe the propositional content of guilt in non-moral terms.

If the propositional content of guilt was (rather than a claim about one’s own blameworthiness) the claim that “one’s own actions or state is of the sort to meet a damaged communal relationship,” and if the propositional content of anger was (as she adopts from Aristotle) taken to be “revenge is something the agent should arrange for,” then neither of these would make the moral judgment in norm-expressivism circular, for neither the content of guilt nor anger need imply the concept of blameworthiness. Thus, Greenspan’s theory need not fail to meet the norm-expressivist’s C3, although it meets this condition only through allowing
us to change the description of the propositional content of both anger and guilt as she takes it.

Greenspan’s perspectivism would also meet C5 or C6. In accordance with C5, it does provide a framework to explain how anger and guilt can be caused by the acceptance of norms that we are more or less aware of at the time. Under Greenspan’s theory, I could “feel” as if one is blameworthy and this could be caused by my finding of warrant for this response which is derived from norms which I (like other animals) lack awareness of. We need not be able to consciously report our reasons for emotions in her view, for they may be pre-linguistic as they are in the case of modular emotions. Thus, her theory provides a framework for explaining how anger and guilt can be caused by the acceptance of norms that we are more or less aware of at the time. In accordance with C6, Greenspan’s theory also provides a framework to explain how broadly moral sentiments can help shape guilt and anger. It is surely possible for sympathetic feelings to help set one’s “evidential requirement” for guilt over harming others. If one tends to feel highly sympathetic toward other beings, then one will likely accept stricter moral norms that call for guilt in response to causing others harm. For example, if one does tend to feel sympathetic toward the suffering of vulnerable beings (such as being physically sickened at the sight of their pain), then this might explain why one accepts that “One ought to feel guilty for breaking social contracts to provide for vulnerable persons, and others ought to feel angry at such behaviors.” We can now translate this judgment in order to reference the (revised) definitions of guilt and anger concepts, and thus we would say, “One ought to feel discomfort at the thought that one has done something to meet a damaged communal relationship whenever one breaks social contracts to provide for vulnerable persons, and one ought to feel discomfort at the thought that one desires to lash out (act aggressively) in response to an obstructed goal whenever another breaks a social contract to provide for vulnerable persons.” In this way, Greenspan’s theory provides a framework that is capable of explaining how broad moral sentiments shape guilt and anger.

Her theory would also meet C2 and C4. According to C4, the theory must provide a framework to explain how guilt and anger motivate action. If emotions are felt evaluations with distinct propositional content, then (as Greenspan maintains) emotions will have their own “characteristic phenomenological symptoms” (“Emotions, Rationality” 129). Because they have their own phenomenological symptoms and their own propositional content (which constitutes the object of emotion), emotions manifest in recognizable behaviors – at minimum, physiological behaviors which constitute their felt qualities. Greenspan does argue,
as we have seen, that emotions promote our long-term ends by functioning as barriers to ordinary rational deliberation. Given the evolutionary bend of her theory, it would seem that emotions (are taken to) promote our long-term ends in that they (generally) promote socially coordinating behavior through acting as “commitment devices” (“Emotions, Rationality” 126). In this way, then, her theory provides a framework confirming Gibbard’s contention that emotions motivate behavior. That is, emotions motivate us to act upon their evaluations even when other types of reasons exist to act otherwise. According to C2, the theory must provide a framework to explain why some emotions (guilt and anger) are stronger motivators of action than are others. Greenspan’s theory would be consistent with some emotions being stronger motivators than others. If emotions are felt judgments which promote our survival and well-being, then emotions that result in behaviors which are more significant to our survival would conceivably have strong action-guidingness than other less critical/fruitful emotions. If Gibbard is right, and anger and guilt are the emotions which best motivate social coordination, then there is nothing in Greenspan’s theory to preclude them having strong action-tendencies.

Though Greenspan’s characterization of emotion “warrant” is distinct from the norm-expressivist’s characterization of “rationality,” her theory does meet C7. According to C7, the theory of emotion must explain how guilt and anger respond to considerations of their consistency with other norms, their grounding reasons and their plausibility. This means that anger and guilt can be caused by judgments that are derived from language-infused processes in consideration of logical consistency and cultural perspectives. One might have concerns that her theory cannot meet this condition since she casts the emotion’s rationale quite differently from Gibbard’s rationality. An emotion-judgment on Greenspan’s view does not express the claim that one finds the propositional content to be consistent and warranted by the preponderance of reasons, as in Gibbard’s sense of rationality (Chapter Two). Emotion types respond to their own (emotive) rationales, which requires only a slice of evidence for judging an object as one does. Gibbard and Greenspan propose different meanings for “rational.” Nonetheless, there is no reason to exclude moral judgments (as Gibbard defines them, per Chapter Two) from being among the slices which can warrant anger and guilt on Greenspan’s view. Thus, guilt and anger can respond to such considerations of their consistency, plausibility, and grounding reasons in this way.

Greenspan’s perspectivism, then, may meet all of the norm-expressivist’s conditions in a mitigated way. It meets C3 as do the other
cognitivist theories, though it does so only by amending Greenspan’s characterization of the propositional content. The ability of her theory to be compatible with norm-expressivism, then, would require a departure from some of her arguments around the specific propositional content of (personal) anger and guilt. Greenspan proposes a nuanced framework for the rationale of emotions, and much of this may appeal to the norm-expressivist. Her theory also aims to connect cognitive and somatic considerations found in emotion research into one holistic theory. Perhaps the norm-expressivist can appreciate the progress in proposing states of mind which respond to reason and yet, at the same time, states that are not best explained as ordinary belief.

4.8 Conclusion

After considering a broad spectrum of cognitivist theories of emotion, we find that all of these theories offer frameworks that are compatible with norm-expressivism – should we successfully depict moral emotion concepts in non-moral terms. While the good news for the norm-expressivist is that nothing in the framework of these theories of emotion (insofar as they explain emotion and emotion types) rules out a non-moral definition of guilt and anger, the bad news is that several of these cognitivist theories (Nussbaum’s judgmentalism, dimensional appraisal theories, and Greenspan’s perspectivism) each provide arguments that depict anger and/or guilt in terms of blameworthiness. It is only through proposing revisions, or at least embellishments, to their account of guilt and anger that the cognitivist theories can all meet C3. Thus, while I maintain that such arguments are plausible, the compatibility of these theories with norm-expressivism is perhaps weakened by the need for such revisionist arguments.

While the main contingency in the capacity of these cognitivist theories to meet all conditions was the circularity challenge (C3), other challenges are found for specific theories. Nussbaum’s judgmentalism is weak in its address of C2 and C4, as would be Ortony et al’s dimensional appraisal theory. While other dimensional appraisal theories may fare better in meeting C2 and C4 (such as seen in Lazarus’s theory), the norm-expressivist should also find that attributional theories, Solomon’s judgmentalism, and Greenspan’s perspectivism address all seven conditions with respect to the action-guidingness and motivational strength of emotion types.

Let us now consider whether somatic theories of emotion will offer compatible frameworks in the analysis of emotion for the norm-
expressivist. Should somatic theories of emotion meet all of the norm-expressivist’s conditions for a compatible theory of emotion, and should these theories better address the circularity challenge without the need of revision or such embellishment in defining moral emotion concepts (as was needed by the compatible cognitivist theories), then these theories would offer the most directly compatible theory of emotion for norm-expressivism. A most ideal theory of emotion for the norm-expressivist would be one that offers a framework that is consistent with all of the norm-expressivists and a theory that provides its own strong case for characterizing guilt and anger without the use of moral concepts.
5. Somatic Theories of Emotion

5.1 Introduction

William James, the father of modern psychology, suggested that if we take the feeling out of an emotion and leave only the thoughts, then we lack any emotionality (“What is Emotion” 193-194). Contemporary somatic theories of emotion share this belief, emphasizing the necessity of a bodily response to any emotion. On the face of it, what distinguishes somatic theories from cognitivist theories of emotion is that somatic theories claim that appraisals can take place without cognition – through bodily changes themselves or through the perceptions of such bodily changes. What distinguishes them from behavioral theories (even paradigmatic behavioral theories) is that somatic theories (like the cognitivist theories) aim to account for conceptually laden evaluations in the typology of emotions – which cannot be accounted for in behaviorist terms. In this chapter, I will discuss two major contemporary somatic theories: Robinson’s “process theory” and Prinz’s “perceptual theory.” Jesse Prinz argues that the perception of bodily changes in response to the perception of stimuli is *sufficient* for emotion, whereas Jenefer Robinson argues that bodily change is necessary to emotion though we will also find an initial non-cognitive affective appraisal (or evaluation) and cognitive monitoring of these changes as well. In considering somatic theories of emotion, I sought theories that were well recognized in this contemporary field. In the end, I find that norm-expressivism can realize its most compatible theories of emotion to be such somatic theories of emotion. I suggest that these somatic theories do not neglect cognitivist concerns and, yet, they will account for these concerns in a way that is more consistent with norm-expressivism than are cognitivist theories as a whole. While other theories can meet the norm-expressivist’s conditions via plausible modification or embellishment, these theories require the least amendments.

By a “somatic theory of emotion,” I mean a theory that defines emotion (at least in part) in terms of embodied appraisals. An emotion, in this view, requires bodily change or the perception of bodily change.
Emotion may be identified with such bodily changes (in part, at least, as in Robinson’s theory), or the emotion may be the perception of such changes (whether the changes occur or not, as in Prinz’s theory). I refer here to bodily changes in a very broad sense. Here, Prinz explains what he intends to include in “somatic” changes:

> Emotion researchers tend to the use the term “somatic” broadly …. In this context…the term “somatic” encompasses any part of the body. Somatic states include states of the respiratory system, circulatory system, digestive system, musculoskeletal system, and endocrine system …. Damasio expands the range of bodily states underlying our emotions to include…changes in the levels of chemicals in the brain, such as changes in hormone levels caused by the endocrine system. (Gut Reactions 5)

Robinson and Prinz both argue that emotion entails at least the perception of such broadly construed somatic changes.

Both theorists aim to explain how (the perception of) bodily changes can evaluate the objects of our emotions in ways that account for emotion typology. In this chapter, I will explain how Robinson’s process theory of emotion and Prinz’s perceptual theory of emotion address these aims. To best understand these theories, we should begin with an overview of some work in neuroscience which is taken (by Robinson and Prinz) to particularly support their theories. Understanding the theories in some detail will allow us to then understand the way in which each meets the norm-expressivist’s conditions for a compatible theory of emotion short of modification or (in the case of Prinz’s perceptualism) even embellishment.

5.2 The Beginnings of Embodied Appraisal Theories in Science

Contemporary somatic theories of emotion claim the support of neuroscience and psychology. In this section, I will explain some of the work in the sciences which both Robinson and Prinz draw upon. An understanding of these empirical contributions to the study of emotion should illuminate the differences between Robinson’s and Prinz’s theories, as well as the shared focus upon bodily responses found in emotion according to both theorists.

William James famously argued that the common sense view of emotion has things backwards. Though we may tend to think of an emo-
tion as the cause of bodily reactions, James argued that bodily changes (including behaviors) occur first and we then have the emotion. Emotions are feelings of bodily changes that respond to perception. The emotion is, in his view, the “feeling of the [bodily] changes as they occur,” and the cause of the bodily changes is the “PERCEPTION of an exciting fact” (“What is Emotion” 189-190). Thus, a perception of an exciting stimulus causes changes in the body, and when those changes are felt, the emotion exists. The emotion just is the feeling of bodily changes in response to perceived stimuli. An understanding of this is especially important, given that a central strength of somatic theories is their connection to the sciences. In James’ view, emotions are always conscious in that they “must occur within hemispheres that indubitably cooperate,” and the mind is then clearly aware of the emotions in some way that allows for this cooperation between the perception of stimuli and changes in the body (Principles 165). Nonetheless, this consciousness does not imply that one is able to report or accurately remember one’s emotional states. Though emotions are always conscious in this broad sense, James admits that the consciousness of them “is split off from the rest of the consciousness of the hemispheres” (Principles 165). There are times at which emotion is accessible only to certain parts of the mind, and when other parts of the mind cannot access these perceptions or perhaps even the feelings in response to the perceptions. Guilt and anger, then, would be the feeling of bodily changes in response to the perception of their objects. As we will later see, both Prinz and Robinson propose somatic theories of emotion that pay tribute to these initial arguments put forth by James.

Recently, the neuroscientist Antonio Damasio defended a Jamesian account of emotion. An emotion itself, in Damasio’s view, is the brain’s mapping of a body-state (111-112). He explains that these body-states include “a complex collection of chemical and neural responses forming a distinctive pattern” (53). Damasio maintains that, whereas an emotion is initially caused by the brain’s detection of a stimulus (often referred to as an “emotionally competent stimulus,” or ECS) and whereas this detection amounts to an “appraisal” (or evaluation) of the perceived object in terms of one’s well-being, this detection then automatically triggers the brain to “map” bodily changes (53- 54, 111-112). That is, the perception of an ECS triggers new bodily-states which are mapped out, so to speak, in the brain. It is the brain’s mapping (or registering) of bodily change in response to the perception of ECSs that actually constitutes the emotion on his view. Damasio explains that “executed responses themselves … constitute emotions” (65). Damasio’s view is similar to James’ view in that they both take the emotion to be
some sort of awareness of bodily changes that result from the perception of stimuli. Still, there are some differences between James and Damasio’s views.

Damasio argues that the bodily changes are sometimes but mapped in the brain, even when these changes are not actually occurring in the body. Through an “as-if-body-loop,” as Damasio calls it, we can map bodily changes even when they don’t exist (115-121). We can, for example as he explains, experience empathy as an emotion, even though we need not actually have the same bodily changes that others have; in feeling empathy for another, we map such bodily changes in the brain and thus perceive such bodily changes even though we may lack the bodily states (115-116). Thus, Damasio infers that emotional experiences need not be (as James thought) perceptions of actual bodily changes. One may register bodily changes in the brain that are not in fact occurring, and in this way have an emotion without the actual physiological changes in the body itself. Thus, for Damasio, it is the perception of a body-state in the mind which actually constitutes an emotion.

Also different from James, Damasio clarifies that emotions are distinct from feelings. The (body-state) maps which constitute emotion result in feelings only when they arise to a “mental level” of biological operations” (177). He explains that emotions entail feelings whenever there is a conscious recognition of the body-state map; that is, feelings occur when one perceives that one is having a bodily response to the object of the emotion. Damasio explains that this “mental event” allows us to then reason about the emotions (realizing that one is having a particular emotional response) and it prompts reasoning about the “objects that triggered the emotions” (177). Not all emotion states will be felt (for even insects have emotions in Damasio’s view, and they have no mental events as he sees it), but felt emotions will engage mental deliberation in ways that support “flexibility and high power information gathering that mental processes can offer, as well as the mental concern that feelings can provide” (43, 178). Human emotions do result in feeling, which supports our sophisticated problem-solving in response to emotion (42). Feelings facilitate flexibility in coping with ECSs. Damasio explains that “The process of learning and recalling emotionally competent events is different with conscious feelings from what it would be without feelings. Some feelings optimize learning and recall. Other feelings, extremely

---

33 Damasio notes that, not only can the mind map bodily changes that do not occur, but it can block bodily changes that do occur such as in cases of “natural analgesia”. In these cases, too, the brain generates “false body maps” (113). In such cases, one’s brain maps different body-states which are more conducive to our well-being or survival – such as in running fiercely from a predator, without noting one’s injured leg (113-114).
painful feelings in particular, perturb learning and protectively suppress recall” (178). Unlike James, then, Damasio does not regard emotions as “feelings,” even though this body-state mapping (which constitutes emotion) results in feeling upon the perception of the body-state maps (86). Because emotions have evolved from simpler creatures, we should recognize that emotions themselves do not include feelings; in humans, though, emotions can lead to feelings. This does not imply that human emotions always lead to feelings, but it does imply that human emotions can (and sometimes do) lead to feelings and thus higher level reasoning about ECSs. In spite of Damasio’s argument that emotions are body-state maps (and not feelings), he seems to support a Jamesian view for the most part. For both James and Damasio, an emotion is taken to occur when the perception of stimuli causes a state of mind that registers bodily change.

