FRAMING EMOTION: CONCEPTS, CATEGORIES, AND META-SCIENTIFIC FRAMEWORKS

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI'I IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN PHILOSOPHY

MAY 2008

By Kyle R. Takaki

Dissertation Committee:
Ron Bontekoe, Chairperson
Mary Tiles
James E. Tiles
Arindam Chakrabarti
Ben Bergen
MISSING PAGE NO.

AT THE TIME OF MICROFILMING / Scanning
ACKNOWLEDGEMENTS

Thanks to my dissertation committee, Ron Bontekoe, Mary Tiles, James Tiles, Arindam Chakrabarti, and Ben Bergen, for the time they invested into this project. Special gratitude goes to my three main “co-advisors,” Ron Bontekoe, Mary Tiles, and Jim Tiles, under whom I have apprenticed through my graduate years. Without your support and critical feedback, I would not have been able to grow as a scholar and philosopher. And even though I may go off to explore new fields of philosophical inquiry, I will always carry—tacitly or explicitly—the innumerable lessons gleaned from my years of apprenticeship.

Finally, thanks to my family, especially Mom and Dad, who supported me (with trepidation) in spite of the fact that I decided to become, of all things, a philosopher. This work is dedicated to you.
ABSTRACT

The following work is in part about framing various sciences of emotions. In addition to some modest interest in organizational issues pertaining to differing scientific research agendas, my further interest in focusing on emotions, specifically scientific perspectives on emotions, is that emotions are a type of "borderland" phenomena that, on the one hand, concern the felt experience of things—consequently they have an air of ineffability about them, making emotions the sort of topic apparently not amenable to appropriate scientific investigation—and on the other hand, concern neuro-physiological processes, observable behaviors, etc. which are to an extent "quantifiable"—consequently they appear to be the sort of topic a sophisticated science (pertaining to consciousness, artificial intelligence, and so on) would strive to understand. In other words, emotions are both a central part of human experience, and a "grail" which cognitive science seeks to more fully grasp.

What prospects might science offer us in way of illuminating various emotive phenomena? The question I seek to explore (but not definitively answer) is whether a science of emotion is possible. To be sure, there are particular sciences, each of which investigates a restricted range of emotive phenomena. However, a science of emotion is broader than these particular investigations. Hence the question is really, what sense, if any, does it make to speak of a "science" of emotive phenomena, most generally speaking? There are two sciences that I bring to bear on this issue: biology—or more properly, the evolutionary framework by which biology "hangs together"—and the new "science" of complexity. I place most of the emphasis on evolutionary considerations; the concepts associated with complexity make their appearance in the final two chapters.
What is at stake in exploring whether there can be a science of emotion is the question: Can there be a science of mind? Given the contemporary acknowledgement that apparently all "mind processes" are emotive-cognitive processes—or simply "affective systems"—modern research programs seek to better understand these affective systems in the search for a "science" of "the mind."
TABLE OF CONTENTS