The psychologist Robert Zajonc is well known for disputing Lazarus’ cognitivism. As discussed in Chapter Four, Lazarus argues that empirical data sufficiently proves that cognition is causally necessary for emotion. Zajonc disputes this view, maintaining that the empirical data evidences that “emotion precedes cognition” (119). Zajonc notes research showing that encoded data (in response to a stimulus) can travel direct pathways to the amygdala prior to cognitive processing, and that emotion can be caused via this non-cognitive pathway alone (119). (The amygdala, as we will see below, is a region of the brain that is of critical importance in emotional memory and emotion regulation. More will be said on the nature of this pathway below, with respect to LeDoux’s studies on the same topic.) Zajonc argues that affective states can be induced short of cognitive processes, through the use of drugs such as valium for example (120). Zajonc admits that, “it is possible as Schachter and Singer (1962) have shown, that some qualities of the valium induced states may be altered by cognitive input” (120). He claims, nonetheless, that “in the final analysis, at least some very significant aspects of the change in the emotional state will be caused directly by the valium, regardless of what information the subjects are given and what justifications they themselves offer afterwards” (120). Thus, Zajonc maintains that cognition is not necessary for emotion. Zajonc is also well known for his exposure studies, in which subjects are shown to have preferential responses to stimuli that they were exposed to at a rate which cannot be recalled

or reported by the subjects. He claims that “Preferences for stimuli (tones, polygons) can be established by repeated exposures, degraded to prevent recognition (Kunst-Wilson and Zajonc, 1980; Takenishi, 1982; Wilson, 1979)” (Zajonc 120). Zajonc even claims support for his non-cognitivism in the work of Lazarus and McCleary, whose studies he takes to show that people can “make autonomic discriminations” between even “nonsense syllables” (120). Whereas Lazarus infers that such discriminations imply that “some form of appraisal occurred prior to the emotional excitation,” Zajonc responds that such an appraisal cannot be proven (120). To Zajonc, these studies only prove that emotions exist prior to cognition. Of course, one might respond that such affective responses to drugs and unrecognizable exposures are not genuine emotions, because they do not entail cognition. Zajonc, though, argues that framing the issue in this way is “circular” in that it cannot possibly be “falsified” (117). Whatever bodily response occurred is taken to be caused by cognition by the mere fact that it was a response. In this case, there is no way to prove whether cognition is necessary for a response. Zajonc maintains that, if we avoid circularity and look only to the empirical data, we find what he calls a “Primacy of Affect” over cognition. Emotions, in his view, require no cognition, but rather an affective response to the recognition of stimuli.

Neuroscientist Joseph LeDoux has published results of his studies on the inner workings of the emotion process. Whereas Damasio focuses on the brain’s mapping of bodily changes, LeDoux focuses his work on distinct pathways in the brain that impact (change) the bodily messages received upon the perception of stimuli. Upon this data generating initial changes in the body, the lower level evaluations in the brain then reappraise the data based upon information received upon the bodily changes and higher level thought processes. Thus, an emotion ultimately entails feedback loops between different levels of information processing. Based on LeDoux’s studies (most famously with rats), he identifies neural circuits of distinct pathways in the brain system that cause and modulate fear (150-159). As we will later see, Robinson speculates that


LeDoux’s studies will have implications for the study of emotion in general.

LeDoux argues that, in cases of fear at least, the brain has two pathways that, via the amygdala, trigger physiological changes. He calls these pathways the “the Low and the High Roads to the Amygdala” (164). In studies on fear in rats, LeDoux found that the initial perception of stimuli (such as the sound of a buzzer) is received by the thalamus and then transported to the cortex (which, itself, interprets the data and then transports it to the amygdala as well), but he found that data is also sent to the amygdala directly by the thalamus (by-passing the cortex). LeDoux describes these two pathways as the “thalamo-amygdala route and cortico-amygdala” route (165). He explains that the cruder, and more immediate, pathway from the thalamus to the amygdala in cases of fear is then followed with the finer-tuned recognition of stimuli provided to the amygdala via the cortex (150-163). The physiological changes stimulated in cases of fear are caused by the amygdala; it is the amygdala that generates outputs to the body that control “freezing, blood pressure, stress hormones, and startle reflex” upon recognizing stimuli as emotionally relevant (160). The amygdala is largely responsible for emotional memory, and upon receiving the representation of stimuli which is found to carry an automatic emotional memory, the amygdala immediately elicits bodily responses. The thalamo-amygdala pathway provides the amygdala with information directly and very immediately in modular form – taking approximately “twelve one-thousandths of a second” in rats (163). The thalamus is involved in sensory perception and motor functions, though its encoding of information is much less detailed than the information carried by the cortex to the amygdala. The cortex encodes data in more detail, occurring at a speed “almost twice as long” (163). LeDoux illustrates the distinction between these two pathways as follows:

(S)uppose there is a slender curved shape on the path. The curvature and slenderness reach the amygdala from the thalamus [at which time the body immediately defends against a snake], whereas only the cortex distinguishes a coiled up snake from a curved stick. If it is a snake, the amygdala is ahead of the game. From the point of view of survival it is better to respond to potentially dangerous events as if they were in fact the real thing than to fail to respond. The cost of treating a stick as a snake is less, in the long run, than the cost of treating a snake as a stick. (165)
LeDoux’s findings with rats show that, not only do rats have fear responses through both pathways, but *both pathways are receptive to learned stimuli*. That is, LeDoux discovered “a pathway that could transmit information directly to the amygdala from the thalamus [which] suggested how a conditioned fear stimulus could elicit fear responses without the aid of the cortex” (158). Thus, even very crude and modular emotional responses (such as certain instances of fear) can be learned responses to new stimuli. Furthermore, such emotional responses do not rely upon “explicit or declarative memory system” but rather upon an “emotional memory” (182).

LeDoux is not arguing that fear results from but these two pathways. Other pathways provide data to the amygdala as well. He explains that the hippocampus serves a distinct and later role as well in providing a contextual appraisal of the stimuli to the amygdala (167). The hippocampus recognizes and maintains memories of contexts; it does not identify individual features, but rather pools of features that are emotionally relevant. Thus, he explains how being in the place where a traumatic event took place (even when the threat is no longer directly present) can bring back emotional responses (167); the context is recognized through the hippocampus, and this data is transported to the amygdala which generates bodily responses. He explains that “The amygdala is like the hub of a wheel. It receives low-level inputs from sensory-specific regions of the thalamus, higher level information from sensory-specific cortex, and still higher level (sensory independent) information about the general situation from the hippocampal formation” (168). A more complete understanding of the neural pathways that explain a fear response generated in the amygdala, though, should recognize that the amygdala receives input not only from the thalamus, the cortex, the hippocampus, but also the “rhinal (transition) cortex” (which retrieves explicit memories) and the “medial prefrontal (extinction)” (which can weaken a response to threatening stimuli upon repeated exposure) (170). LeDoux’s conclusion, clearly, is that there are many neural pathways that contribute to emotional response and that a full understanding of an emotion will consider the feedback loops between distinct neural pathways.

Like Damasio, then, LeDoux focuses on an understanding of the interactive bodily processes which cause (and recall) the distinct homeostatic states involved in each type of emotion. LeDoux’s work, though, focuses more upon the various levels of appraisals (and feedback loops) than the mapping of bodily change. He argues that we are best to study, not emotions as one phenomenon, but as distinct emotion processes since (in his view) they have independent neural pathways which function to impact different bodily states (16).
The work by such scientists as James, Damasio, Zajonc and LeDoux are taken, by Robinson and Prinz, to offer empirical support for somatic theories of emotion. Robinson and Prinz use such data to support their theories that define emotions (at least in part, anyway, for Robinson) as noncognitive bodily appraisals – which we may or may not recognize in any declarative way.

5.3 Robinson’s “Process Theory” of Emotion (Weak Somaticism)

Robinson argues that it is important to conceive of an emotion as an episode rather than a static state, and thus she frames the emotion as a distinct “process” (57, 59). The experience of an emotion on any given occasion will not just exist in one way because (at the very least) the physiological arousal present in the emotion changes throughout the course of the experience. Robinson argues that an emotion is a process that entails three stages:

Generalizing from LeDoux’s results, then, we can say that in the simplest, bare-bones case of an emotion process, there is (1) an initial affective appraisal of the situation that focuses attention on its significance to the organism and causes (2) physiological responses of various sorts – especially ANS activity and changes in the facial musculature – and motor responses, which get the organism dealing with the situation as very broadly appraised by affective appraisal, and which give way to (3) a further more discriminating cognitive appraisal or monitoring of the situation. In other words … it is a process, a sequence of events. (59)

While a noncognitive appraisal is taken to initiate physiological changes, cognitive appraisals are taken to then monitor this initial affective appraisal. Robinson argues that emotion includes a noncognitive appraisal that causes the bodily responses, and thus an emotion is constituted by the combination of an affective appraisal, bodily change, and the cognitive reappraisal of the object. This second stage of an emotion is bodily change, and Robinson claims that this change “gets my body ready for appropriate action” (59). While Robinson’s theory here denotes actual physiological changes that take place in an emotion, perhaps her theory is also quite consistent with the view that such body-states could be but mapped in the brain through Damasio’s “as-if loop” as well. Robinson holds that a cognitive analysis of the situation (occurring after the initial
affective appraisal and bodily change) then monitors and changes these bodily responses through “feedback loops” with the affective appraisal that causes the physiological responses (76, 86). Through these feedback loops, the cognitive appraisals thus monitor changes. While she offers a somatic theory of emotion (in that she believes bodily changes can occur in an emotion without cognitive activity), I would note that her theory is a weak (or soft) somatic theory in that it includes cognition in the emotion as well. Robinson makes a unique case for regarding emotion as this specific three stage process and, in doing so, she offers both intuitive and scientific support. In this section, I will summarize her theory so that we can then consider how it pairs up with the norm-expressivist’s conditions for a theory or emotion.

Robinson believes that we can intuitively recognize the reality of an emotional process that involves different levels of appraisal, in that “We infer that a cognitive appraisal or reappraisal occurs because our behavior changes – from sitting quietly to bursting into tears, from freezing to fleeing, from an incipient movement of flight to sitting quietly again, depending on the circumstances” (76). This does not mean that we have two emotions, given the noted changes, nor that part of this is not an emotion (even though it still feels like it and science suggests it in her view). She finds such observations to illustrate that the same emotional episode includes such reappraisal. The reappraisal is cognitive, in that recognitions are taking place in a slower and more deliberated way.

Robinson specifies her meaning of cognition, claiming that “I think I am on the side of the majority in using the term ‘cognition’ only when we are talking about processes localized in the neo-cortex” (45). Cortical processing alone, then, does not constitute cognition. Information processing that takes place in the limbic system (which would include the hippocampus), the rhinal cortex (as a transition to the neocortex), and the medial prefrontal (which functions in rats, as well as humans, to extinguish threats) would then be non-cognitive processes according Robinson’s criterion, since none of these are localized in the neo-cortex; according to LeDoux (as seen above in 5.2), the amygdala can receive input directly from all of these processes and research demonstrates that these processes impact emotional regulation (LeDoux 170). Cognitive appraisal, on Robinson’s view, requires “higher processing” in the brain, and it modifies physiological arousal and behavioral tendencies through feedback loops with the affective appraisal re-

38 This research is supported by the work of Shin et al.
responses via these lower brain functions (45, 76, 86). Robinson intends for her processes theory of emotion to account for, not only the necessary affective qualities in emotion (as argued for in the subtraction problem by James, Section 5.2), and not only the necessary neurological system (as was argued for by Staats and Eifert), but also the cognitive evaluations that seem to account for adjustments in our affective responses.

Robinson aims to depict the affective appraisal in a way that might help to distinguish it from cognitive appraisal. She describes the non-cognitive affective appraisal as follows:

The answer suggested by Zajonc’s research findings is that there is an affective appraisal that concerns those things that matter to the organism and that occurs very fast, automatically, and below the threshold of awareness. This affective appraisal is non-cognitive in that it occurs prior to and independently of any cognitive evaluation. It serves to pick out and focus attention on those things in the internal or external environment that matter to the animal or human being and to appraise or evaluate them in terms of how they matter, for example, whether they are a good or bad thing, a threat or an offence. (42)

Robinson uses bold letters in her text in order to highlight the distinction between a mere perception or even a thought (e.g. that “I have been mistreated”) and an affective appraisal of the thought (“Offence!”) (90). Affective appraisals are not cool evaluations, but rather evaluations that have a phenomenological quality to them, and they result in bodily change. Here, Robinson argues that Zajonc’s studies evidence this affective non-cognitive appraisal:

Zajonc himself interprets his results as evidence that affect precedes cognition and that cognitive evaluation is therefore not necessary to emotion. Yet the data he cites can just as plausibly be construed as showing that at least some emotions involve primitive evaluations that occur below consciousness and independently of higher cognitive processing. Examples would include the preferences evinced by subjects in the mere exposure experiments, in which subjects “evaluated” the tones or polygons they had been exposed to more often as pleasanter than the others. Similarly, in the subception experiments, subjects “evaluated” some syllables as threats. (42)
Robinson holds that Zajonc is right to see such automatic responses to stimulus as both evaluative and non-cognitive. They are, she maintains, non-cognitive appraisals. According to Robinson’s theory, these “primitive evaluations that occur below the level of consciousness” just are the affective noncognitive appraisals, and she maintains that all emotions include such appraisal (42). Affective appraisals are taken to be unconscious, automatic, fairly “immediate” and relatively unsophisticated appraisals of stimuli that cause bodily change.

Robinson’s theory emphasizes her position that there are two types of appraisals that take place in an emotion – the noncognitive (affective) appraisal and the cognitive appraisal. She notes that LeDoux’s studies demonstrate “the brain circuits that are responsible for these two levels of processing [at the affective and cognitive levels], at least in the fear system” (51). LeDoux’s work, and Zajonc’s work as well, are taken to support making such general distinctions, and LeDoux’s work is taken to provide a specific example of one such mechanism.