Acknowledgements ..................................................................................................................................... iii
Abstract .................................................................................................................................................. iv
List of Figures ....................................................................................................................................... xiii
Preface ................................................................................................................................................... xiv
Chapter 1: A Science of Emotion? ........................................................................................................ 1
  Prospectus ............................................................................................................................................. 1
  1.1: Natural Kinds, Projectibility, and Causal Homeostasis ............................................................... 2
    Natural Kinds..................................................................................................................................... 3
    Causal Theories of Reference ............................................................................................................. 4
    Problem of Induction .......................................................................................................................... 6
    Induction and Projectibility ............................................................................................................... 7
    Causal Homeostasis ........................................................................................................................... 9
    Putting the Pieces Together .............................................................................................................. 10
  1.2: Can Cognitive Science be a Science? ............................................................................................ 12
    Structure of the Argument .................................................................................................................. 14
    Belief Fixation is Isotropic and Quinean ............................................................................................ 15
    The Analogy Between Confirmation and Central Processes ........................................................... 16
    Natural Kinds and Folk “Kinds” ......................................................................................................... 18
    Different Sciences, Different “Kinds” ................................................................................................ 21
  1.3: Summary........................................................................................................................................ 23
Chapter 2: Science and Heterogeneity ........................................................................................................ 26
  Recapitulation ...................................................................................................................................... 26
  Prospectus ............................................................................................................................................. 26
  2.1: Autonomy and Biology .................................................................................................................. 29
    Heterogeneity and Ecology .................................................................................................................. 29
    The Qualified Autonomy of Biology .................................................................................................. 30
    Autonomy and Nomic Generalizations ............................................................................................... 32
  2.2: A Conceptual Space for Questions of Autonomy ........................................................................... 34
    The Causal-Phenomenological Axis ................................................................................................. 34
    The Concrete-Abstract Axis ................................................................................................................ 35
    Capacities and Tendencies ................................................................................................................ 37
    The Fidelity Axis ............................................................................................................................... 39
    Implications of the Three-Dimensional Space .................................................................................. 40
  2.3: Paradigms, Research Programs, and Philosophical Frameworks .................................................. 42
    Kuhn and Paradigms ......................................................................................................................... 43
    An Overview of Kuhn ....................................................................................................................... 47
    Lakatos and Research Programs ....................................................................................................... 48
    Biology and Research Programs ....................................................................................................... 50
    Appropriating Research Programs .................................................................................................... 51
    Laudan’s Reticulational Model .......................................................................................................... 52
    Appropriating the Reticulational Model ............................................................................................ 53
    Philosophical Frameworks ................................................................................................................ 54
  2.4: Summary......................................................................................................................................... 56
  2.5: Looking Ahead .............................................................................................................................. 58
Competing Research Programs within “Evolutionary Psychology” (Broadly Construed) ................................................................. 106
The Santa Barbara School, Behavioral Ecology, and Developmental Systems Theory (“DST”) ................................................................. 109
A General Overview of the Santa Barbara School’s Research Program: Spaces for the Human Sciences? ........................................................ 110
4.2: The Functional Stance to Information ................................................................. 112
Organizing Affect Programs and FIIPS ................................................................. 113
A Conceptual Approach to “Function” ................................................................. 113
Some Brief Historical Background ................................................................. 114
The Functional Stance ......................................................................................... 114
Terminological Clarifications ................................................................................ 115
The Intentional Stance ......................................................................................... 116
4.3: Summary ........................................................................................................ 119
Chapter 5: Cognitivist Approaches to Emotions ................................................................. 121
Recapitulation ........................................................................................................ 121
Prospectus ............................................................................................................. 121
5.1: The Cognitivist Program ................................................................................... 123
Distinguishing the Modern Cognitivist Program from Two Other Traditions ........................................................................................... 123
An Overview of the Argument ........................................................................... 125
What is “Dimensional” About Dimensional Approaches? ................................ 125
Linking the Cognitivist Program to Dimensional Approaches ................................ 128
Appraisals as Modes (Systems) of Activation ................................................... 131
The Debate Between Zajonc and Lazarus ........................................................... 132
Appropriating Zajonc and Lazarus into the Cognitivist Program ............................. 133
The “Continuum” Between Discrete and Dimensional Approaches ......................... 135
The Tacit Communicative Dimension of Appraisals .............................................. 135
Emotions and the Commitment Problem ................................................................ 136
Emotions as Game-Theoretic Equilibria ............................................................... 137
The Communicative Dimension of Game-Theoretic Equilibria (“Strategies”) ................................................................. 139
5.2: Developmental Systems Theory and the Strategic Stance ................................ 141
Distinguishing “Static” Equilibria and Evolutionary Equilibria ................................ 142
Summing Up ......................................................................................................... 143
Overview of the Remaining Argument .................................................................. 144
Developmental Systems Theory .......................................................................... 145
Some Differences Between DST and the Cognitivist Program .............................. 147
The Strategic Stance ............................................................................................. 149
5.3: Summary: Relating the Strategic Stance to the Four-Dimensional Framework ........................................................................................... 151
Chapter 6: Social Constructivist Approaches to Emotions ........................................ 154
Recapitulation ........................................................................................................ 154
Prospectus ............................................................................................................. 154
6.1: The Social Constructivist Program .................................................................... 155
Overview of the Argument .................................................................................... 155
Strong Constructivism and the Conservative Dimensional Approach .................. 156
Strong Constructivism and the Less Conservative Dimensional Approach ........ 157
Dynamic Nominalism and Higher-Order Entities .................................................. 159
Weak Social Constructivism and Overt Construction ................................................................. 161
Weak Social Constructivism and Covert Construction: The Need for a Higher-Order
Perspective on Emotion-Systems ............................................................................................... 162
Social Syndromes and Dynamic Nominalism: Emotions as Covert Social Constructs
(Higher-Order Appraisals) ........................................................................................................ 163
Interpreting Social Syndromes: The Importance of Group-Level and Individual-Level
Perspectives ............................................................................................................................... 164
The (Perspectival) "Interests" of Syndromes: Group-Level, Individual-Level, and "Mixed"
Perspectives .............................................................................................................................. 165
Averill’s Three Systems Situating Emotive Social Syndromes (as Modal Systems) .......... 165
Developmental Systems Theory (DST) and Social Constructivism ....................................... 167
Dual Inheritance Theory (DIT): Distinguishing the Cognitivist Program from the Social
Constructivist Program ........................................................................................................... 168
Distinguishing Dual Inheritance Theory (DIT) from Developmental Systems Theory
(DST): Creating a Space for Differing Philosophical Frameworks ...................................... 173
6.2: The Semiotic Stance ........................................................................................................ 175
Peirce on Semiotics .................................................................................................................. 176
Inquiry and Dispositions ......................................................................................................... 176
A Triadic Framework for Understanding Dispositions: Sign, Signified, and
Interpretant ............................................................................................................................... 177
Semiotics as a Dispositional Framework ................................................................................. 178
Inquiry and Semiotics: A General Framework for Projectibility .......................................... 179
The Semiotic Stance Encompasses the Functional and Strategic Stances ......................... 180
The Semiotic Stance and Philosophical Frameworks .............................................................. 181
6.3: Overview .......................................................................................................................... 182
Chapter 7: Biology and Psychology: The Link Between Evolution and the Four Emotions
Research Programs .................................................................................................................. 184
Recapitulation .......................................................................................................................... 184
Prospectus .............................................................................................................................. 184
7.1: The Parallel Between Biology and Cognitive Psychology .............................................. 186
Three Reasons for the Parallel Between Biology and Cognitive Psychology ....................... 186
The Psychological Side of the Diagram .................................................................................. 187
The Biological Side of the Diagram ...................................................................................... 188
The Parallel: Two Interrelated Senses of "Function" .............................................................. 190
Relating the Diagram to the Functional Approaches to Emotions ....................................... 191
7.2: Evolution and Function ................................................................................................ 192
Ariew’s "Reconstruction" of Mayr ........................................................................................ 193
The Reconstructed Proximate-Ultimate Distinction ............................................................... 194
Reconstructing the Ultimate Side of the Distinction ............................................................... 195
Does Evolution Range Over Individual Level Causal Properties? ....................................... 196
Defending Mayr ...................................................................................................................... 198
An Architectural Overview of the Received View: A Propensity Interpretation of
Fitness ...................................................................................................................................... 199
Summary ................................................................................................................................. 201
7.3: The Received View and Functional Approaches to Emotions ................................... 202
A Complete Account of an Adaptive Explanation ................................................................ 202
Linking the Parallel Between the Received View and Cognitive Psychology with the
Functional Approaches to Emotions ...................................................................................... 203
Emotions and Systems ................................................................. 257
Levels of Investigation ................................................................. 258
Complex-Systems as Representations ........................................... 258
Conceptualizing FIIPS as Complex-Systems ................................. 259
Conceptualizing Dynamical Modes as Complex-Systems ............... 260
9.2: Tools for Complex-Systems Thinking ....................................... 261
A Sketch of the Conceptual Complexity Framework ..................... 262
Individuating Complex-Systems on the Basis of Their (Dispositional) Behaviors .................................................. 263
Phase Spaces and Dispositions ...................................................... 265
Information and Classifying Systems as Complex .......................... 266
9.3: Emergence From an Epistemic Point of View ............................. 269
Elemental Descriptions and Constituent Descriptions ..................... 269
Auyang’s “Left-Over” View of Emergence ...................................... 270
Non-Resultant Characters as Emergent Characters ....................... 271
The Second Interpretation: The Denial of One Condition but not the Other ................................................ 272
Case (A) ....................................................................................... 272
Case (B) ....................................................................................... 273
The First Interpretation: Denying Each Condition ......................... 273
The Denial of (1): Epistemically Disclosing Dispositions ................ 274
The Denial of (2): The Category of Disposition as a Condition for Understanding Emergence ............................... 276
9.4: Levels, Thresholds, and Stances ............................................... 278
Levels, Perspectives, and Causal Thickets ....................................... 279
Mapping Levels-Perspectives-Causal Thickets to the Three Informational Stances ............................................. 280
What is “Information”? .................................................................. 281
What is a “Stance”? ...................................................................... 282
Why Three Informational Stances? ............................................... 283
Complexity and the Three Informational Stances ............................ 284
9.5: Toward a Science of Emotion ..................................................... 286
Overview of Chapters Two Through Nine ...................................... 286
Overview of the Remaining Argument ......................................... 287
9.5.1: Emotion and Affect .............................................................. 288
Two Reasons for the Shift from “Emotion” to “Affect” ...................... 288
“Emotion” ................................................................................. 289
“Feelings” ............................................................................... 290
“Attitudes” .............................................................................. 290
“Mood” ................................................................................... 291
“Affective Style” ...................................................................... 292
“Temperament” ........................................................................ 292
9.5.2: Affect and Complexity: Borderland Research Programs ........ 293
The Functional and Developmental-Social Approaches to Affects and the Three Informational Stances .............. 293
Speculations about Future Borderland Accounts ......................... 294
9.5.3: The Category of Emotion: Complexity as a Framework for Heterogeneity? ........................................... 295
The Qualia Objection ................................................................. 296
Structure of the Remaining Argument .......................................... 296
The Bigger Picture: The Four Affect Research Programs and Their Tangled Relations to the Human Sciences .... 297
DIT and DST: Differing Conceptions of "Science" ....................................................... 298
Evolution as an Indefinitely "Evolving" Science .......................................................... 300
Consilience of the Category of Emotion? Emotion as a Regulative Ideal .............. 304
Glossary ....................................................................................................................... 308
Bibliography ................................................................................................................ 321
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The (Modified) Parallel Between Biology and Cognitive Psychology</td>
<td>28</td>
</tr>
<tr>
<td>2. The Causal-Phenomenological Axis and the Concrete-Abstract Axis</td>
<td>37</td>
</tr>
<tr>
<td>3. The Four-Dimensional Framework</td>
<td>57</td>
</tr>
<tr>
<td>4. The Three Informational Stances and the Four-Dimensional Space</td>
<td>59</td>
</tr>
<tr>
<td>5. Reverse Engineering and the Four-Dimensional Space</td>
<td>81</td>
</tr>
<tr>
<td>6. The Three Informational Stances and the Four-Dimensional Space</td>
<td>82</td>
</tr>
<tr>
<td>7. Encapsulation, Inaccessibility, and Domain-Specific Inputs</td>
<td>97</td>
</tr>
<tr>
<td>8. Tweaking Encapsulation and Inaccessibility</td>
<td>98</td>
</tr>
<tr>
<td>9. The Functional Stance and the Four-Dimensional Space</td>
<td>120</td>
</tr>
<tr>
<td>10. The Three Informational Stances and the Four-Dimensional Space</td>
<td>183</td>
</tr>
<tr>
<td>11. The Parallel Between the Received View and Cognitive Psychology</td>
<td>187</td>
</tr>
<tr>
<td>12. The Architecture of EvoDevo</td>
<td>209</td>
</tr>
<tr>
<td>13. The Parallel Between DST and Developmental Orientations in Psychology</td>
<td>216</td>
</tr>
<tr>
<td>14. A Reductive View</td>
<td>229</td>
</tr>
</tbody>
</table>
As the title “Framing Emotion: Concepts, Categories, and Meta-Scientific Frameworks” suggests, the following work is about framing various sciences of emotions. Besides some modest interest in organizational issues pertaining to differing scientific research agendas, what further interest motivates the focus on emotions, specifically scientific perspectives on emotions? The answer is (relatively) straight-forward: emotions are a type of “borderland” phenomena that, on the one hand, concern the felt experience of things—consequently they have an air of ineffability about them, making emotions the sort of topic apparently not amenable to appropriate scientific investigation—and on the other hand, concern neuro-physiological processes, observable behaviors, etc. which are to an extent “quantifiable”—consequently they appear to be the sort of topic a sophisticated science (pertaining to consciousness, artificial intelligence, and so on) would strive to understand. In other words, emotions are both a central part of human experience, and a “grail” which cognitive science seeks to more fully grasp. (For example, the “grail” of understanding consciousness—a project much in vogue these days—apparently has the deep problem to address: how do mechanistic neuro-physiological processes give rise to the what-it-feels-like of consciousness? Even though consciousness is distinguished from emotions, it remains that consciousness is often presumed to centrally involve affective states.)

More generally, this “borderland tension” between lived experience and understanding such experience may in part account for the diverse range of conceptions of emotions. For instance, witness the peculiar status of emotions in the history of philosophy. They have alternately been viewed as irrational, and as (quasi) rational; as lower types of ("non-cognitive") motivational factors, and as judgments; as something that ought to be the slave of reason (along with the “converse,” that reason is, and ought to be, the slave of the passions); as potentially
noble, and as "bestial"; as subject to cultivation, and as "fixed"; and so on. Perhaps part of the 
reason for these variegated characterizations of emotions has to do with the variegated "nature" of 
emotive experience, as well as the variegated ways in which we represent, or conceive of, 
"emotions."

My concern is less with the experiential and phenomenological features of emotions, and 
more with what it means to speak of scientific understandings of emotions. Accordingly, I focus 
mainly on third-person accounts of emotions provided by various research programs; I will not 
discuss aspects of emotional experience that may be out of the reach of scientific inquiry. For 
example, if first-person experiences of emotions are centrally about the "ineffable" quality of 
what-it-feels-like, akin to the qualia of consciousness, then perhaps no scientific account can 
capture such emotive qualia. Indeed, debates in philosophy of mind concerning consciousness 
often turn on the qualia problem—briefly, how could materialistic processes give rise to the 
qualia, the "what-it-feels-like," of conscious states? Interpreting "materialistic processes" as the 
domain of scientific investigation, and interpreting "qualia" as the domain of felt, first-person 
experiences of emotions, it may be that "science," by its very nature (as more generally providing 
third-person accounts), cannot capture these experiences.¹ Thus while scientific understandings 
of emotions often utilize reports regarding first-person experiences, I focus primarily on the 
(third-person) conceptual underpinnings of research programs.

What prospects might science offer us in way of illuminating various emotive 
phenomena? The question I seek to explore (but not definitively answer) is whether a science of 
emotion is possible. To be sure, there are particular sciences, each of which investigates a 
restricted range of emotive phenomena; however, a science of emotion is broader than these 
particular investigations. Hence the question is really, what sense—if any—does it make to speak

¹ Note that the claim isn't merely that third-person accounts differ from first-person accounts, which is 
trivially true. Rather I have in mind debates concerning the "hard problem" of consciousness.
of a "science" of emotive phenomena, most generally speaking? There are two sciences that I bring to bear on this issue: biology—or more properly, the evolutionary framework by which biology "hangs together"—and the new "science" of complexity. I place most of the emphasis on evolutionary considerations; the concepts associated with complexity make their appearance in the last two chapters.

I examine specific emotions research programs from an evolutionary viewpoint because of several conceptual parallels between the programs and evolutionary research agendas. Broadly speaking, I examine two types of approaches to emotions paralleling a general debate within biology: the parallel concerns "functional" approaches to emotions, and developmental approaches to emotions. I explain how two functional approaches to emotions—Paul Ekman et al.'s work on affect programs, and evolutionary psychology—adopt an evolutionary perspective (specifically, the "Modern Synthesis" view). By contrast, the two developmental approaches I consider—the "Cognitivist Program" and the "Social Constructivist Program"—exhibit significant parallels to another evolutionary view, namely "Developmental Systems Theory."