5.3.1 Nature of Modular Affective Appraisals

Gregory Johnson cautions against any argument which uses the thalamus-amygdala pathway to explain a feature of all human emotions. He argues that the thalamus-amygdala “sub-cortical circuit cannot have a role in the explanation of emotions in humans,” and he particularly aims this criticism at Robinson’s illustration of the affective appraisal (739, 741). Johnson argues that, while the thalamus-amygdala route (the “Low Road” as LeDoux calls it) might work to explain some cases of modular and learned threat appraisals in response to sound and with respect to fear, this pathway would not be able to explain the causal mechanism for all emotions. Johnson reviews the properties of this neural pathway, arguing that the neural pathway involved in the thalamus-amygdala route would not be neurologically capable of carrying information about an object of one’s emotion. He notes that this pathway cannot recognize objects after all. One may try to argue that, though this route cannot recognize complete objects, it can nonetheless functionally perceive a threat. However, in Johnson’s detailed overview of the thalamus-amygdala process (in the auditory and visual system), it seems clear that the cells involved are not able to be even set off by complex stimuli (like complex cognitions such as belief). Johnson interprets the details of this process to demonstrate that “the cells in these areas of the thalamus that project directly to the amygdala are not set up to respond to stimuli that are very complex – a possibility that is essentially ruled out by the way in which stimuli are encoded at the periphery and then transferred to the
thalamus” (746). For this reason, he argues that “this pathway … does not have the capacity to respond to the types of stimuli that are generally taken to trigger emotion responses” (739). While we have evidence that sounds and sights can take this thalamus-amygdala route, we have no evidence that judgments can do so. Johnson reasons that if we assume that at least some emotions (if not all emotions) can respond to judgments, then we cannot adopt Robinson’s theory which purports that all emotions begin with a modular appraisal of this sort. After all, LeDoux chose to study sound on rats (which may also have systems quite different from humans) because he claimed that “the processing of sentences is a much more complex, and less well understood, brain operation” (148). Except for the most basic affective response to sights or sounds, many (if not most) emotional responses require higher order processing of the stimuli. Johnson maintains that a theory of emotion, then, cannot be modeled on the thalamus-amygdala because perceptual systems are unable to carry the type of information that would often constitute the object of an emotion.

I do not find Robinson’s theory to require that the affective (non-cognitive) stage of appraisal must always occur through the thalamus-amygdala pathway. There may be other ways in which one can make appraisals “very fast, automatically, and below the threshold of awareness,” aside from the thalamus-amygdala pathway. (Information processing in the neocortex would be excluded from this affective appraisal on her view, but not all cortical processing need be excluded.) She does note, after all, that LeDoux’s studies demonstrate that the “low road” (Section 5.2) is the thalamus-amygdala pathway and that the “high road” (Section 5.2) is the cortical-amygdala pathway in the fear system only (51). She does not claim that the thalamus-amygdala pathway is the means of every affective noncognitive appraisal. While (as LeDoux notes) the cortical-amygdala pathway takes twice as long as the thalamus-amygdala pathway, it is clear that two times as long as “twelve one-thousandths of a second” (the speed of the cortical-amygdala path) would be still be quite automatic and far below the threshold of awareness (163). There is nothing, in my view, to preclude Robinson’s theory from adopting a meaning of modularity which is in some way broader than the thalamus-amygdala pathway alone. (Prinz, as we will see, does just this.) So long as this appraisal results automatically and rapidly in physiological change (or perhaps the perception of it anyhow), and so long as it functions to appraise the relation of the stimulus to one’s well-being, it is an affective appraisal under her view. In the spirit of charity, then, we can take Robinson’s theory to propose that the thalamus-amygdala path-
way is but one example of how such non-cognitive affective appraisals can take place.

Still, Robinson’s theory may have some ambiguity about the exact nature of these affective appraisals. She states, for example, that “I will call such appraisals ‘affective appraisals,’ and I shall assume that they are ‘non-cognitive,’ in the sense that they occur without any conscious deliberation or awareness, and that they do not involve any complex information processing” (45). Later, though, she admits that “It is not surprising … that typically human cognitive appraisals and reappraisals, like affective appraisals, take place below the level of consciousness” (76). If she admits that affective and cognitive appraisals can take place below the level of consciousness, then we cannot use that as the basis to distinguish them from one another. Cognition, as noted earlier for Robinson, takes place in the neocortex (beyond, then, such brain functions found in the thalamus, hippocampus, rhinal cortex and medial prefrontal), and thus it involves higher level processing in this sense, although it need not entertain declarative thought. (The neocortex, also found in other mammals, is responsible for linguistic thought, although it is also involved in other higher functions such as spatial reasoning and other forms of conscious thought.) What would seem to distinguish the affective appraisal from the cognitive appraisal for Robinson, given her discussion of Zajonc and LeDoux’s work above, is its relative speed (though the details on this speed are left rather open on her account), its direct cause of bodily change in contrast to higher cognitive thought which need not (directly) cause physiological change, and (as we shall now see) its function in terms of Lazarus’s dimensional appraisal theory.

5.3.2 Robinson & Lazarus

Robinson’s theory is likened to Lazarus’s theory in that she takes the initial evaluation of an object to carry into the physiological state of the emotion. Thus, she takes there to be stages (or, as Lazarus put it, “dimensions”)\(^\text{39}\) to emotive evaluations. Robinson proposes that the affective and cognitive appraisals function to make dimensional appraisals likened to the primary and secondary appraisals proposed by Lazarus (as discussed in Chapter Four, Section 4.5.2). She recognizes that her theory is not consistent with Lazarus’s theory, in that she denies that the affective appraisal entails cognitive processing. Her intent is not to make her theory consistent with Lazarus’s theory in all ways. Nonetheless, she finds that her theory can account for Lazarus’s general findings with respect to

---

\(^{39}\) See Chapter Four, Section 4.5.2.
primary and secondary appraisals and she even uses this distinction to explain the roles of affective and cognitive appraisals in her theory. Her claim is that the different types of appraisals function to make the different types of evaluations as proposed by Lazarus.

While the initial affective appraisal is not taken to be a form of cognition on her view, she likens these appraisals to Lazarus’s primary appraisals in that his primary judgment also functions to determine the “relevance” of the stimuli to one’s “well-being” (199). Robinson describes Lazarus’s dimensional appraisal theory as follows:

On the one hand there are ‘primary appraisals’ that address ‘whether and how an encounter is relevant to a person’s well-being’ and are ‘based on the personal relevance of what is happening, which in turn depends on goal commitments and transactional stakes in a particular environmental context’\(^40\). In my terms these primary appraisals play the role of the initial process as a whole. Indeed Lazarus recognizes that these primary appraisals may be rapid, automatic, and unconscious. (198-199)

Robinson notes that Lazarus would likely regard her affective appraisal as a form of cognition, although he would certainly note that this cognition (in his view) occurs at a much faster speed than higher cognition and this lower cognition always occurs below the threshold of consciousness.

Robinson likens the cognitive monitoring stage of an emotion to Lazarus’s secondary judgments (199). She explains these appraisals as follows:

On the other hand there are ‘secondary appraisals’ [in Lazarus’ view] that have to do with the options for coping and expectations.’ The three components of secondary appraisals are said to be ‘blame or credit, coping potential, and future expectations.’ (199)\(^41\)

Robinson proposes that, in the cognitive monitoring stage of an emotion process (through which cognitive appraisal and reappraisal of the circumstances occur), we make secondary appraisals that function to evaluate such concerns.

Though there is disagreement between various dimensional appraisal theories regarding what the dimensions are in various emotion

\(^40\) Robinson cites Lazarus, *Emotions and Adaptations*, 145.

types. Robinson does not find this information exceedingly relevant to her theory. Even with respect to some of the details in Lazarus’ theory, she claims, “We do not need to examine Lazarus’s theory in detail. The aspect I want to focus on is his idea that emotions involve a sequence of appraisals, in which a primary appraisal of how some situation is relevant to our well-being gives way to a reappraisal or secondary appraisal of that situation as appraised by the primary appraisal” (199). Robinson purports that emotions result from an organism’s ability to identify in some automatic, non-cognitive way emotionally competent stimuli. For Robinson, this ability just is the affective appraisal. The result is a bodily change (or at least the perception of such bodily change). The secondary appraisal may consist in reflections on any number of factors related to one’s coping mechanisms, future expectations, or considerations of blame and credit (or perhaps even other points of deliberate reflection). With respect to other dimensional appraisals, such as that proposed by Ortony and Clore, Robinson notes that these theories also note that emotions are evaluations of matters of concern to one’s well-being in specific ways, followed by more nuanced evaluations (26). Robinson takes the affective appraisal stage of an emotion to establish the “concern” that an emotion entails, whereas the cognitive monitoring stage would seemingly be capable of considering different types of features (with whatever logical structure they may have) (41).

Should a hybrid cognitivist theory of emotion include somatic change within the emotion, and should it also include such affective appraisals within its treatment of “cognition,” then one might suggest that such cognitivist theories are quite similar to Robinson’s theory. It is perhaps difficult to clearly categorize Robinson as a cognitivist or a noncognitivist. In some ways, Robinson would seem to be a cognitivist in that she includes cognition in emotion and in that she recognizes that her proposed “non-cognitive appraisal” is utilized in other cognitivist theories (at least as proposed by Lazarus). On the other hand, Robinson aims for her theory to distinguish between the immediate appraisal and the higher cognitive appraisals, and she argues that affective appraisals are distinct from cognitive evaluations (so distinct that they are not even describable in propositional terms) (62). Though she admits that Lazarus recognizes this low-road appraisal to be a type of cognition, she maintains that this is a non-cognitive appraisal. She claims that “I will call such appraisals ‘affective appraisals,’ and I shall assume that they are ‘non-cognitive,’ in the sense that they occur without any conscious deliberation or awareness, and that they do not involve any complex infor-

mation processing” (45). Because of this, as well as her focus on the somatic change in emotion, I will regard her as a somaticist although I would suggest that she recommends a weaker form of somaticism than does Prinz as we will see.

Lazarus’s theory of emotion may appear to be quite compatible with Robinson’s weak somaticism. Emotions, on both accounts, are a combination of bodily change (which include action tendencies) and dimensional appraisals on both views. Though Lazarus maintains that adult human emotions utilize a distinct form of (higher) cognition which is unavailable to infants and other animals (as discussed in Chapter Four, Section 4.5.2), his analysis could perhaps be modified (“amplifying Lazarus’s theory” as Robinson puts it) in a way more consistent with her process theory (68). This proposal is perhaps not so objectionable to Lazarus. Robinson notes that, in response to LeDoux’s work in demonstrating the automatic thalamo-amygdala pathway, Lazarus admits that such (high road and low road) processes would then be but forms of cognition. Here, Lazarus shows his consideration of such possible forms of cognition:

If, indeed, one takes seriously that there is more than one kind of cognitive activity in emotion, the automatic and the deliberate, one would no longer need to equate cognitive processing with the relatively slow … or with the deliberate, volitional, and conscious reasoning that is often found …, but would recognize that emotional meanings can be generated in more than one way. I believe this possibility strengthens my position that appraisal is a necessary and a sufficient condition of emotion. (“Cognition and Motivation” 359)

He claims that “Cognitive psychologists of all stripes are beginning to take seriously at least two modes of meaning generation, one that operates automatically and without volitional control, the other deliberate and volitional” (“Progress on a Cognitive-Motivational-Relational” 824). Still, while emotions may be generated in both ways, Lazarus argues\(^\text{43}\) that adult emotions are distinct from other animals and infants because adult human emotions are caused by and constituted by higher cognition. Robinson’s “noncognitive affective appraisal,” may well be addressing a function that cognitivist theories identify as emotional cognition, since she explicitly includes higher cognition in emotion processes as well as lower forms of cognition in all emotions. Lazarus’s theory does not spec-

---

\(^{43}\) See section 4.5.2., in Chapter Four.
ify that all emotions must include both forms of cognition. To the contrary, adult human emotions are generally taken by Lazarus to be forms of higher cognition. He proposes a typology of adult human emotions couched in terms of higher cognition, with both the primary and secondary appraisals being regarded (in adult human emotions) as a form of higher cognition, whereas Robinson’s theory has no such implication. Robinson’s theory is distinct in that she specifically includes lower (affective/non-cognitive) appraisals in all emotions, whereas Lazarus’s theory would seem to but accept the possibility.

Unlike Lazarus, Robinson argues that an emotion is not a judgment, but rather a series of appraisals (only some of which are cognitive judgments) in combination with bodily change. She makes the bold claim that we cannot even describe, in an accurate way, the actual affective appraisal because this appraisal is not itself a propositional judgment (62). We may be able to say how the affective appraisals function (in terms of primary appraisals), but this does not make the affective appraisal a primary judgment in Robinson’s view. She maintains that it is misleading to characterize emotions as propositional judgments due to their nature as processes which include bodily response and affective appraisals. Still, while Robinson does not take propositional judgments to constitute emotive appraisals, she notes that cognitive acts may be necessary (but not sufficient) to the identity of some propositional judgments in characterizing the identity of some emotion types.

5.3.3 Robinson’s Account of Emotion Typology

Robinson argues that emotion types can probably be best understood as a variation of “basic emotions” (68). She finds that, based on research, “there are a small number of basic emotion systems, usually identified by means of their uniquely distinguishable physiological symptoms” (68). She argues that the diversity of basic emotions (such as happiness, sadness, anger, surprise, disgust and fear) provides evidence that affective appraisals allow for more diverse evaluations than merely positive and negative ones. Robinson maintains that affective appraisals are not propositional judgments, and that cognitive acts are necessary but not sufficient for the identity of some emotion types.

44 Interestingly, some cognitivists (like Salmela) would apply a Dretskean analysis when characterizing emotion types in terms of function as well. Perhaps it would amplify Robinson’s theory to use a Dretskean analysis to support her methodology on this point similar to cognitivists (Salmela, “Can Emotions be Modelled on Perception?” 9). Salmela argues that, in keeping with Dretske’s point, if an analysis functions in a way that can be described in propositional terms then there must be some actual propositional structure to the analysis. If this argument is right, then it would seem that the affective appraisal is describable in propositional terms (and perhaps that it even is a propositional judgment in this sense). Robinson, though, would have to (somehow) reject this part of such an analysis of course.
negative responses (65).\textsuperscript{45} Robinson argues that basic emotions evidence distinct basic evaluations, in that these bodily changes would be caused by affective appraisals which are much more rough and ready than we find in the evaluations made through cognitive monitoring, and yet these basic emotions are more differentiated than merely positive and negative effects. She proposes that “affective appraisals might be part of what Ekman calls the ‘affect programs’ that identify the basic emotions” (see Chapter Three, 3.3.1) (68). Thus, since all emotion types include an affective appraisal, emotion types are at least derivatives of these basic emotions. As she notes, “Clearly, there are not different emotion systems for every emotion named in the English language” (69).

Non-basic emotions are identifiable by their distinguishing forms of cognition – whether this cognition is causal and/or constitutive at the stage of a cognitive appraisal. She herself considers that “anger, guilt and shame”\textsuperscript{46} may carry a distinct relationship with cognition (200). She supposes that cognitively complex emotion types may require some sort of distinct belief or thought as their cause, and/or it may be that these emotions require such distinct thoughts during cognitive monitoring. She provides the following argument:

We know that rough and ready, quick and dirty evaluations are enough to trigger emotional responses, but at the same time many human emotions seem to have cognitively complex content: they are reactions to beliefs about morality and politics or responses to women in authority or my friend’s winning a prestigious prize. And it also seems as if emotions are distinguished one from another by means of these complex cognitions, since rough and ready, quick and dirty evaluations are not enough to distinguish shame from guilt or jealousy from envy. (59)

Distinct forms of cognition distinguish non-basic types of emotion from one another. Still, this cognition should be understood in a broad way – as implying a thought, at least, but not necessarily a belief (91). Her reference to such cognition tends to be quite inclusive. She states that “More complex cases of emotion in human beings might involve affective response not to a perception but to a thought or belief, and the cognitive monitoring may be correspondingly sophisticated…” (59). Bearing that in mind, the difference between basic emotions and cognitively complex emotions is that in cognitively complex emotions there is a spe-

\textsuperscript{45} She borrows the concept of “basic emotions” from Ekman’s article “Constants Across Culture” (127).