And yet in spite of these parallels between developmental emotions research programs and developmental approaches within evolution, there are also elements in the former making irreducible use of ideas stemming from the "human sciences" (which includes an open-ended array of humanistic disciplines). In brief, I will describe 1) "borderland" functional accounts of emotions whose borders are situated between cognitive psychology and the Modern Synthesis view of evolution; and 2) "borderland" developmental accounts of emotions whose borders are situated between the human sciences and certain biological considerations. I argue that concepts associated with complexity potentially frame these borderland accounts.

---

2 I expost this claim in Chapter 2 and Chapter 7.
3 Another way to speak of these "borderlands" is in terms of Dilthey's distinction between Naturwissenschaften ("natural sciences") and Geisteswissenschaften ("human sciences"). Briefly, the
The structure of the argument divides the issue of whether a science of emotion is possible into several parts. In Chapter 1 I discuss the nature of the problem; that is, I expound several arguments for why a science of emotion probably cannot be established. Chapter 2 then steps back from these critiques and addresses the broader issue of what is meant by “science.” Chapter two gives a conception of science adopting a “heterogeneous perspective,” which differs markedly from the picture of science expounded in chapter one. In particular, chapter two offers a conceptual space to organize different kinds of sciences (e.g., paradigms, research programs, etc.)—and potential disputes over differing conceptions of “science.” I then apply this framework, in Chapter 3 through Chapter 6, to several of the major scientific approaches to studying emotions. In other words, chapters three through six examine actual emotions research programs, using the framework from chapter two to align their differing aims and disciplinary structures.

Chapter 7 proceeds to link the emotions research programs covered in chapters three through six with the concepts and tools coming from biology and cognitive psychology (as well as other human sciences); this chapter attempts to justify why the links make sense. Chapter 8 introduces complex-systems thinking, and its potential application to framing certain architectural issues in biology. Additionally, given the links discussed in chapter seven between the emotions research programs and biology, the complexity framework also offers a set of tools for conceptually framing aspects of the research programs covered in chapters three through six.

In Chapter 9 I explore whether a science of emotion is possible. The previous chapters suggest a theoretical map for why there is reason to hope that a science of (particular) emotions may be possible. The transition to the general category of emotion is much more difficult, and of

natural sciences emphasize explanation and the search for cause-effect relations; the human sciences emphasize understanding human experience in terms of various part-whole relations. The borderland accounts mentioned above concern \textit{variegated} conceptions of how the natural sciences (specifically the biological sciences) relate to the human sciences.
necessity, highly speculative. Thus chapter nine speculates about what bearing complex-systems thinking has on a science of emotion. I engage in some semantic “spadework” pertaining to the use of “emotion” and its relation to the emerging field of affective science. I then revisit the question, what sense does it make to speak of a “science” dealing with a subject matter as general and variegated as emotion? Complexity and evolutionary considerations occupy center stage in addressing this question. The upshot will be a deflationary (fallibilist) conclusion: it is not clear if the category of emotion even exists; all we can say is that there may be an indefinite, “evolving” science of various emotions—a science containing particular research programs and the borderland relations between these programs. If the general category of emotion serves any purpose, it would be as a regulative ideal guiding the search for consilient relations between fields of research. 

As mentioned above, in the last chapter I do some semantic spadework on “emotion.” I save this discussion for the last chapter since it is most profitably viewed after much of the “heavy lifting” done in the previous chapters. However, I should flag the reader’s attention to an important point: “emotion” is a term of art employed by various research programs. Much of the exposition that follows may seem to suggest that “emotion” is not quite the proper term to apply to what is under discussion. Actually, I employ “emotion” not only as a term of art, respecting its usage by the major research programs, but also as a placeholder for the recent and more appropriate term “affect”—especially since most modern research programs claim that it is misleading to strongly distinguish “emotion” from “cognition”/reason. The widening range of scientific investigations about “emotive” phenomena has led to a recent shift in terminology reflecting this expanding demand. There are two general reasons why the shift from “emotion” to “affect” makes sense. First, because “emotion” carries certain pre-theoretic associations

---

4 By “consilience” I mean the unexpected coherence and convergence of results from independent lines of investigation from (semi) autonomous fields of research.
concerning felt subjective states—in particular, “emotion” usually connotes an affective state, of short duration, whose felt quality has an intensity and “texture”—these phenomenological aspects of emotion exhibit tangled relations to scientific senses of “emotion.” The second reason for the shift from “emotion” to “affect” is due to the internal demands of scientific practices, which require an increased range of distinctions concerning affective phenomena. In other words, given the modern acknowledgement that “emotion” is not strongly distinct from “cognition”—even stronger, it appears that “the mind” consists of heterogeneous emotive-cognitive processes—scientific intuitions have shifted about what emotions are; for while “emotion” is still a term of art, as research has evolved, so has the need for better working characterizations of various “affective phenomena.” Since “affect” connotes a wider category compared to pre-theoretic intuitions about “emotion,” the new field of affective sciences includes a diverse array of research programs.

To reiterate, explicit discussion of affect terminology will be saved for the final chapter. However, to give a foretaste of what’s to come, both in the final chapter and other chapters, what is called “emotion”—according variegated uses of the term by research programs exploring different emotive-cognitive systems—may include moods (a “diffuse affective state”), personality traits and habits (complex affective-socialized systems), specific types of “knee-jerk” reactions (what are called “affect programs”), belief-desire complexes (“affective attitudes”), emotions in other animals (ethological studies concerning the expression of “emotions” in non-human animals), and so forth. In other words, given the contemporary acknowledgement that apparently all “mind processes” (especially what will be called “centralized systems,” discussed in Chapter 1) are emotive-cognitive processes—or simply “affective systems”—modern research programs seek to better understand affective systems in the search for a “science of the mind.” Is the mind
nothing but affective systems? Nobody knows, and perhaps nobody will ever know. Still, the hope to better understand the dimensions of (centralized) affective systems persists.

It may be objected that I am exposing myself to a methodological hazard by using “emotion” as a term of art, and also as a placeholder for the recent and more appropriate term “affect.” The related objection might be raised: can’t I define or characterize more tightly (via conceptual analysis) what ”emotions” are, to ”locate” my project, and to give some sense of what I am contributing to philosophical scholarship on emotions? Part of the problem with these objections, while legitimate, is that 1) nobody really knows what ”the emotions” are; and 2) the question “what are emotions?” may not be an answerable question, because in small form answering this question amounts to answering the ”million-dollar” question: Can we establish a ”true” science of the mind? (Or to use variations on this theme: What is the solution to the “Frame Problem”? What is the solution to the “Binding Problem”?) To reemphasize, since nearly all modern researchers acknowledge that ”emotion” and ”cognition” cannot be divided—indeed, all mind-processes apparently are emotive-cognitive processes—being able to define or tightly characterize emotions would amount to solving perhaps the hardest problem in all of science, namely, can we establish a genuine science of the mind?