\textsuperscript{46} Here, Robinson cites Lazarus, Emotions and Adaptations, 112.
cific type of cognitive cause (as the stimulus for the emotion) and/or there is a specific sort of cognitive appraisal taking place in the cognitive monitoring stage of the emotion type.

With respect to non-basic emotions, it may be that certain emotion types are always responses to specific forms of cognition. Perhaps, for example, certain types of beliefs (or ways of seeing things) are necessary for certain types of emotion. Robinson notes that “In a case of emotional resentment, for example, the thought that I have been badly treated in a situation in which I have been led to believe I will be well treated may be part of the causal chain leading up to an affective appraisal of Offence!” (90). Even when a specific type of cognitive activity is a necessary cause for an emotion type, though, Robinson maintains that “thoughts and beliefs and ways of seeing a situation can figure in the causal chain that leads to an affective appraisal, but they are never the whole story” (91). While certain emotion types may require cognitive activity as their causes, the emotion itself still begins with an affective appraisal of that cognition (62). The emotion itself begins not with the belief (or thought or way of seeing a situation), but rather with the appraisal that this thought (about the mistreatment) indicates an Offence!

Still, it is possible that some emotion types are best identified according to their cognitive antecedent.

On the other hand, Robinson proposes that even when certain cognitively complex emotions respond to a variety of initial objects (objects that may or may not be specific forms of cognition), emotion types may also be distinguishable according to their secondary appraisals in the cognitive monitoring stage. She proposes that, “some emotions can be distinguished chiefly by differences in secondary appraisals” (202). Robinson provides the following example of such possibilities:

Thus both fear and anger may occur when someone has done something to us which we regard as a wrong or a threat. Fear is associated with an appraisal that one is unable to deal with the situation, as when I have been insulted by my boss, and anger is associated with an appraisal that I can deal with the situation, perhaps by exacting revenge, as when I have been insulted by an underling or subordinate. Fear and anger are usually proposed as totally distinct emotional states, but this analysis suggests that they might be quite closely related …. How I emerge at the end of the process will crucially depend on how successfully my coping mechanisms – avoidance, denial, and the rest – have functioned. (202)
In both cases, there is an original (rough and ready) affective appraisal of a circumstance (such as an off-putting comment). Perhaps the affective appraisal is of “An Accost!”, which then causes the physiological changes (typical of such alarm). Very quickly, though, cognitive processing will turn this upset into either fear or anger. Because emotions include bodily states as well (as basic emotions, in Robinson’s view, or perhaps even as “blends” of basic emotions as Ekman calls them), emotion types can be identified (and defined) according to their bodily states and/or their distinct cognitions (88).

Thus, Robinson’s theory proposes that that we ultimately categorize emotion types according to the combination of their bodily states (their basic emotions) and, in some cases, their distinguishing cognitive causes and appraisals (36). While proposing that emotion types may be identified in any of these ways, her theory of emotion does not specify a clear structure or mechanism by which to conceive of emotion concepts. She focuses her work more on the nature of emotion rather than emotion typology. In terms of the independent merits of her theory, it is probably a strength (as it was with other theories, such as Solomon’s engaged judgmentalism for example) that her theory is loose enough to accommodate different proposals for emotion typology. We should consider, though, whether her framework provides an adequately clear answer to the nature of moral emotions.

5.3.3.1 Robinson’s Account of Moral Emotions

As noted, Robinson suspects that shame and guilt have cognitively complex content (59). In this case, they are non-basic emotions. Nonetheless, if the facts indicated otherwise, then Robinson’s theory would of course allow for moral emotions to be basic emotions. In that case, anger and guilt would need to have their own emotion-specific physiology, and these emotions would be bodily responses to the initial non-cognitive appraisal of a variety of cues (both judgments and percepts) that elicit these responses. Because basic emotions are found in other primates, if guilt and anger are basic emotions, then the stimuli that elicit guilt and anger would have to include both percepts and judgments (Ekman “Basic Emotions” 56). (Cognitive monitoring would, on her view, follow of course.) If the moral emotions were basic emotions, then their distinct bodily responses would be triggered by an automatic (non-cognitive) appraisal of certain types of objects (or objects that are somehow appraised in such a way anyhow), for on her theory basic emotions function to appraise (in a rough and ready way) different sorts of relevance in the object of the emotion (e.g., danger, offences, surprise, etc.). (Basic emotions cannot, then, require higher order thinking or complex judgments as
their *cause*.) In this case, guilt and anger would be the automatic appraisal (of circumstances) and the corresponding physiological changes that distinguish the moral emotion types in this case, rather than higher cognition. Perhaps automatic bodily responses can occur upon the affective appraisal (in Robinson’s view) of specific, and morally relevant, objects independent of higher level thought. Robinson, though, offers no argument for this being the case in moral emotions.

While Robinson is comfortable supposing that anger is a basic emotion, with its own typical physiological state, she admits that it is quite questionable whether many other emotions (including “shame”) have their own distinct physiological states and affective appraisals (69). She proposes that guilt and shame are distinguished by their (either causal or constitutive) necessary cognition. While she takes anger to be a basic emotion (thus carrying a distinct physiological state), Robinson again cites Lazarus for support in maintaining that cognitive reappraisals particularly shape anger, guilt and shame (200). Though there is no mention of “disdain” here, it would seem that her logic follows equally for disdain. (Granted, though, that is yet another empirical question.) We should consider how the moral emotions are to be explicated, on her view, in the case that they are not basic emotions.

If moral emotions are non-basic emotions, then there are two ways in which they may be distinguished as cognitively complex emotions, and it may even be (on her view) that some combination of the two determines their identities. On the one hand, these emotions could be distinguished by their objects. If the object that elicits these emotions is always a specific thought (e.g. a belief or desire about a specific propositional content), that may distinguish these emotions from one another. Perhaps guilt is always a response to entertaining the thought of a personal transaction, or *seeing things as if* one’s actions are of the sort to meet a damaged personal relationship. In this case, it could be (as it could be with the attributional theories) that the thought which causes the moral emotions has a sort of propositional content that is describable in non-moral terms. That is, as discussed, one might appraise the belief that one has acted in such a way (so as to meet a damaged communal relationship) as indicating something like “**Bad Interpersonal Behavior (by Oneself)!**”. Many actions by a person could be interpreted in terms of such a thought and then appraised to be a matter of concern. A child might, in some way, entertain the thought that her failure to practice piano today was a “bad interpersonal behavior (by oneself)” in that she is aware that it will cause a rift with her parent that evening. Thus, she be-

---

believes (or sees things as if) the failure to practice is the kind of behavior that tends to meet such a riff, and (if she feels guilty about it) she then appraises this thought as a “Bad Interpersonal Behavior (by One-self)!”. One would not then just find that one acted in such a way, but one would (upon appraising it) find it to matter in this specific (concerning) way. Without the appraisal of this thought content, the child may realize that her choice not to practice was a behavior of the sort to meet a disconnection, or riff, with the parent, but just not care. The affective appraisal involved in guilt could be the appraisal, that is, that one’s thought content (that one acted in a way that tends to meet a damaged personal relationship) is now of concern. Thus, if guilt and anger are caused by specific thought content, then it is plausible (as it was with cognitivist theories) that this thought content can be conceived of in non-moral terms.

On the other hand, it may turn out that these emotion types are distinguishable from other emotion types according to some unique cognitive appraisals that occur after the bodily change begins – that is, in the cognitive monitoring stage of the emotion process. In this case, these emotions would clearly be non-basic. It may be that the moral emotions (or guilt, shame, and disdain anyhow if we grant that anger is a basic emotion) are non-basic emotions and they require no special thought content as their cause. In this case, these emotions may respond to percepts as well as judgments and that they would share body states with other emotion types. If this is so, then it can only be the cognitive monitoring that determines the emotion type. Nonetheless, there is nothing in Robinson’s theory to require that such cognitive monitoring in these emotions must employ moral concepts. While Robinson admits that we can only speculate about the exact nature of appraisals, we can recall that she proposes that the cognitive appraisals function to make Lazarus’s secondary appraisals (76). Thus, she takes secondary appraisals to evaluate “blame or credit, whether he or she can cope with the situation, and what his or her future expectations are” (75). Considering the way in which the cognitive appraisals are taken to function, and the fact that these functions can be realized in a pre-linguistic and unconscious way, I

48 Of course someone may respond that such an example is – if anything – more likely to cause fear rather than guilt, but (based on his response to Sinnott-Armstrong’s critique) Gibbard would likely note that we can recognize the difference between such emotions in part due to their distinct phenomenological qualities. Fear feels different than guilt. In noting this recognizable difference between basic body states, Robinson would also be in keeping with Prinz’s argument in highlighting the basic body state (of sadness) that is then cognitively elaborated upon in forming feelings of guilt in particular. Surely, some children are not so much frightened in such situations as they are recognizably sad at the disappointment of a parent.
do not find that these cognitive appraisals *need* be moral judgments. As argued earlier (Chapter Four, Section 4.5.3), it is plausible that other non-moral judgments may serve to explain one’s assessment of responsibility for an act without employing the concept of blame. Perhaps such judgments (at the cognitive monitoring stage) focus on the directness of causal locus in specific (and descriptive rather than moral) terms. We may have thoughts that function to assess features that turn out to be relevant to an assessment of blame and credit, even though these features (themselves) are not assessed in terms of blameworthiness. Perhaps guilt entails the cognitive appraisal that “I cannot easily repair this harm to the person I injured” (thus reflecting an evaluation of my ability one to cope with, or address, the transgression) and/or “It is highly likely that these actions will meet a damaged communal relationship” (thus addressing future expectations relevant to the appraisal).

Whereas a moral judgment is something that we would seemingly (if Gibbard is correct) be willing to defend in moral discourse with others (if uncensored), a cognitive appraisal on Robinson’s view need not be of this sort, for it may be processed using mechanisms that are (like the affective appraisal) below the threshold of such declarative reasoning. Norm-expressivism, of course, proposes a very narrow and stipulated meaning for the moral judgment (per Chapter One), and this meaning entails an acceptance of the moral norm as derived in response to normative discussion and a willingness to defend it in response to demands for consistency. Clearly, that is not the type of appraisal that need take place in the cognitive monitoring stage of a moral emotion. A cognitive appraisal need not be something that one can accurately describe in linguistic terms at all – which is why the emotions of other animals also include this stage of cognitive appraisals on Robinson’s view (51, 54, 58). If there are brain structures, which operate prior to the neocortex that can process data *relevant* to blame (such as perhaps tracking one’s level of effort in avoiding harm to others and/or the degree of suffering that we identify in others), then it would seem that we may in this way cognitively appraise circumstances that are relevant to blame without making an actual moral judgment (in the norm-expressivist’s sense) with respect to *blameworthiness*.

It may turn out that the moral emotions are basic emotions – with their own distinct, unique and universal physiological states. If they are not basic emotions, then their typology will be explicated in terms of their necessary cognitive evaluations (whether causal or constitutive). Either way, the moral emotions will consist of non-cognitive appraisals, physiological changes, and cognitive monitoring. If it is possible to explain the cause and mechanisms of moral emotions without the use of
moral judgments (through non-moral descriptions of any such thought content or the way in which the emotions function), then it would appear that we can identify the emotion concepts (of guilt and anger) without necessarily employing these moral concepts. Still, while such arguments may be made, they are not specifically made by Robinson. After all, it is plausible for her theory to be embellished in a number of ways.

5.3.4 Would Robinson’s Process Theory Suffice for Norm-Expressivism?

If we embellish Robinson’s theory in some specific (and, yet, plausible) ways, then her process theory meets all of the norm-expressivist’s conditions for a compatible theory of emotion. This does not prove that her theory is an ideal theory of emotion for the norm-expressivist because the theory is only “ideal” for norm-expressivism if it turned out to, in fact, be a cogent theory upon an analysis of the independent merits (or demerits) of it and if we add the needed embellishments. As with Solomon’s engaged judgmentalism, her theory could be quite compatible with the norm-expressivist’s conditions without modifying any of her arguments. We need only add arguments that are quite consistent and plausible with her theory.

An understanding of the way in which her theory meets C1 reveals the way in which it can meet C3. With respect to C1, “The theory of emotion must provide a framework capable of distinguishing between guilt, anger, shame and disdain (and other broadly moral emotions).” As shown above, Robinson’s theory proposes several possible ways for distinguishing between emotions. If the moral emotion types are basic emotions, then they can be distinguished from one another (and other broadly moral emotions) through their distinct physiological states and the rough and ready primary (affective) appraisals that these states function to make. If the moral emotion types are non-basic, then they can be distinguished by their causal and/or constitutive cognition. If some of the moral emotions are non-basic (as is the case with guilt and shame in Robinson’s view), then they may at least be distinguished according to either their distinct cognitive causes (should the emotion turn out to require a specific cognitive activity, such as a specific belief or way of seeing things) and/or they may be distinguished from one another (which Robinson seems to find most likely) according to the content of their cognitive monitoring.

C3 requires that “The theory of emotion cannot imply that anger and guilt are, nor that they are necessarily responses to, moral judgments,
nor can moral emotions be defined in terms of moral concepts.” While anger and guilt may turn out (upon Robinson’s theory) to be distinguished through a number of ways, none of these ways requires guilt and anger to either consist of moral judgments, or to necessarily be caused by moral judgments, or to be defined according to moral concepts. Still, Robinson does not give arguments to this effect. Her theory does not require modification in order to meet C1 and C3 (that is, we need not change any of her arguments or positions), though her theory does require (as would Solomon’s engaged judgmentalism) that we embellish her theory with arguments that she does not make. It is certainly conceivable (on her view) that moral judgments are among the causes of anger and guilt, even though they need not be caused by moral judgments alone.

Her theory also provides an account of the relationship between emotions and behavioral tendencies, and in doing so it meets C4 and C2. C4 requires that “The theory of emotion must explain how specific types of emotions (e.g. guilt and anger) motivate actions.” Emotions do, as Robinson says, “get us ready for action” through physiological changes (which get us ready for action) as well as through the (both affective and cognitive) appraisals that modify these physiological states. According to C2, moreover, “The theory of emotion must provide a framework capable of explaining how some emotions (such as anger and guilt) tend to be stronger motivators of action than are other moral emotions.” Because all emotions include cognitive monitoring on Robinson’s view, and because certain cognitive acts can surely be more motivating than others, it is perfectly conceivable that (given her theory) some emotions will tend to be stronger motivators of action than are other emotions.

Because the moral emotions are subject to the influence of cognitive monitoring (as are all emotions) on Robinson’s view, her theory would also meet C5, C6 and C7. C5 would require that “The theory of emotion must explain how emotions can be caused by the acceptance of norms that we are more or less aware of at the time.” Robinson admits that both the affective appraisal and the cognitive appraisal within an emotion process may take place below the threshold of awareness. Emotions can, then, be caused by the acceptance of norms that we are more or less aware of at the time. C6 requires that “The theory must provide a framework that explains how broadly moral sentiments shape guilt and anger.” Though Robinson does not mention this explicitly, her theory would seem to be consistent with secondary (broadly moral) emotions influencing the way in which both affective and cognitive appraisals take place – thus impacting the very appraisals that constitute anger and guilt. For example, if one is a particularly sympathetic person, then one may
just tend to make the affective appraisals that (in part) constitute the emotion guilt in response to less serious harms, and this person may do so more frequently than someone who generally experiences less sympathy toward others. Our tendency to experience certain emotion types (such as sympathy) may well influence our tendency to find concern (affectively appraising) objects of concern to other emotion types as well (such as anger at less serious offences). C7 requires that “The theory of emotion must explain how guilt and anger respond to considerations regarding their consistency with other norms, their grounding reasons, and their plausibility.” Upon Robinson’s view, it would seem that considerations regarding an emotion’s consistency with other norms, its grounding reasons, and its plausibility would at least be entertained at the level of cognitive monitoring. Perhaps the affective appraisal level of an emotion is even particularly helpful in determining the plausibility of an emotion in response to a situation. For example, if someone does make the affective appraisal that someone else’s words are an “Offence!” then it would probably appear to be quite plausible to the person that an offence occurred. Still, the consistency of the emotion with other norms and a consideration of grounding reason would seem (on my view of it anyhow) to occur at the level of cognitive monitoring.

Robinson’s process theory of emotion is, then, consistent with the norm-expressivist’s seven conditions. None of this solves the question as to the independent merit of her theory, nor the merits of her theory in comparison with other theories of course, but her process theory would seem to provide a framework that is quite plausible and also as consistent with norm-expressivism as is paradigmatic behaviorism, Solomon’s judgmentalism and attributional theories. These theories are all consistent with the seven conditions, though the consistency in each case is demonstrated only by embellishing these theories with additional arguments pertaining to the identification of guilt and anger. Robinson’s theory, though, maintains the unique position that emotions all begin with a modular and non-cognitive appraisal of an object. Her theory maintains that physiological changes always occur in response to immediate evaluations without any higher level cognition. Unless it turned out that moral emotions are responses only to specific thoughts, and unless these thoughts must just be moral judgments, then this feature of her theory might strengthen the norm-expressivist’s claim that moral emotions need not be caused by moral judgments nor conceived of in terms of moral concepts. While Robinson leaves but this possibility, Prinz offers arguments to prove that the experience of moral emotions does not in fact require any such moral judgments (nor even the use of moral concepts) as its cause. The norm-expressivist will still have the need to iden-
tify moral emotion concepts in non-moral terms (in order to fully satisfy Sinnott-Armstrong’s critique\(^49\)), but Prinz will offer responses to that which also reinforce the ability of Robinson’s process theory (and other theories) to do so.

### 5.4 Prinz’s “Perceptual Theory” of Emotion (Strong Somaticism)

In his book *Gut Reactions*, Prinz argues for a perceptual theory of emotion based largely upon the empirical work of such researchers as James, Damasio, Zajonc and LeDoux (among others), in addition to his criticism of alternative theories.\(^50\) In accordance with Damasio’s findings, Prinz believes that an emotion is the mind’s perception of bodily change. Prinz also draws upon philosophical distinctions and Lazarus’s proposed core relational themes\(^51\) in psychology in order to account for the intentionalty of emotional episodes and the seemingly shared appraisals among emotion types. Prinz’s theory of emotion can be regarded as a somatic theory in that he finds the perception of bodily changes (whenever these responses represent core relational themes anyhow) to be necessary and sufficient for emotion. He maintains that even enduring emotions (such as long-lasting love and guilt, which may last for years) are explainable as dispositions toward such bodily perturbations in response to their objects (“Which Emotions?” 79-80). In this section, I will summarize the main tenets of his theory so that we can then consider how this theory meets the norm-expressivist’s conditions for a theory of emotion. As we will see, Prinz offers not only a framework by which to explain emotion typology in a way that is consistent with norm-expressivism, but he also offers explicit arguments regarding the nature of guilt and anger which should especially appeal to the norm-expressivist.

Prinz argues that cognition is neither necessary nor sufficient for emotion. When he claims this, as we will see, he means that “acts of cognition” are neither necessary nor sufficient for emotion (46). He does not deny that emotions require information processing, although he disputes that this truism proves them to be (or require) any cognitive act. Prinz rejects proposals that describe cognition in such broad terms as “transformations of inputs,” noting that “transformations of input occur

---

\(^49\) Review Chapter Four, Section 4.3.1.1.1.
\(^50\) Throughout the remainder of this chapter, all page citations from Prinz are from *Gut Reactions*, unless otherwise noted.
\(^51\) See Chapter Four, Section 4.5.2.
all the way through the nervous system” (43). (Even vision would satisfy this criterion, and vision is not cognition.) Such a description of cognition is, on his view, too broad. Prinz proposes that, “we call a state cognitive just in case it includes representations that are under the control of structures in executive systems, which, in mammals, are found in the prefrontal cortex” (47). Prinz argues that “A cognitive act is an act of generating a thought under top down control. Thoughts produced automatically in the course of perception are unthought thoughts. We have them without thinking. They are cognitions, but not acts of cognition” (46). When Prinz rejects cognitivist theories of emotion, he largely rejects the claim that emotions require an act of cognition in this sense. He doesn’t deny that there is a transformation of input, but he denies that emotion requires top-down control under the executive systems in the prefrontal cortex (47). He claims that “If emotions are cognitive, they must be under cognitive control” (49). Clearly, cognitivists may disagree with this in that they admit that the necessary cognition for emotion may be prelinguistic and below the level of consciousness. Still, if we accept Prinz’s definition of cognition as such, then we might more readily understand the sense in which he denies that emotions entail cognition.

Given strong evidence that affective responses can result from processing prior to the prefrontal cortex, Prinz concludes that emotions can occur prior to (and thus without) such cognition. LeDoux’s study with rats (discussed above) is but one example of bodily response to the perception of stimuli that occurs prior to any top-down conscious thought in the prefrontal cortex. Indeed, it occurs prior to any cortical processing. Prinz references LeDoux’s work evidencing this modular pathway of emotional response via the thalamus-amygdala pathway (34). Prinz notes, for example, that when a person sees a curled shape on the ground in the woods, the person need not think “that’s a snake”; the fear response in the body responds directly to a perception of the shape. The objects that elicit an emotional response may be acts of cognition (such as beliefs), but Prinz finds such cognition to be unnecessary to emotion since emotions can also be elicited by non-cognitive objects as well. That is, one could be frightened by the belief that one’s bank account is overdrawn, but one could also be immediately frightened by the perception of a snake-like shape. Prinz maintains that we can have emotional responses to stimuli (whether percepts or judgments), and he maintains that these objects of emotion elicit emotion when they are perceived through modu-

---

52 Mikko Salmela critiques Prinz’s narrow depiction of cognition. Salmela argues that many cognitivist theories of emotion frame cognition in a much broader sense, and thus he argues that Prinz commits a “straw man argument” in his critique of many cognitivist arguments (“Can Emotions be Modelled on Perception?” 8).
lar (immediate) processes in the brain (77). Thus, cognition is neither a necessary nor sufficient cause of emotion.

Likewise, Prinz argues that behaviors (or at least macro-level behaviors) are unnecessary to emotion. While he admits that emotions provide “a reason for action,” he maintains that this is different from “compelling us to act” (193). Anger may provide a reason to take revenge, but it does not compel one to take revenge. Because emotions do not compel us to act, Prinz maintains that actions are unnecessary for emotion. Emotions are neither actions nor “action tendencies” in Prinz’s view. He maintains that the tendency to act on an emotion occurs only upon a choice being made, and prior to someone choosing to act the emotion is but “action enabling” (194). That is, an emotion may provide one of many motivations to choose an action.

We can note that Prinz’s arguments against behaviorist theories of emotion (theories that claim emotion is identical to observable behavior, as seen in Chapter Three) are directed at claims about macro-level behavior (like taking revenge, or striking out). He need not deny that the perception of bodily change has its own behavioral manifestations on some neurological level. To define emotion as but such inner behaviors, though, would not seem to offer as complete an account for emotion as he intends. Prinz’s theory may be quite consistent with paradigmatic behaviorism (per Chapter Three, Section 3.4), but Prinz’s somatic theory proposes an analysis of representation and the causal mechanisms of emotion types that we do not find in paradigmatic behaviorism. Prinz’s theory would seem to amplify paradigmatic behaviorism, and in doing so Prinz’s theory addresses the limitation of paradigmatic behaviorism in meeting the norm-expressivist’s conditions.

5.4.1 Emotion as Perception

Prinz’s theory is a “perceptual theory” of emotion in two senses: emotions are taken to be direct perceptions of bodily changes, and they are taken to be indirect perceptions of their core relational themes (224-225). In keeping with Damasio’s as-if-body-loop findings, Prinz argues that emotion is the perception of bodily change (58). That is, Prinz claims that emotion is a perception of bodily changes – rather than actual bodily changes. Prinz’s theory is a perceptual one, not only with respect to his claim that an emotion is the perception of bodily change, but also because of his claim that emotions are perceptions of the “core relational themes” (to use Lazarus phrase53) that they represent (64-66, 224-225).

53 See Chapter Four, Section 4.5.2.
Obviously, emotions are not perceptions of core relational themes in all of the same ways in which we perceive visual or auditory objects. We ought not think of the perception of a core relational theme as the perception of a “mental image” or some such thing (230). Nonetheless, Prinz maintains that there are enough similarities between the way in which sensory perception occurs and the way in which emotional responses occur so as to warrant regarding emotions as perceptions of both bodily changes and core relational themes.

It is no doubt easier to understand the claim that emotions are perceptions of bodily change than it is to understand the claim that emotions are perceptions of core relational themes. Certainly, Prinz argues for this second claim in a qualified sense. Prinz argues that emotions should be conceived of as perceptions for two reasons: 1) “Perceptions must occur in a perceptual system,” a system which in the case of emotions just is the body’s responsiveness to the objects of emotion, and 2) “Perceptions pick up information in a distinctive way,” which in the case of emotions is the body’s ability to carry information about core relational themes (224). An emotion is the perception of a bodily change, and this (perceived) bodily state is set up to carry different sorts of information about one’s relation to the object of one’s emotion. The information that these states carry (as we will later see in more detail, in the next section) is none other than the emotion type’s core relational theme. If one is afraid of a shadow in an alley, for example, then the mind’s perception of one’s racing heart and grimacing face just is the evaluation of this shadow. These experiences represent the danger of the shadow to one’s own well-being, whether one is able to recognize one’s fear or not.

Prinz maintains that all perceptual systems (and thus emotional systems) are characterized by their modularity. He argues that the perception of these bodily changes is a modular response to the (perceived) object in an emotional experience. Thus, an emotion is the perception of a bodily change which directly carries information about its object. The bodily change tells us that the eliciting object of a given emotional experience (e.g. a snake or the judgment that one’s bank account is overdrawn) falls within a core relational theme. One’s racing heart and grimacing face can tell one that the shadow is a case of danger. Prinz argues that emotions are perceptions in that they take place via a modular system (being the immediate, and automatic, perception of bodily changes) and they pick up information in distinctive ways (in that these body states represent core relational themes).

Like Robinson, Prinz holds that the perceived bodily change, and thus the representation of information by such a state, is a modular response to the perceived object of an emotional experience. Prinz main-
tains that even “higher cognitive emotions are not entirely nonmodular. They contain embodied appraisals, and embodied appraisals are modular” (96). In comparison with Robinson, though, Prinz more distinctly argues that this modular response is not limited to one specific pathway, nor even one necessary feature. He emphasizes a broad conception of this modularity. Prinz explains that “modularity may come in degrees. A system that fails to satisfy all … criteria,\(^5\) or that satisfies them only to a limited degree, might be classified as relatively modular nonetheless” (233). Thus, whatever merit Johnson’s criticism of Robinson’s theory might have with respect to an overreliance upon the thalamus-amygdala pathway as a model of emotion processing (although, as noted, I think her theory can address such criticism), Prinz’s theory certainly presumes no such criterion.

Interestingly, Prinz argues that we can perceive core relational themes without employing (or even comprehending) such concepts. In perceiving a core relational theme, one is not making some sort of unconscious and abstract judgment. Prinz argues that diverse species and even young children can track such themes in a modular way independent of conceptual understanding. He explains that “Core relational themes are what our emotions represent. But core relational themes need not capture the inner structure of emotions or the inner structure of any mental states that lead up to our emotions” (66). Rather, Prinz argues that there are mechanisms that track core relational themes without employing any concepts in the process. Arguably, one can perceive the color red without grasping the concept of redness; so long as there are mechanisms which pick up and represent such information (visual systems in this case), the data is perceived. (After all, children are taught to identify colors, in terms of color concepts, only after they perceive them.) Our ability to perceive core relational themes implies, in Prinz’s view, that we have a mechanism dedicated to the perception of core relational themes. As he puts it, “Just as vision is used to detect objects through light, emotions detect relational themes through the body” (233). Prinz proposes that “Just as the visual system subdivides into hierarchical pathways for detecting color, form, motion, and position, the somatosensory system subdivides into pathways for detecting textures, shapes, temperature, injuries, and core relational themes” (225). One might wonder whether science would tend to support such claims. As LeDoux

---

\(^5\) Prinz cites J.A. Fodor’s criteria for modularity. Modules are said to have the following properties: They are localized, subject to specific breakdowns, mandatory, fast, shallow, inaccessible, informationally encapsulated, ontogenetically determined, and domain specific. (Prinz, 232) He cites: Fodor, J.A. *The Modularity of Mind*. Cambridge: MIT Press. 1983.
claims, after all, “The brain does not have a system dedicated to perception ... Each system evolved to solve different problems that an animal face. In a similar vein, the various classes of emotions are mediated by separate neural systems” (16). It would seem that Prinz must respond with a functional definition of such a perceptual system. That is, a system would be said to exist insofar as something must serve the common function of recognizing (in a modular way via perceived bodily changes) the core relational themes that promote our well-being. Core relational themes are perceived through bodily responses, regardless of whether one conceives of such concepts at the time or not.

In arguing that we (emotively) perceive core relational themes, Prinz clarifies that he does not agree with Lazarus that core relational themes “capture inner judgments,” nor that these themes represent evaluations along unified “appraisal dimensions” (65). (Thus, Prinz would disagree with Robinson’s theory, in that she adopts a model of primary and secondary appraisal dimensions as well.) While it may be that some emotion types necessarily correlate with specific cognitive judgments, Prinz argues that this relationship is a causal one rather than a constitutive one (76). In framing emotions as direct perceptions of bodily changes, and as but indirect emotive perceptions of core relational themes, Prinz does not find emotions to be judgments of the body (as do Solomon and Greenspan for example). Prinz claims that perceived bodily changes are sufficient for “embodied appraisals,” in that “the discrete motions of our body convey how we are faring in the world” (78). While we can describe the core relational theme of an emotion in propositional terms, this does not imply (in Prinz’s view) that the appraisal is constituted (even in part) by cognitive acts.

In Prinz’s view, the perception of bodily changes in response to the perception of emotionally competent stimuli just is the sole evaluation. In contrast to Robinson, Prinz argues that the perceived bodily changes in response to an emotionally competent stimulus (such as a snake or bad news) represent core relational themes, and no preceding affective appraisal nor cognitive monitoring after the bodily change need accomplish this evaluation. The perception of bodily changes that represent a core relational theme (and thus, in this way, is also perceived) just is the emotion.
5.4.1.1 Perceptions of Core Relational Themes (Emotions as Representations)

Drawing upon Fred Dretske’s theory of representation\textsuperscript{55}, Prinz argues that the perceived bodily changes (elicited in an emotion) come to “represent” these core relational themes (“Emotion, Psychosemantics” 78-79; Gut Reactions 52-55, 66). To explain the ability of bodily changes to represent nonbodily content, Prinz employs Dretske’s theory of mental representation. Though Prinz notes that theories of intentionality were developed to “explain how concepts refer,” Prinz believes that Dretske’s criteria can “apply to mental states quite broadly” (“Emotion, Psychosemantics” 78). Other mental states, and not just concepts, can refer in like fashion. According to Dretske’s theory of representation, the reference (or “representation”) of any mental state (such as perception) only requires that such states are reliably caused by events that serve a real function. Mental states will represent their reliable causes whenever the causes were set up (whether through evolution or learning) to carry specific information. Prinz argues that, “Mental states refer to those reliable causes that they have the function of detecting. More succinctly, a mental state refers to what it is set up to be set off by” (“Emotion, Psychosemantics” 78). Since the perception of bodily changes has been set up to be set off by core relational themes (themes that carry information about one’s wellbeing), Prinz concludes that such bodily changes represent these core relational themes.