Instead, what I offer is a more prudent conceptual extraction: there are different research programs investigating what they think emotive-cognitive systems, of certain sorts, are (I discuss these types of systems in chapters three through six). I also attempt to organize their differing structures, and I engage in some minimal speculation about the possible interrelations between the programs. I try to keep such speculation minimal, since philosophers, these days, probably ought not to offer ”prescriptive” views of what ”science” ought to do. Furthermore, attempting to locate ”what emotions are,” via conceptual analysis satisfying pre-existing philosophical
intuitions, would, I think, make such philosophical analysis even more out of touch with the modern explosion of interest and research on affects.

In sum, figuring out “what emotions are” is messy because the term “emotion” may tap into folk intuitions, philosophically and culturally informed intuitions, scientific intuitions, and so on. My focus, again, will be on various senses of “emotion” from the standpoint of several major scientific research programs. Accordingly, since these programs utilize a fair amount of technical language, I will appropriate technical terms to reference the major concepts of these programs. In addition, a number of technical philosophical terms will be used to frame and interpret some of these concepts. Important technical terms, whether philosophical or “scientific,” will be in bold text when introduced. (I put in bold mainly those terms/concepts that have import for later chapters.) For assistance, brief characterizations of these terms will be given in the Glossary. Since these technical terms will be employed, and built upon, in later chapters, the Glossary is intended to help the reader follow the “long argument” to come.
Chapter 1: A Science of Emotion?

Prospectus

Can there be a science of emotion? If we consider that a colloquial sense of “emotion” involves an affective state that is usually of short duration and whose felt quality has an intensity and “texture,” it may appear (pre-theoretically) that the prospects for such a science are not bright. Indeed, the subtle and variegated nature of what may count as an emotion is often taken to be a reason why providing a “hard” scientific account of emotion might be an implausible—if not impossible—undertaking. But surely, for those who are sympathetic to a possible science of emotion, this project is not only possible but in fact is already under way in certain areas of psychology, endocrinology, neuroscience, and so forth. The relevant sub-disciplines investigate particular emotions or particular emotive processes. The question, then, is given that there are sciences dealing with particular emotions, can there be a science that establishes a general theory of emotion? That is, can there be an account of various particular emotions and an account of how they hang together under the general category of emotion—if there is even a single unitary thing?

In the sciences, as elsewhere, there are “splitters” who isolate phenomena and then proceed to analyze the phenomena into their parts, and “groupers” who prefer to start with the “forest” and then proceed to figure out how the “trees” contribute to the whole. Both approaches have problems. For a skeptic of a science of emotion, there are the respective objections that even if splitters investigate particular emotions, that still does not give us any comprehensive view of the general category of emotion as such (whatever this—if it exists—may be); and even if we start with the general category of emotion, groupers still give us no systematic way to organize the putative parts of the category. It is clearly artificial to claim that any practicing
scientist is either exclusively a splitter or a grouper. More relevantly, most scientists who research emotions are a little bit of both; the issue really has to do with where a balance is struck. Still, whatever balance between splitting and grouping is given, the question remains whether there can be a general theory of emotion.

Paul Griffiths, in his important work *What Emotions Really Are*, argues that while there is a particular scientific account of certain emotions (the “psychoevolutionary approach”; see fn.1 below), a science of emotion probably cannot be established—that is, there (probably) can be no general theory of emotion. A helpful (although not wholly accurate) picture of what he argues for is that the psychoevolutionary research program, which strikes a balance more on the splitters’ side, offers a genuine scientific account of certain emotions; but as research programs lean further toward the groupers’ side (what he calls “ecological” approaches to emotions), their claims to being scientific become increasingly dubious. What I will do in this chapter is first to outline Griffiths’ argument as to why there can be no general theory of emotion.¹ Then in section 1.2 I situate some of these ideas in light of the debate over whether “cognitive science” is—and whether it can be—a science.

1.1: Natural Kinds, Projectibility, and Causal Homeostasis

Why can there not be a science of emotion? The argument for this claim appeals to certain intuitions about what science does, or ought to do, and how emotions generally do not conform to these intuitions. There are two intuitions in particular that form the backbone to Griffiths’ argument. One intuition is that scientific categories ought to refer to (natural) kinds in

¹ Aspects of Griffiths’ account of the psychoevolutionary approach receive discussion in Chapter 3. Also aspects of his coverage of other approaches receive discussion in chapters four through six. The main concern in this chapter is with Griffiths’ critique of a science of emotion.
the world, and the other intuition is that the way in which science investigates these kinds is by looking for "projectible" categories which exhibit something called "causal homeostasis."

**Natural Kinds**

Natural kinds have a long history in philosophy, although, for the former intuition above, the spin Griffiths gives to natural kinds and the role they play in science is a more modern contribution from philosophy of science. I discuss in Chapter 8 a particular view of how kinds and scientific laws interrelate, so here I shall only be concerned with a broad overview of natural kinds. A useful illustration of a natural kind that Griffiths gives is (the referent of) the category *reptile*. The stereotypical, "commonsense" features of being a reptile include the characteristics of being cold-blooded, not giving birth to "live" young, and so on. Most people, when asked to explain what a reptile is, probably think of examples of reptiles—like alligators—as falling into this category without fully knowing the conditions for membership (although the commonsense features associated with alligators are usually known). Building on this commonsense account of category use, an older philosophical view holds that the world has "natural joints"\(^2\) that enable us to systematically group individuals of whatever sort into neat and tidy categories. Ideally, these categories do not overlap with one another. Additionally, these categories would accommodate a complete list of all individuals. As applied to an antiquated philosophy of science view, an ideal science would specify, for example, a complete biological taxonomy that may contain categories like *reptile* which have all and only those predicates defining the conditions for inclusion. Assuming we could do this, the conditions for being a reptile might include predicates like "is cold-blooded," "gives birth to non-live young," etc. Generalizing this strategy, we ought to search for all scientific categories, where we give the complete conditions for membership—the

\(^2\) Think of the ditty "...the hip bone is connected to the thigh bone...." Finding the "bones" of Nature is what carving Nature at its joints is about.
necessary and sufficient conditions—for each category. The resulting perfectly “tiled”
categorization scheme would be one thing that an ideal science yields.

Unfortunately this picture is wrong. Nature doesn’t yield necessary and sufficient
conditions for all categories, even assuming that an exhaustive classification scheme is what
science ideally strives for (or which it ought to strive for). Taking the above example about
reptiles, Griffiths argues that not all reptiles are “cold-blooded,” as birds are “warm-blooded” (see
fn.3 below). So we need a revised picture to account for theory revision in the sciences as it bears
upon the issue of categorization. The revised picture that Griffiths offers has its roots in Saul
Kripke’s causal theory of reference and in Nelson Goodman’s work on the problem of induction.
Let me give a brief exposition of each of these ideas.