In explaining the relationship between a bodily state and its representation, Prinz discusses a “smoke detector” (“Which Emotions?” 82-83). A smoke detector is set up to function in a way that transmits information about the presence of smoke caused by fire; its alarm, then, can be said to represent smoke caused by fire. What constitutes the smoke detector, though, is its mechanisms – distinct from what it represents. Similarly, Prinz argues, an emotion consists of (perceived) bodily changes even though these changes were set up (evolutionarily) to carry information about our relation to the environment with respect to our well-being. Of course, evolution has also provided us with the ability to learn. Thus, new core relational themes emerge. Prinz argues that it would be, after all, implausible that the perception of bodily changes was set up (through evolution) to represent themselves. He maintains that “Evolution chooses things that confer a survival advantage. If evolution furnishes us with emotions in order to detect bodily changes, then detecting bodily changes must confer a survival advantage. This is a strange hy-

thesis. It is not clear why it is advantageous to know when my blood vessels are constricting” (59). Even though emotions are the perception of bodily change, then, emotions are not about bodily changes. They are about their objects of concern. They are beneficial in that they promote focusing on these objects of concern, not focusing on one’s bodily state. Prinz suggests that the (perception) of such bodily change functions to represent information about our organism–environment relation. He proposes that the perceived bodily states represent information which, as psychology suggests (such as in the work of Lazarus), can be used to identify emotion types according to these core relational themes. Since core relational themes are put in place (within our psychological makeup) to carry specific information about how we fare in the world, these perceptions of bodily changes are then taken to represent these advantageous and specific appraisals. Thus, Prinz rejects the idea that emotions should be construed as dimensional appraisals (or judgments); he claims that “changes in the body are the nominal contents of emotion, and core relational themes are their real contents” (225). Emotions are not judgments, but rather perceived bodily changes that represent core relational themes. Nonetheless, cognitivists are right to point out the intentionality of emotions.

In order to account for the intentionality (or “aboutness”) of emotion, Prinz draws upon Anthony Kenny’s56 distinction between “formal” and “particular” objects (62). Prinz proposes that emotions are intentional (that is, they are “about” something) in two senses: they are about their objects and they are about their core relational themes. Prinz likens the core relational theme to Kenny’s formal objects, and he likens the particular objects to those specific stimuli that elicit an emotional response (62). The particular object of one’s joy on a given occasion, for example, might be a raise in one’s pay at work; in this sense, one’s joy on this occasion is about the raise. The emotion “joy,” though, is always about a more universal theme – conceivably a “fortuitous event.” Likewise, while a smoke detector is intended to be set off by specific instances of smoke caused by fire, the mechanism was set up to carry information about fire – not just (perhaps) today’s fire. The alarm on a smoke detector represents fire (not this fire), even though the representation must be only indirectly caused through the perception of individual cases. Likewise, Prinz maintains that (in an emotional experience) bodily states are taken to represent their core relational themes. It is the (indirectly) perceived core relational themes that reliably cause such states.

We should recall that, for Prinz, an emotional experience does not require an understanding of emotion concepts. If a young child experiences joy, this does not imply an understanding of the concept of fortuitousness; it reveals only a capacity to perceive bodily changes which function to (somewhat metaphorically) carry this information about particular objects. Handing a cookie to a toddler may trigger the perception of the core relational theme for joy (via the perceived bodily change) because the cookie is associated (by one who feels joyful about it) with other fortuitous events. While joy on this occasion is about both the cookie and a fortuitous event in general, the reliable cause of joy is the emotive perception of its theme (being fortuitousness). On Prinz’s view, it would be this indirect perception of fortuitousness that reliably causes the joy. The intentional objects are, then, the particular object and the formal object of the emotion, but the reliable cause of the emotion is said to be the formal object (the core relational theme), and this is what joy then represents. We can certainly imagine being joyful about other particular objects, in that a child’s joy could be about things other than cookies of course, though joy would always reside in the perception of fortuitousness. To experience the emotion does not imply that the child is aware of the emotion concept. While emotions are about both particular and formal objects, emotions are taken to represent their formal objects (their core relational themes) because the perception of core relational themes are their reliable causes, and the perceived bodily changes are set up to benefit us by carrying such information.

Prinz proposes that emotion types, then, can be explained (or identified) according to their core relational themes. That is, emotion types can be defined/categorized according to their (indirect) reliable causes. Thus, we can identify emotion types according to what they represent – their core relational themes.

5.4.2 Prinz’s Account of Emotion Typology

On Prinz’s view, all emotions are perceptions of bodily changes which represent their core relational themes. While each emotion type will obviously have diverse particular objects that elicit it and while different emotion types will sometimes share the same bodily state, Prinz argues that emotion types can be accounted for in terms of their distinguishing core relational themes.

The diversity found between particular objects of emotion and the diversity found between culturally specific emotion types (thus diverse core relational themes) are accounted for in terms of what Prinz
calls (using “Dretskean terminology”)

calibration files” and “recalibration” (99). Prinz explains that “Calibration files are data structures in long-term memory. Every calibration file contains a set of representations that can each causally trigger the same (or similar) patterned bodily response. The perceptions of the bodily responses caused by representations in a calibration file are emotions” (100). Representations of the particular, and certainly diverse, objects which can trigger a specific core relational theme are taken to comprise calibration files distinct to each emotion type. The representations that comprise a calibration file are coined “elicitation files” (“Which Emotions?” 86). A calibration file that represents “loss,” for example, will contain many elicitation files within it (“Which Emotions?” 85); Prinz speculates that, while sadness may have its origins in a response to “separation distress,” sadness can in time (through enculturation) come to respond to other elicitation files (“Which Emotions” 85). The elicitation files which comprise a calibration file for sadness may include such diverse objects as the death of a loved one, job loss, break-ups, etc. It is the (indirect) perception of calibration files (which is triggered by a motley of elicitation files) that causes the perceived bodily changes which constitute the emotion.

Prinz also maintains that the same (perceived) bodily state can be recalibrated to represent new core relational themes (that is, new calibration files) through cultural learning. Thus, he explains that jealousy may be the recalibration of the bodily state that comprises anger (representing the core relational theme, perhaps, of “offence”) to a new theme of “infidelity” (99-100). He proposes that whenever there is a theme that (when perceived) reliably causes a bodily state, a calibration file exists. Through cultural learning, new themes are added (recalibrating body states to represent new information), and thus new emotion types become possible. Emotion types can be explicated, then, in terms of their core relational themes, whether these themes are provided for by nature itself (as is the case in basic emotions) or whether the themes are learned and thus represented via recalibrated body states (as is the case in non-basic emotion types).

Critical to Prinz’s perceptual theory of emotion is his claim that emotions are elicited by both percepts and judgments (Gut Reactions 66-67). Percepts are taken to be objects that are immediately perceived independent of any judgment, such as the perception of a “sudden sound” or “gun pointing” (77). Judgments are taken to be cognitive acts, such as the belief that “He’s a fascist!” or “That’s poison!” (77). Prinz maintains that all emotion types consist of the same structure. They are patterns of

---

(perceived) basic bodily change (or blends of such changes) in response to the (indirect) perception of a core relational theme. Here, Prinz explains the common structure of all emotions:

Each is structurally analogous. Each is simply a perception of a patterned bodily change. Even emotions that we acquire by blending have this simple structure. They are simply perceptions of blended bodily patterns. Some emotions are attained by adding conceptually sophisticated judgments to our elicitation files, but this does not alter their structure. Elicitation files are content-determining causes of our emotions, not constituent parts. (“Which Emotions?” 85-86)

Prinz argues that while we may be able to identify themes represented by basic emotions (in that we can describe the information that basic emotions carry), “basic emotions” are best understood as basic bodily changes (provided by nature) rather than as representations of distinct emotion concepts (“Which Emotions?” 84-85). These basic bodily patterns (and even blends of them) come to respond to percepts and judgments, and these states come to represent diverse core relational themes acquired through enculturation. Like Robinson, Prinz admits that some emotion types may be “cognitive elaborations of basic emotions” (93). That is, some emotion types may be particularly “caused by disembodied judgments” (such as emotions that are caused by propositional beliefs) (76). Nonetheless, he claims that these cognitive elaborations are not part of the emotion but rather a way of identifying the emotion type in such cases. If anger and guilt turned out to require moral judgments or the use of moral concepts as their cause, then (as discussed) this would certainly pose circularity problems for the norm-expressivist. Fortunately, for the norm-expressivist, Prinz’s theory does not imply that anger and guilt require such causes. In fact, he argues to the contrary.

5.4.2.1 Prinz’s Account of Moral Emotions
Since emotion types are taken to be types of perceived bodily states that represent core relational themes, we should consider what bodily state constitutes the moral emotions and what core relational themes would identify these emotion types.

Prinz argues that, in spite of the critics who maintain that it is known that guilt is “not associated with bodily perturbations,” there is evidence to the contrary. He claims that “existing functional imaging studies of [guilt] tell against the constructionist conjecture” (“Which Emotions?” 80). That is, functional image studies of the brain show that
guilt is not but a social construct, for it also carries its own distinct neurological patterns. Prinz notes that Shin et al.58 “found activation in the anterior cingulate cortex and the insula when subjects recalled episodes of guilt” (“Which Emotions?” 80). Referencing Damasio’s work, Prinz concludes that “These structures … are known to play a role in perception and regulation of the body” (“Which Emotions?” 80). We can also recall, along these lines, other studies on the neurological patterns being discovered in guilt as discussed in Section 3.4.1 of Chapter Three. Still, while it may turn out that the moral emotions have their own distinct bodily perturbations after all, this is not necessary in order for his theory to account for these emotion types. It is ultimately the core relational theme (represented by the bodily change in response to a cluster of stimuli) that distinguishes an emotion type (128).

As discussed, Prinz maintains that recalibrated emotions are perceptions of basic bodily states (or blends of these basic bodily states) in response to objects that comprise a new calibration file (thus constituting a new core relational theme). While (on Prinz’s account) new core relation themes are learned through enculturation, we can recall that the experience of emotions (on his view) does not imply that one comprehend the emotion concepts. In keeping with this, Prinz proposes that guilt, in its more primitive form anyhow, may be the recalibration of the bodily state associated with “sadness” to a new core relational theme specific to guilt (128-129). He argues that this can occur even in one who lacks such an emotion concept. Prinz notes evidence that “By the time they are two years old, children show signs of guilt and shame when they do something wrong.” At this time, children lack any understanding of these emotion concepts (The Emotional Construction 36, 117). Thus, Prinz notes that children may experience moral emotions before they are able to make moral judgments in the norm-expressivist’s sense, and thus “A person can have moral attitudes without grasping concepts that refer to moral emotions. The crucial thing is not emotion concepts, but the emotions themselves” (The Emotional Construction 117). A young child may show signs of guilt prior to even understanding (or being able to identify) her own emotional experience. Thus, Prinz argues that we can experience moral emotions without grasping these moral emotion concepts. The body can represent core relational themes even when we have yet to understand these emotion concepts.

Regardless of whether guilt entails a unique bodily state or whether it is the learned use of another bodily state in response to learned

behaviors, Prinz argues that guilt is “conceptually inexpensive” (128). Even if guilt is a learned emotion, Prinz maintains that guilt may still “have analogues in dogs and nonhuman primates” (128). While he admits that some emotion types may require cognitive elaborations as their cause, he does not suspect that this need be the case with guilt. We can recall that Prinz argues that there is a dedicated and non-cognitive perceptual system which allows us to track core relational themes (upon the recognition of the objects that elicit these bodily responses) and this perceptual system need not employ any top-down reasoning (64-66, 224-225). He suggests that if other animals can track personal transgressions (as do dogs), without making moral judgments, then humans would also have the ability to track such primitive themes without entertaining these concepts. Prinz provides the following argument for his position that emotions track core relational themes independent of conceptualization:

Emotions get their meaning by reliably tracking core relational themes. Guilt may get its meaning by reliably tracking personal transgressions. This does not require a self-concept or a transgression concept. A creature can track its own transgressions in other ways …. When Fido pees on the carpet in the presence of his human caregiver, he enters a negative emotional state. This could be regarded as a primitive form of guilt because it is reliably caused by personal transgression. Fido does not enter this state by thinking “Gee, I violated a moral norm.” He perceives the event, which happens to be a norm violation, and he responds by feeling bad. (128)

Thus, Prinz argues that dogs and people would seemingly be able to learn the types of circumstances which beget a loss of affection by those who matter, without entertaining normative concepts. While Prinz here discusses only primitive guilt (for there may be other more conceptually expensive forms of guilt, ones that are only experienced by those who *can* entertain moral concepts), this is a form of guilt nonetheless. Prinz’s point in discussing this primitive guilt is to propose the origin of guilt. All feelings of guilt seem to have their roots in this primitive experience.

Prinz proposes a theory which may explain the existence of such primitive guilt. He proposes that the recalibration of sadness into guilt is quite plausible, since sadness responds to loss, and guilt would seem to be a response to particular objects (such as transgressions, or harming others) which beget the loss of others’ affection (129). Thus, such specific behaviors come to be associated with one another and, thus, guilt (as a distinct emotion type) comes to be a response to losses that are especially
interpersonal and resulting from one’s own deeds. We come to associate certain deeds with a damaged interpersonal relationship and thus feel sadness in response. (Since canines can also, perhaps, anticipate the loss of affection upon doing certain acts and experience the feeling of sadness, they may then also experience such primitive guilt.) Because this feeling of sadness occurs in response to learned behaviors that tend to meet a damaged interpersonal relationship, the emotion “guilt” exists insofar as it has its own core relational theme. That is, the body comes to respond to the recognition of circumstances that just do tend to meet such damaged interpersonal (or Baumeister may say “communal”) relationships. Insofar as the body, through its response to these elicitation files, then carries information about such “transgressions” (as Prinz calls it), a new core relational theme is born even if the experiencer has yet to grasp the emotion concept of guilt, self, or transgression. Prinz considers the following explanation for the development of such primitive guilt:

When a young child does something wrong, parents and caregivers become upset. They may respond in various ways: getting angry, drawing the child’s attention to the harm she has caused, or withdrawing love and affection. Children are gregarious and want their parents’ affection. Being scolded or cast off naturally cause sadness, because they lead to loss of that affection. Being oriented toward someone who has been harmed also causes sadness through emotional contagion. Thus each form of ordinary parental response leads to sadness. This forges a link between sadness and transgression in memory, and guilt is born. Subtract sadness from the child’s emotional repertoire, and guilt goes too.