Causal Theories of Reference

The two ideas that Griffiths highlights in causal theories of reference are that reference
can be preserved across theories for certain theoretical terms (“transtheoretic identity”), and that
reference can be partially preserved as theories change. As an example of the former, Griffiths
cites the category of stars, which in the past was thought to refer to fixed points on crystalline
spheres, but which we today think of differently. The point is that even though modern
astronomical theories give a very different worldview by which stars are understood, the
reference is “the same”—namely, “actual” stars. (Griffiths knows this is not quite right, as I will
discuss later.) As for the latter idea, the reptile example shows that certain biologists changed
how they view classification in light of phylogeny, which changed the reference of the term
“reptile” by increasing its scope (to accommodate birds), and also its sense /intension because the
predicates used to describe members of that class significantly shifted by including certain
endothermic creatures in addition to the traditional ectothermic organisms. Griffiths rightfully places emphasis on the latter idea, as the issue isn’t whether science in general preserves reference across theory change, but rather how partial reference is to be accounted for when theories change (hopefully—and usually—for the better). In particular, the way in which causal theories of reference apply to theory change in science is through “causal homeostasis.”

The “causal” aspect of “causal homeostasis” gets its name from causal theories of reference. A part of the way in which theoretical terms can preserve reference across theories has to do with their causal history. That is, if the category stars in part refers to our experiential phenomena of twinkling stuff in the night sky, then the ways in which theories change in accounting for the stars’ location and constitution still rely on a discourse surrounding such phenomena; and such discourse in turn relies upon a historical fixing of the reference. One can imagine someone, say “Astronomical Eve,” pointing up at the heavens and naming those points of light. This historical fixing then forms a causal history when chains of communication are formed that “pass the reference” from one speaker to the next. Also, causal theories are causal in the added sense that we fix certain terms of reference (again, in part) by the causal role of the referent in question—for example, the causal effect of the “actual” stars on how we perceive star phenomena. What causal homeostasis shares with the above discussion of older versions of categorization is the presumption that Nature can, in some sense, be cut at its joints, even if the resulting picture is not as tight as finding necessary and sufficient conditions. So even if there aren’t very precise conditions for delineating the category of, say, reptile, there is still some

3 Strictly speaking, from a cladistic perspective classification changed. Also strictly speaking, from this perspective reptile does not form a real kind since it isn’t a monophyletic group (i.e., a taxon with a single phylogenetic ancestral group); so let us say for the purposes of the above discussion that “reptile” is shorthand for the (surviving) monophyletic groups snakes, crocodilians, lizards, turtles and birds.

4 I ought to note that the way I am unfolding the notion of causal homeostasis takes some liberties with Griffiths’ presentation, but the overview is still generally accurate.

5 Tracing the genealogical process is clearly a retrospective enterprise, which significantly complicates matters.
transtheoretic preservation of what the category refers to (causally) when using a commonsense, mid-level classificatory scheme. After all, stereotypical examples like lizards and snakes are still what many people think of when asked to name a reptile. They have become historically fixed in our discourse, and continue to be “causally efficacious” (e.g., through media or zoos).

Problem of Induction

As for the “homeostasis” part of “causal homeostasis,” first note that the concept of homeostasis is generally the idea of the maintenance of a stable state. Philosophically, the idea is that categories doing genuine work in describing the world are “projectible.” So in the reptile example, tracking how sense and reference change in lieu of theory change is still about tracking more stable ways of representing the world. The notion of projectibility, which is so crucial to causal homeostasis and to Griffiths’ argument about why there cannot be a science of emotion, requires a background context in order to understand its application. The notion of projectibility comes from *Fact, Fiction, and Forecast*, where Nelson Goodman discusses what he calls the “new riddle of induction.” The older version of the problem of induction is put forth by David Hume. Briefly, what establishes Hume’s problem is that if one is given a set of premises and a conclusion, there does not appear to be any non-question-begging way in which to derive the conclusion from the premises for any conclusion whose information goes beyond what is contained in the premises, strictly speaking. A homely example illustrating the problem goes as follows. Take the premises that (1) the sun has risen in the past; (2) the sun rises today; therefore (3) the sun will rise in the future. The conclusion does not deductively follow from the premises, for we can imagine that it can be false, as in the case where we, say, destroy (literally) our planet. The plausibility of the conclusion comes from our past experience and the expectations we have regarding those regularities that deal with our experience of the sun’s behavior. The conclusion could be made into a deductive consequence of the premises if the additional premise is added
that all future events will be like past events, or more weakly, that all future experiences will be like past experiences. But the only way we could know if this premise is true is by induction itself—that is, taking this as a conclusion, the premises would be that all past future events were like past past events, and that the same thing holds "currently." But how can we know that future events will be like past events without begging the question? Thus Hume’s problem of induction, in a nutshell.7

Induction and Projectibility

Goodman enters the picture by acknowledging Hume’s problem, but raising the further query as to how Hume even knows in the first place what inductive inferences are worthy inferences that get subjected to the problem of induction. For any number of inferences can be made from premises like the ones above. From premises (1) and (2), why not infer that the sun will not rise tomorrow (or in the future)? Someone who holds that the prospects for the human species are grim indeed might say with exasperation and awe: “my goodness, the sun has risen in my past, and I experience another sunrise today...it is a wonder that we haven’t blown ourselves and the planet to oblivion.” The example Goodman employs to illustrate his new riddle of induction is more “exotic,” at least in the sense that it usually makes non-philosophers who encounter it scratch their heads about the sheer artificiality of the set-up.8 But what matters is that the form of the set-up raises the problem of why certain inductions are scientific inductions versus those that are merely artificial, merely a matter of opinion, and so on.

6 I suppose more accurately (or pedantically), the proper conclusion would be that all future futures will be like future pasts.
7 Note that the general problem of induction also includes problems concerning counter-inductive inferences such as “future events will not be like past events.” The literature on induction is rather large; for further discussion, see, for example, Salmon (1966).
8 Goodman uses the predicate “grue” for “is green” (up until time t) and “is blue” (after time t), and contrasts this with the predicate “is green” (up until and after time t). He then asks: why is the predicate grue not something we are inclined to induct upon when claiming that all emeralds are P? That is, why do we claim that all emeralds are green, and not that all emeralds are grue?
Using our sun example again, the set-up goes as follows. Suppose that there are two predicates, “x rises” and (the ungainly predicate) “x risots,” where “x risots” is taken to mean that “x rises up until time t” and “x does not rise after t” for some future time t. The corresponding inductive inferences given premises (1) and (2) are, respectively, that for all times, the sun satisfies the predicate that x rises, and that for all times up until and after t, the sun satisfies the predicate that x risots. Clearly we are inclined to induct on the former claim and not the later claim. Why? The puzzle of why certain predicates are “good” predicates—what Goodman calls “projectible” predicates—for inductive inferences is what he calls the “new riddle of induction.”

Hume starts from the assumption that certain predicates are good and then raises his problem of induction; Goodman moves the skeptical attitude a step back by asking how we know what predicates are good in the first place. Projectible predicates take (past) correlated properties that are representative of what such predicates can be applied to; what makes them projectible is that they hold up in new instances—they are “good” inductions in that they reliably work in these new instances. This is why the predicate “x rises” is projectible and “x risots” is not—the former has held up in past and current instances while the latter has not.\(^9\)

Why projectibility matters for natural kinds is that if we assume that Nature can be (more or less) cut at its joints, then we get a “solution” to—or more accurately an avoidance of—Goodman’s and Hume’s problems. Both of these are epistemic problems—for Goodman, the problem of how we are justified in knowing what inductions are projectible inductions, and for Hume, the problem of what the proper scopes of inductions are. The ontological import of natural kinds is that the categories which science proffers are projectible, and that the referents of

\(^9\) In slightly different terms, Goodman’s riddle of induction is the question of how to distinguish lawlike or confirmable hypotheses from accidental or nonconfirmable hypotheses. To answer that lawlike hypotheses have projectible predicates whereas the latter to do not is to beg the question, for the issue is about picking out what predicates will reliably work in new/future instances. As Goodman notes, projectibility is a dispositional mark of good inductions, not a solution to the riddle (p.49).
natural kind categories are the kinds in Nature. But this is too strong an interpretation of natural kinds, as Griffiths is well aware. Not only does it miss the (epistemic) point of Hume's and Goodman's problems, it also raises concerns about the previous account of causal theories of reference. For in what sense is the transtheoretic identity of, say, the (category of) stars preserved across theories? What exactly is the reference? Are the "actual" stars the points on crystalline spheres? (Not from a contemporary view.) Planets? (Perhaps, according to current colloquial—but not scientific—usage.) Other "suns"? (OK, but assuming that stars are self-luminous bodies—which includes bodies other than suns.) Our shared experience of the phenomena, even assuming they were fixed historically? (Phenomena for who?) This flat reading of causal theories of reference has built into it a crude realism about "actual" stars. Additionally, as the sense and reference of terms shift when theories change, simply chalking this up to a causal history (i.e., "Astronomical Eve" fixed the reference and we've inherited the relevant discursive practice) does not pay sufficient attention to the internal mechanics of scientific theory change. So how do we preserve some of the intuitions about causal theories of reference—namely that something like a causal history makes sense, as does a "softened" notion of transtheoretic identity (see Griffiths, pp.192-6)—without missing the point of Hume's and Goodman's problems?

**Causal Homeostasis**

The answer lies in emphasizing homeostasis with respect to internal changes in scientific theories. Griffiths offers a revised view of natural kinds (utilizing work by Richard Boyd) that is not dependent upon a metaphysical realism. The amendment to natural kinds that causal homeostasis recommends lies in the claim that projectibility is judged by looking at scientific practices and then asking what the relevant background theories warrant as the best guess for some domain of investigation. This basic insight means that scientific practices try to find
projectible inductions that are "stable" with respect to background theories, evidence, etc.; they are "homeostatic." So the revised picture is that causal homeostasis looks for scientifically projectible inductions relative to an established scientific context, whereupon the causal aspect of causal homeostasis is the postulation of an underlying system of causes explaining why such inductions are projectible in the first place. It is important to note that Boyd's account does not assume realism; it is intended as a rapprochement between realism and empiricism. That is, a realist would be happy with causal homeostasis and would add that the reason why it works is because Nature really has joints; an empiricist, who is by temperament more conservative, would also apparently be satisfied with causal homeostasis because stable projections are stable insofar as they save the phenomena. However, Boyd's account is strongly suggestive of a realist position, at least to the extent that the postulation of a system of underlying causes which explains the projectibility of an induction suggests that a realist view provides the best explanation for why an induction works.

Putting the Pieces Together

The upshot of the above discussion is that Griffiths extends causal homeostasis in two directions to then argue for why there can be no science of emotion. “First, the theory of natural kinds not only gives a good account of certain elements of scientific practice, it captures an important aspect of the formation and use of concepts by humans in general” (1997, p.175). This first move links projectibility in natural science (via “categories”) with projectibility in

---

10 The causal-historical dimension is still there as well, since past postulated systems of causes inform current projections.
11 For the empiricist, I suppose the postulated system of causes can be weakly read as just the manner in which projections save the phenomena.
12 Indeed, Griffiths draws attention to the similarity of Boyd's notion of causal homeostasis with Arthur Fine's "natural ontological attitude" (p.188), which is basically an abductive argument for a kind of realist stance.
13 I will not cover the details of this argument, since it is not crucial to my aim of providing an overview of why there can be no science of emotion. For details, see Griffiths (1997), Chapter 7.
psychology (via "concepts"). The way this link works is by arguing that a certain view of how concept formation occurs is more or less the same as how projectibility works in the natural sciences. "Second, Boyd's rapprochement between realism and empiricism can be extended to a rapprochement between realism and all its rivals. If the theory of natural kinds is a central part of the best scientific account of concept formation and use, then an ability to make sense of this becomes an adequacy condition on any account of how thought and language relate to the world" (p.175). This second move claims that given the first move, any adequate account of how a science describes aspects of our world—including our psychological states—ought to satisfy the projectibility intuition. As it relates to emotions, any purported science of emotion must have concepts describing the relevant psychological entities, and if such concepts are to be scientific they need to be projectible. The projectibility intuition that should be met is one adopting a (soft) "realist attitude"—i.e., even if we cannot know if Nature really has natural kinds, Nature still behaves as if it did. But the problem is that the general category of emotion is too varied, and can be cut-up in too many ways to be genuinely projectible—it doesn't behave as if it has natural kinds. Thus there cannot be a scientific account of "the category" of emotion because it (probably) doesn't exist. We can have concepts about emotion(s), and we can even have a science of certain emotions (e.g., the psychoevolutionary approach), but we cannot have a science of emotion.15

For now this general presentation suffices as a picture of why there cannot be a science of emotion. What I want to highlight in outlining Griffiths' argument is how the question of

---

14 Griffiths notes: "I try to consistently use category to refer to the aspect of reality to which some concept is supposed to correspond. I use concept to refer to a psychological entity" (p.175-6). For expository purposes, I have used "category" as a concept which, when projectible, is purported to refer to natural kinds. Henceforth, I will use both terms interchangeably: categories or concepts, when projectible, are purported to refer to natural kinds.