(129)

Guilt, then, can be accounted for even if it has no unique, or innate, bodily response. It may be that guilt is the perception of its own distinct bodily perturbation in response to the indirect (and emotive) perception of a core relational theme, or it may be the recalibration of sadness (or possibly some other innate bodily state or a blend of them) to this new core relational theme.

To establish that persons can experience guilt without grasping or entertaining moral judgments or moral concepts is one thing, but we should also consider how we are to conceive of the emotion concept. As Sinnott-Armstrong pointed out (Chapter One and Chapter Four59), we can only recognize and make moral judgments on the norm-expressivist’s

59 See Section 4.3.1.1.1.
view if we grasp these moral emotion concepts. We are to do so, according to Prinz’s theory, by identifying the distinguishing core relational theme. It is the core relational theme that defines the particular emotion types.

Prinz proposes that the core relational theme, for guilt, may be described by us as something like “personal transgression” (128). He proposes that “Guilt emerges when sadness gets recalibrated to personal transgressions” (“Which Emotions?” 84). One might ask whether transgressions (or even personal transgressions) is specific enough to be the core relational theme for guilt. Prinz cites Baumeister et al. (who, as discussed in Chapter Four60 proposed that guilt is about one’s actions that harm communal relationships), noting that “Baumeister et al. (1994) have shown that guilt is most frequently associated with actions that threaten individuals to whom one has an attached relationship” (The Emotional Construction 76). Prinz notes that “bona fide instances of guilt” have an “interpersonal character,” and thus perhaps it would be better to describe the core relational theme as “personal transgression of interpersonal relationship” (The Emotional Construction 117). Even so, if we are to identify the core relational theme for primitive guilt, anyhow, it is odd to attribute “transgressions” to the theme, since this might be taken to imply a moral concept to the theme, and Prinz has argued that primitive guilt can be experienced even by canines and young children (hardly moral agents). While it may be possible to conceive of transgressions in a non-moral terms (perhaps as behavior that is against the typical expectations of others in a relationship), I think it best to avoid that complication. What Prinz seems to be getting at is that even primitive guilt can function to represent the types of behaviors that just tend to be interpersonal transgressions. In this case, though, it may be better to say that the core relational theme for primitive guilt is “behaviors by oneself that tend to meet a damaged communal relationship.” (After all, while it may seem odd to say that dogs experience personal relationships, I have no problem conceiving of the canine as communing with others. If we are to characterize primitive guilt, then we ought avoid limiting the core relational theme to persons only.) Given Prinz’s theory and his arguments, it seems quite right to say that primitive guilt can be conceived of as the perception of a negative bodily state (one likened to sadness) that represents “behaviors by oneself that tend to meet a damaged communal relationship.” To identify guilt in such a way does not appear to be either a modification to his theory, nor even an embellishment.

60 See Chapter Four, Section 4.3.1.1.2.
In contrast to guilt (which is taken to recalibrate the bodily state for the sadness), Prinz infers that shame borrows its (perceived) bodily state from the basic emotion embarrassment ([The Emotional Construction](p. 76)). Based on a study that Prinz conducted, he infers that shame has a distinct core relational theme in that it tends to respond to different sorts of objects than does guilt. Prinz asked the subjects of his study to report how they would feel in response to having committed different sorts of “moral crimes” (76). The questions were intended by Prinz to reflect “crimes against nature” (such as “Suppose, in a moment of weakness, you allow a person who is really old to kiss you romantically”), “crimes against persons” (such as “Suppose you take something from someone and never return it”), and “crimes against community” (such as “Suppose you unthinkingly start to eat your dinner at a family gathering before everyone has sat down at the table”) ([The Emotional Construction](p. 77)). Prinz reports that, “For the crimes against nature subjects overwhelmingly chose shame, and for the crimes against persons, they overwhelmingly chose guilt. There was no statistically significant difference, however, between guilt and shame for the crimes against the community” (77). As a species of embarrassment, Prinz claims that shame arises “when we receive unwanted attention from others,” and yet (given the results of his study) he also explains that shame responds to acts by oneself that are associated with social impurities. He claims that shame “could arise only after our species began metaphorically to extend the class of physical impurities into the social and behavioral domain. Shame arises inevitably once a culture starts to label certain acts as unnatural or deviant” (78). Prinz does not discuss disdain, though perhaps we can see disdain as the response that others feel toward one who behaves in shameful ways.

If this is true, though, then it would seem that Prinz is characterizing shame and disdain in ways Gibbard did not. These emotions, being responses to circumstances that constitute “impurities” in the social and behavioral domains, would seem to be defined in terms of moral wrongdoing. To conceive of one’s actions as being “socially or behaviorally impure” is certainly to conceive of one’s acting wrongly. ([The Emotional Construction of Morals](p. 78), Prinz argues that shame is among the moral emotions.) Should shame also be extended to states (such as appearances), then the core relational theme would not seem to imply any concept of wrongdoing. In the way Prinz conceives of it, though, shame is about impure acts, and this would then seem to imply moral concepts.

---

If shame and disdain are conceived of in terms of wrongdoing (wrongdoing of a specific sort, with respect to purity that is, but wrongdoing nonetheless), then the norm-expressivist would find these emotion concepts to be conceived of in terms of moral concepts. Still, even if Prinz has it right in describing shame and disdain in this way, this does not undermine the norm-expressivist’s analysis of the moral judgment. The moral judgment, on the norm-expressivist’s view, is analyzable in terms with reference to the rationality of guilt and anger. This need not preclude shame and disdain concepts from being defined in terms of moral concepts. While shame and disdain are (on Prinz’s account) defined in terms of moral concepts, this need not alter the norm-expressivist’s analysis of the moral judgment. The norm-expressivist could still analyze the meaning of a moral judgment in terms of (primitive) guilt and anger, because only those emotion types are broad enough responses to wrongdoing to account for the moral judgment. Shame, on Prinz’s account, responds only to acts of impurity, and surely moral judgments are about more than just that. Guilt and anger would seem to better serve as the references of moral judgments.

Given Prinz’s theory, anger would also be the perception of a bodily state in response to an emotively perceived core relational theme which is affected through enculturation. The bodily state could be a unique state unto anger alone, it could be a state that is shared with other emotion types, or it could even be a blend of basic bodily states. Whatever this bodily state, it would represent the core relational theme of anger. Prinz proposes that the core relational theme of anger might be something like “goal frustration” or “aggressiveness” (16). Again, as with guilt, Prinz’s theory would maintain that one can track cues that comprise this calibration file without thinking (or judging) that “An offense (or aggression or goal frustration) has occurred.” After all, other animals get angry and, in Prinz’s view anyhow, they do not engage in (top-down) cognitive acts that judge the objects of their anger. Anger can be elicited by both percepts and judgments. As Prinz says, “Anger can be stirred-up by a glare or by a chain of high-level moral reasoning” (“Which Emotions?” 77). If anger can be elicited by percepts as well as judgments, then it is not an emotion that requires a moral judgment as its cause. Likewise, we can conceive of this basic/primitive anger concept in non-moral terms. Prinz proposes what we previously considered (on behalf of the norm-expressivist) as a compatible embellishment to cognitivist theories. He proposes that anger can be defined (or conceived of) as the emotive perception of “goal frustration” or even “aggressiveness”; we might even propose, in no way that even modifies Prinz’s claims, that anger is the “tendency to lash out in response to goal frustration.” On
Prinz’s theory, anger can be experienced and defined without any use of moral concepts.

Like Solomon, Prinz does propose a characterization of *moral anger* (*The Emotional Construction* 80-71). *Moral anger, as a distinct emotion type, would be identified according to a specific type of judgment (a moral judgment) as its cause. Moral anger, though, is a species of the more primitive anger. The norm-expressivist is not analyzing the moral judgment in terms of moral anger, but rather a more general sense of anger. Thus, we need not fear that Prinz’s characterization of moral anger would impact the norm-expressivist’s analysis of the moral judgment. The norm-expressivist need only find a compatible way to conceive of primitive guilt and primitive anger.

5.4.2.2 Perceptualism & Cognitivism: Shared Compatabilities for Norm-Expressivism

Notwithstanding the unique and holistic appeal of Prinz’s theory, it is interesting to consider whether his perceptual theory may have more in common with cognitivist theories of emotion than he would admit. Salmela (in his article “Can Emotion be Modelled on Perception?”) critiques perceptual models of emotion, pointing out many of the similarities between Prinz’s theory and cognitivist theories of emotion. Salmela’s analysis of Prinz’s theory would reveal that (if he is right) Prinz’s theory implies that emotion includes cognition after all. Salmela argues that emotions are not analogous to perception, for any system that can pick out core relational themes would require the use of cognition in order to truly pick out stimuli that elicit such representations. Salmela notes that emotions only *function* to represent core relational themes *if* their elicitation files are indeed fitting to the cause of concern (6-7). (He notes, for example, that according to the DSM-IV manual of psychopathology, phobias and panic disorders are disfunctions due to one having irrational fears in response to stimuli) Salmela argues that cognition must then be included in emotion after all, via the “initiation pathways” which function to house correct calibration files (6). Whereas Prinz would likely respond that such initiation pathways are not venues for cognitive activity (since they can operate in perhaps even sub-cortical and automatic ways), Salmela responds that cognitivists would include such transfers of data as cognition (10). Prinz and Salmela would perhaps agree on the anatomy of how emotions can proceed, though Prinz’s narrow depiction of “cognition” implies (in his view) that such processes (from the initiation pathway to the emotion itself) are all explainable within a perceptual model of emotion.
Since my aim is not to evaluate the theories of emotion so much as the compatibility of their frameworks with norm-expressivism, I will just note that such criticisms might imply that (should Prinz’s theory work with norm-expressivism), then it may again support the possibility that broadly construed hybrid cognitivist theories are equally compatible with norm-expressivism.

5.4.3 Would Prinz’s Perceptual Theory Suffice for Norm-Expressivism?

I find that Prinz’s perceptual theory of emotion meets all of the norm-expressivist’s conditions for a compatible theory of emotion. Our understanding of the way in which his theory meets C1 (distinguishing moral emotions from one another) reveals the way in which it meets C3 (the circularity challenge). His theory also accounts for the relationship between emotions and the behaviors that emotions enable, and in doing so it meets C4. Bearing in mind how the theory meets C4, and how it explains emotion typology in general, the theory also meets C2 (explaining how some emotions are stronger motivators of action than are others). Because moral emotions are subject to normative governance on Prinz’s view, through the role of enculturation in emotion typology, Prinz’s theory will also meet C5 (allowing for the impact of norms that we are more or less aware of), C6 (allowing for broadly moral emotions to shape anger and guilt), and C7 (allowing for judgments of rationality to causally impact the moral emotions).

C1 requires that “The theory of emotion must provide a framework capable of distinguishing between guilt and anger, as well as shame and disdain (and other broadly moral emotions).” As discussed above, Prinz’s perceptual process theory does provide a framework to distinguish between the moral emotions, in that we can distinguish between the emotions by either their distinct and unique bodily states (if those exist) and/or their core relational themes. C3 requires that “The theory of emotion cannot imply that anger and guilt are, nor that they are necessarily responses to, moral judgments, nor can moral emotions be defined in terms of moral concepts.” Since moral judgments would be expressions of higher level thought processes attained through normative discourse and considerations of their rationality, the moral judgment is not nor would it necessarily be caused by such moral judgments on Prinz’s view. Given Prinz’s perceptual theory, he maintains that the core relational themes for both guilt and anger may be perceived without any top-down cognitive activity. He maintains that guilt and anger are both conceptually inexpensive, being experienced even before one grasps the
concepts that they function to represent. While this condition also requires that we must be able to identify guilt and anger concepts, Prinz provides arguments to support non-moral descriptions of such emotion concepts. Unlike the other theories of emotion surveyed, Prinz argues that we ought to conceive of guilt (or primitive guilt anyhow, which is all the norm-expressivist need reference in the moral judgment) as the representation that one’s own behavior is of the sort to meet a damaged communal relationship, and anger (or general, primitive anger anyhow, which is again all the norm-expressivist need reference) represents one’s own goal frustration or aggressiveness. Therefore, Prinz’s theory can distinguish between anger and guilt according to their core relational themes without implying that these emotions require moral judgments or moral concepts as their cause, or that they need be defined in terms of moral concepts.

C4 would require that “The theory of emotion must explain how specific types of emotions (e.g. guilt and anger) motivate actions.” While emotions are neither necessary nor sufficient for actions on Prinz’s account, they do enable actions, and this would explain our tendency to find certain types of actions to manifest in response to certain emotions. Even if emotions are not such behaviors and even if they cannot by themselves tend one toward such behaviors (short of the choice to engage in the behavior), emotions nonetheless support behaviors on Prinz’s view in that they enable them. Emotions could seemingly enable actions without being necessary to the actions, for emotion could be one among other disjuncts which together form the necessary disjunctive conditions for the behavior. If, for example, I either believe that I should do X or I enjoy doing X, then I can choose to do X. The emotion may, here, enable my doing X even though it is neither necessary (since a belief might have also allowed me to choose the action) nor sufficient (since I would still need to choose to do X in order to do X) for doing X. Emotions, then, are neither sufficient nor necessary for action, even if they enable actions. In these ways, Prinz’s theory would seem to account for the behavioral specificity we tend to see in emotion types without implying that such macro-behaviors are sufficient or necessary to emotion.

C2 requires that “The theory of emotion must provide a framework capable of explaining how some emotions (such as anger and guilt in norm-expressivism) tend to be stronger motivators of action than are other moral emotions.” Because emotion types carry their own affective responses, certain emotions will be more perturbing than others. If an emotion type manifests in bodily responses that are more pronounced (or uncomfortable, such as the negative emotions), it could turn out (as Gibbard thinks) that these emotions motivate people more so than other
emotions to do something to change the situation. (Of course, it could also turn out that positive emotions are in fact more motivating toward behavior, or equally motivating anyhow, but in that case Gibbard’s theory has a different sort of problem.) It would also be consistent with Prinz’s theory that people have behavioral norms that dictate choosing behavior to address feelings of guilt and anger more so than they dictate choosing behavior to address other feelings. Prinz, himself, notes that emotions may enable, but not dictate, our chosen actions. We may choose to act upon some emotions more so than others, and norms may certainly impact our choices. Therefore, Prinz’s theory offers a framework to address C2 in that his theory allows for the perturbation of some emotion types to potentially motivate us more than others and in that his theory would allow for us to just choose to act upon some emotion types more than others.

C5 requires that “The theory of emotion must explain how emotions (e.g. anger and guilt) can be caused by the acceptance of norms (and the fact that one has accepted norms) that we are more or less aware of at the time.” Because elicitation files are taken often to be products of enculturation, and because the calibration files which comprise the core relational themes are also products of enculturation, emotions are certainly (on Prinz’s theory) subject to normative influence. Additionally, since he maintains that emotions are responses to non-cognitive perceptions, it would seem that we may be more or less aware of these norms that influence the calibration files and even the elicitation files themselves.

C7 requires that “The theory of emotion must explain how guilt and anger respond to considerations regarding their consistency with other norms, their grounding reasons, and their plausibility.” Because Prinz argues that emotions can respond to percepts and judgments, guilt and anger could then respond to judgments about their consistency with other norms, their grounding reasons, and their plausibility; a judgment about any of these things can certainly cause or mitigate the moral emotions on his view. If one is angry and then comes to believe that the anger is based on inconsistent norms or that there are no grounding reasons for the anger, then one’s anger may diminish (which is not to say that it has to diminish, or that the bodily response won’t hold for a bit longer of course). While Prinz’s theory does not maintain that emotions are the product of such judgments about their rationality, it is completely consistent with such judgments influencing the emotion in that judgments can be included in the elicitation files for emotion types.

C6 requires that “The theory must provide a framework that explains how broadly moral sentiments shape guilt and anger.” To this, Prinz could explain (within the framework of his overall theory) that the
learned societal norms for one emotion can clearly be impacted by norms for other emotions. Empathy, for example, may promote norms for guilt and anger. Perhaps in part because empathy promotes honesty toward loved ones (in that we can imagine how bad it feels to be deceived by a loved one), we then add “one’s own dishonesty toward loved ones” to the calibration file for guilt. It is empathy, in this example, that helped to then shape this norm for guilt. We might similarly suppose that the empathy for others’ feelings upon such deception also shapes our norms for anger at such things – by again adding “others’ deception toward their loved ones” to the calibration file for anger. Equally, of course, feelings of empathy could promote norms for guilt and anger that promote forgiveness. Whether one's experience of other broadly moral emotions (like empathy or sympathy) is the product of enculturation or one’s biological make-up, the broadly moral emotions could (on Prinz’s view) certainly shape the norms for guilt and anger.

In the end, then, I find that Prinz’s perceptual theory of emotion would be consistent with the norm-expressivist’s seven conditions. Notwithstanding that there are challenges for this ingenious theory in its own right, the theory does on its face appear to be consistent with norm-expressivism. In fact, Prinz’s theory offers a unique advantage to the norm-expressivist in that Prinz, alone (of those theories surveyed anyway), provides an argument that anger and guilt may be both experienced and identified without the concept of blameworthiness or wrongdoing. In this, Prinz’s theory of emotion is a most ideal fit for norm-expressivism.

5.5 Conclusion

Both somatic theories of emotion are consistent with norm-expressivism. Like paradigmatic behaviorist theories and cognitivist theories, these theories meet all seven conditions. Neither of these theories requires modification in order to depict the moral emotions in a way that avoids the circularity challenge. Prinz’s perceptualism, though, is the only theory that need not be even embellished in order to accomplish this. In fact, Prinz seems to offer the arguments for the nature of emotion and the nature of both (primitive) guilt and anger that was so needed by even cognitivist theories of emotion if they are to show compatible with norm-expressivism.

Somaticists emphasize the ability of emotions to result from non-judgments, and thus moral emotions need not be caused by moral judgments. According to the somaticists’ theories (and like the
attributional theories of emotion maintain as well), moral emotions could well respond to cues that just happen to be of moral concern without implying that a moral concept was employed. For Prinz, emotions can result from a mental mechanism’s ability to track core relational themes associated with various particular objects of emotion, even when the person has no comprehension of the emotion concepts. For Robinson, emotions begin with an immediate non-cognitive appraisal of either percepts or judgments as well; the emotion itself is not a judgment, nor is it required to be a response to a judgment. Both theories explain how moral emotions can be learned responses without implying that the moral emotions are (or need be caused by) moral judgments. While cognitivist theories may be able to address this point, also, by characterizing the distinguishing judgment (or propositional content) of an emotion type in non-moral terms, somaticists need not be concerned with this in order to deny that moral emotions need be caused by moral judgments. Somaticists make the argument that emotion types can function in terms of certain appraisals (representing them, as Prinz argues) without “describing them” (65).

While somatic theories would (if otherwise cogent) avoid the need for either moral judgments or the use of moral concepts causing anger and guilt, Sinnott-Armstrong’s criticism would no doubt be made with respect to our ability to conceive of or recognize these moral emotions. If we cannot comprehend guilt and anger concepts, then we cannot make a moral judgment or recognize moral judgments made by others. Here, Prinz’s theory takes the lead over Robinson’s theory for the norm-expressivist. Prinz’s perceptual theory is particularly well suited to norm-expressivism, for Prinz offers accounts of moral emotion concepts that are what we have been seeking all along. He argues that not only can the experience of primitive moral emotions occur without one’s use of moral concepts of moral judgments, but also that we can identify guilt and anger concepts in non-moral terms. He characterizes the core relational themes of primitive guilt and anger in non-moral terms, for he maintains that primitive moral emotions are experienced prior to one’s grasping moral concepts. The way in which Prinz characterizes the core relational theme of primitive guilt and anger, and even the arguments that he makes for describing these features as such, are precisely the way in which cognitivist theories of emotion may meet all of the norm-expressivist’s conditions. We just need not embellish or modify Prinz’s theory in demonstrating the compatibility of his theory with norm-expressivism.

Some theories, then, are show to meet the norm-expressivist’s conditions for a compatible theory of emotion. Still, they do this only if a pretty bold claim proves to be true – namely, that moral emotions can in
fact be elicited through mental mechanisms which learn to track such circumstances without employing any moral concepts in the process. If Prinz is right, in maintaining that this is possible, then norm-expressivism would pair up well his perceptualism.
6. Conclusion

As it turns out, several theories of emotion are consistent with norm-expressivism. These theories meet all seven conditions. In different ways, these consistent theories provide frameworks that would allow for moral emotions to be responses to clusters of objects in some way that carries information without ever employing a moral judgment or an understanding of moral concepts. If paradigmatic behaviorism, cognitivist theories, and somatic theories of emotion are cogent, then they will complement norm-expressivism so long as they allow for the possibility that people track and respond to morally relevant objects without ever employing moral judgments. The paradigmatic behaviorist theory, cognitivist theories and Robinson’s process theory would allow for this possibility – being quite silent on the issue. Indeed, an attributional theory and Robinson’s theory of emotion may seem to offer the norm-expressivist some advantage, in providing arguments to explicitly prove the capacity of emotions to arise in response to perceived objects alone – independent of any judgment. Prinz’s perceptualism, though, maintains that this is not only a possibility for some emotion types, but it is the case with primitive guilt and anger in particular. Thus, while several theories are consistent with the norm-expressivist’s conditions, we find that some theories offer arguments that more explicitly support a characterization of anger and guilt in ways conceived by the norm-expressivist. Prinz’s theory, in depicting primitive guilt and anger in explicitly non-moral terms, is especially conducive to addressing the circularity challenge faced by norm-expressivism.

Paradigmatic behaviorist theories of emotion maintain that emotion types can be identified (among other things) by their causes, and these distinguishing causes are left open for other fields of science to determine. Should other fields of science, such as psychology, determine the causes of emotion to be a sort of labeling of a body state in direct response to circumstances (as proposed by attributional theories for example) or a sort of recognition of morally relevant circumstances without any moral judgment, then the paradigmatic theory of emotion is consistent with the norm-expressivist’s conditions. There is nothing in paradigmatic behaviorism to resist such possibilities and its heuristic would
welcome such empirical analysis, and thus in this way it meets all of the norm-expressivist’s conditions. Solomon’s denial that emotions have propositional content and his broad treatment of the emotive “judgment” also allow for the possibility that moral emotions do not employ moral judgments. Should the propositional content of the emotion types, as proposed by other cognitivists, be realized in non-moral depictions of guilt and anger, they too may address the norm-expressivist’s circularity concern. The somatic theorists’ emphasis on the modular (and non-cognitive) evaluations of emotive objects would also imply that, in their view, moral emotions may occur without any moral judgment. Indeed, Prinz even argues that this is so and there is at least some empirical evidence to support his claims. So long as the narrow moral emotions (of anger and guilt) do not consist of moral judgments, nor require moral judgments as their cause, the norm-expressivist’s analysis of the moral judgment is not directly circular.

Of course, only a discovery into the facts about the nature of these eliciting objects and the nature of such psychological mechanisms could prove that moral emotions do not, in fact, employ moral judgments. That, though, is an empirical question. Solomon’s and Nussbaums’s judgmentalism, attributional theories of emotion, Greenspan’s perspectivism and somatic theories of emotion all purport that emotions can, in fact, be responses to the perception of objects without employing higher cognition – cognition seemingly required in the norm-expressivist’s moral judgment. Each of these theories would allow for the narrow moral emotions (or anger and guilt) to have the broad capacities required by norm-expressivism. In these ways, these theories address the norm-expressivist’s central circularity concern (C3), and they do so while offering a framework to distinguish moral emotions from one another (C1). With some modification, even dimensional appraisal theories can offer a compatible framework for emotion to the norm-expressivist. Each of these theories addresses the other conditions in their own ways as well. Because paradigmatic behaviorism is consistent with the norm-expressivist’s seven conditions, and because its heuristic allows for its account of emotion typology to be amplified by these other theories of emotion as developed in psychology, it also meets the norm-expressivist’s needs well.

Still, all except for Solomon, Robinson and Prinz meet these conditions by way of our modifying either their framework itself or their way of specifically characterizing guilt and anger. Solomon and Robinson meet these conditions by way of our embellishing their theories with more exact (and non-moral) depictions of guilt and anger. While these embellishments and modifications are plausible, only Prinz offers explic-
it arguments for primitive guilt and anger being responses which can occur without any moral judgment or use of moral concepts.

As noted throughout, even if moral emotions do not themselves imply moral judgments (they are not moral judgments and they need not be caused by moral judgments), the ability to make a moral judgment on the norm-expressivist’s view would still imply the use of moral concepts in order to understand/reference the moral emotions. Some (like Sinnott-Armstrong) have argued that, in order for one to make a moral judgment on the norm-expressivist’s view, we must be able to understand the concept of these moral emotions without begging the question. Sinnott-Armstrong and Hill have argued that one can only conceive of guilt in terms of moral concepts – in terms of wrongdoing. Thus, on this view, something like moral judgments would seem to be implied within our understanding of the moral emotion concept itself, for we would only comprehend these emotion concepts (as points of reference in the moral judgment) through comprehending the moral concepts. If the understanding of these moral emotion concepts (i.e., the understanding of guilt and anger) requires the use of moral concepts themselves, though, then norm-expressivism would still face problems with circularity in its definition of wrongness. It may not be a direct circularity, in that employing a moral concept in one’s understanding of an emotion concept is not the same thing as making a moral judgment in the norm-expressivist’s very distinct sense. Still, some circularity would seem to be present in such an account nonetheless. After all, Gibbard uses his meta-ethical analysis to explain the meaning of moral judgments in terms of their being judgments of wrongdoing.

I have argued that it is quite plausible that this concern, too, can be met. It is plausible that moral emotions respond to the cues which are describable in non-moral terms. If I can define the content (or core relational theme) of guilt in terms of “actions (by oneself) that tend to meet a damaged communal relationship,” and if the content (or core relational theme) of anger is explained in terms of “a threat to one’s goals by another, of which I have the power to change,” then there need be no moral concepts in either the definition of these narrow moral emotions nor in the recognition of these narrow moral emotions. If this is so, then these (compatible) theories need not run the risk of even a circular definition of moral terms.

While I have argued that it is plausible for guilt and anger to be accounted for in terms of their respective cause, constitution, and identity in non-moral terms, we can note that this is not an entirely non-

---

62 I am grateful to Åsa Carlson for prompting further consideration and clarification on this point.
controversial claim. If we characterize anger in terms of goal frustration (and perhaps the view that it is within one’s power to do something about it through lashing out) and if we characterize guilt in terms of the sorts of actions (by oneself) that tend to meet damaged communal relationships, then norm-expressivism can avoid circularity. Certainly, though, not everyone will agree with this characterization of these emotions – particularly in the case of guilt. A Kantian, for example, may point out that feelings of guilt must be capable of responding to the realization that one has failed to act upon one’s duty, and one may suspect that these actions (i.e. these violations of one’s duty) are not so obviously behaviors that tend to meet damaged communal relationships. A utilitarian, as well, may claim that feelings of guilt must be able to respond to the realization that one has failed to maximize the good, and one may suspect that these actions (which fail to maximize, or aim at maximizing, the good) need not be of the sort to meet a damaged communal relationship. I find, though, that the types of behaviors (by oneself) that are wrong according to Kantian or utilitarian standards just are also the sorts of behaviors that tend to meet damaged relationships. The types of universal duties that exist on a Kantian view (such as abstaining from lying, keeping promises, not killing, etc.) speak to the types of behaviors that also tend to meet damaged communal relationships. The types of behaviors that fail to maximize the good on a utilitarian view (such as, perhaps, failing to tell a sorely needed white lie according to an act utilitarian) are also of the sort that tend to meet damaged communal relationships. Thus, while such an argument is by no means conclusive, I find it quite plausible that this characterization of guilt is compatible with other ethical theories (such as, for example, Kant’s deontology and utilitarianism). I find it quite plausible, after all, that guilt and anger can be sufficiently characterized in non-moral terms.

Nothing in the framework of these compatible theories of emotion implies circularity to norm-expressivism. These theories allow that emotions can be responses to objects through mental mechanisms independent of any conceptual evaluations. It is plausible that we could even understand these emotions concepts (of guilt and anger) in non-moral terms. Should this turn out to be the case for the moral emotions (in contrast to arguments put forth by Sinnott-Armstrong and Hill), then norm-expressivism has found friendly theories of emotion indeed. Not only do these theories of emotion offer frameworks to account for moral emotions without implying the use of moral concepts, but they offer an account of emotion and emotion types that addresses all conditions for a theory of emotion well-suited to norm-expressivism. Thus, I find that that the compatibility of norm-expressivism and moral emotions need not
turn on the debate between behaviorist, cognitivist, and somatic theories of emotion – and that should be good news for the norm-expressivism.

That being said, only the facts can ultimately prove whether or not these moral emotions could work in such a way. That is not a problem with these theories of emotion though. Any cogent theory of emotion (consistent or inconsistent with norm-expressivism) must adapt itself to the facts around emotion typology. If the facts turn out to show, in a rather conclusive way, that children make moral judgments prior to understanding guilt and anger, then the credibility of norm-expressivism would be disproven on any theory of emotion. If the facts somehow show that persons are not able to experience moral emotions in response to circumstances alone (independent of making moral judgments), then no theory of emotion would offer the norm-expressivist an account for guilt and anger as they desire. We can note that these compatible theories of emotion for the norm-expressivist do not require that moral emotions take place in the absence of moral judgments or moral concepts. They just, theoretically, allow for the possibility that this is not the case. Should the moral emotions turn out to require an understanding of moral concepts after all (if there really is no satisfactory way of defining their content/themes in non-moral terms), for example, then such theories may imply a circularity to norm-expressivism at the level of defining “wrongness” anyhow.

There are certainly other facts to be discovered that would impact the credibility of norm-expressivism. Should the facts show that anger and guilt are not better motivators for social behavior than are positive moral emotions (such as compassion, for example), for example, then norm-expressivism loses some grounding as well. Should the facts show that all emotions help to shape each other equally, then this too would undermine norm-expressivism. (In this case, there is no reason to suppose that anger and guilt, in particular, are the references of a moral judgment.) The merit of norm-expressivism will certainly depend upon such empirical facts.

Still, we should (per Chapter One) note that such research on the nature of moral judgments and moral emotions is hardly conclusive. The research itself seems to presume metaethical claims and theories of emotion. Thus, while we can recognize that the compatible theories of emotion are here shown to compliment norm-expressivism in a mitigated way, the threat of contradicting facts does not appear to be a looming blow to norm-expressivism in my view.

What this project hopefully shows is that there are theories of emotion that are theoretically consistent with norm-expressivism. The norm-expressivist need only hope that the necessary factual claims (as
implied by norm-expressivism) turn out to be true and that at least one of these favorable theories of emotion is indeed cogent in its own right.
Works Consulted


217
Gibbard, Allan. “Reply to Blackburn, Carson, Hill, and Railton.”