15 The picture I gave earlier is of help here. Those approaches that lean more toward the groupers' side cannot carve emotion concepts into proper kinds, as there are too many disputable ways to "cut the cake." Thus their putative categories are not projectible, and thus are not genuine, scientific categories.
whether there can be a science of emotion hinges primarily around a certain conception of what
the business of science is about, where projectibility is a crucial component of that conception.
As for the various particulars of Griffiths’ argument, some of these pieces will be more profitably
discussed in later chapters. I now want to turn attention to the broader debate in cognitive
“science” over whether this field can really be a science. Specifically, I will outline Jerry Fodor’s
argument explaining why cognitive science cannot be a genuine science. As I read him, the core
of his argument appeals to certain intuitions about projectibility. There are several reasons for
covering this debate. The first is the point just mentioned, that there are other applications of
projectibility dealing with how one ought to understand science. The second is that Fodor’s
argument is built upon a conception of what the mind is about, and the issues he raises are going
to be helpful in situating some of the conceptual frameworks employed by emotions research
programs that I discuss in later chapters. Lastly, the assumptions regarding what the business of
science is about raises the question of how we ought to think of science. If we are going to
explore whether a science of emotion is possible, we first need to get straight on thinking about
science generally, and in particular with respect to those sciences which deal with heterogeneous
subject matters (like emotions). As will be explained in the following section, Fodor’s argument
implicitly utilizes a distinction between heterogeneous sciences—what he calls the “special
sciences”—and those that are not (e.g., physics). This last issue prompts the need for the next
chapter, “Science and Heterogeneity.”

1.2: Can Cognitive Science be a Science?

Fodor’s answer is “probably not.” What I will do first is to expound his conclusion, and
then work backwards from there to give the reasons for his pessimistic answer. To begin with
unfolding the conclusion, Fodor writes in The Modularity of Mind that “the reason why there is
no serious psychology of central cognitive processes is the same as the reason why there is no serious philosophy of scientific confirmation. Both exemplify the significance of global factors in the fixation of belief, and nobody begins to understand how such factors have their effects. In this respect, cognitive science hasn’t even started” (p.129). Read “no serious psychology” as “no genuinely scientific psychology” and what we have is the suggestion that psychology isn’t scientific to the extent that it is concerned with these things called “central processes.” The reason why psychology isn’t scientific in this respect has to do with a conception of how (successful) science works. As Fodor writes:

The condition for successful science (in physics, by the way, as well as psychology) is that nature should have joints to carve at: relatively simple subsystems which can be artificially isolated and which behave, in isolation, in something like the way that they behave in situ. Modules satisfy this condition; Quinean/Isotropic-wholistic-systems by definition do not. If, as I have supposed, the central cognitive processes are nonmodular, that is very bad news for cognitive science (p.128).

This condition, as employed by Fodor (see his discussion of Goodman, pp.107-8), is broadly the same as what was given in the previous section. (Additionally, in Chapter 8 I discuss Fodor’s implicit use of this condition in arguing for the autonomy of the “special sciences”; specifically, this condition is cashed out in terms of the special sciences’ various uses of “function.”) That is, the intuition here is that genuine science requires projectible categories which ought to refer to natural kinds, and since Fodor argues that psychology has certain categories which are projectible (“modules”)—i.e., which exhibit the above characteristics of being “relatively simple subsystems [etc.]”—but other categories which are not projectible (“central systems/processes”), psychology is not yet serious. Additionally, because “nobody begins to understand how such [central systems] have their [global] effects,” it looks like the prospects for psychology are not bright. And since cognitive psychology in particular is an important subfield of cognitive science, and cognitive science crucially aims to understand central processes, this also spells bad news for the prospects of cognitive “science” ever truly becoming a science.
Structure of the Argument

A brief overview of Fodor's conclusion is as follows. Imagine that parts of the mind-brain operate like factories. Each of these factories has specialized mechanisms that are geared toward taking a certain range of raw materials/inputs and transfiguring them into standardized products/outputs. These factories are called “modules” or “complied transducers” (p.41)—they take the raw material of certain kinds of information, and process the information into ready-made cognitive outputs. The general view of the mind is that it consists of peripheral systems and central systems, where the former are exhibited by modular systems (like the visual system), and the latter deal with “folk psychology”—i.e., beliefs, desires, etc. Modules are localized in that they operate only on certain kinds of information, and yield stereotypical outputs. They are “peripheral” in this sense plus in the added sense that they serve to transduce and compile, say, sensory information getting passed to the central systems. Central systems are global in that they are not localized, and may in principle access any part of the “web of information” coming from peripheral or other central processes. As with the argument in the previous section explaining why there can be no science of emotion, there probably cannot be a science of central processes since there are too many putative ways of “parsing” these folk states—they don’t form genuine natural kinds. So while there can be a science of localized processes, there (probably) cannot be a genuine cognitive science that seeks a robust science of the mind.16 Now for some of the details of Fodor’s argument.

The strategy of the argument first employs a particular view of scientific confirmation, which claims that such confirmation is “holistic.” Fodor then makes an analogy between this view and how central systems operate. Lastly, Fodor interprets the “irreducible nature” of such

---

16 Note that given the rather poor record of philosophical prognostication about what science can and cannot do—as well as philosophy’s dubious portrayals about what science ought to do—Fodor is apparently on shaky ground. If anything, the history of science suggests that we really don’t know the full capacities of “science” (or, for that matter, what future generations will count as “science”).
holistic processes as implying that no localized account (peripheral or otherwise) can yield an account of central processes. There are actually two gaps here: first, if a localized and projectible account were to be given, it wouldn’t carry over to holistic processes like central processes because what the localized account projects “falls short” of the array of phenomena that central processes manifest. Secondly, given a holistic account of central processes, it would not really be projectible because it is underdetermined by the multiplicity of ways in which those central processes can be realized. To put these two points colloquially, even if the splitters account for each of their trees *qua* trees, aggregating those trees don’t reveal the forest(s)—the higher-order pattern(s)—we are after, but just more trees; and if we start with the grouper’s whole, the problem is that there are too many (seemingly arbitrary) ways in which to reconstruct what and how the putative parts contribute to that whole. So in either case we don’t have a projectible account of central processes, supporting Fodor’s negative conclusion.

**Belief Fixation is Isotropic and Quinean**

The first part of the argument about holism claims two closely related things: that scientific confirmation—“the nondemonstrative fixation of belief”—is “isotropic”; and that scientific confirmation is Quinean. As for the former point, what Fodor has in mind by the phrase “the nondemonstrative fixation of belief” is that what “we believe depends on the evaluation of how things look, or are said to be, *in light of background information* about (inter alia) how good the seeing is or how trustworthy the source. Fixation of belief is *just* the sort of thing I have in mind as a typical central process” (p.46). The fixation of belief occurs relative to a network of background information, and how beliefs get fixed in scientific discourse depends on how well they fit with the relevant network. What makes the fixation of belief nondemonstrative, generally, is that the inference which the fixing of the belief represents is a “new” inference—it goes beyond whatever information is thought to be (strictly) contained in the background
information. By "isotropic" Fodor means that scientific confirmation occurs relative to a field of previously established truths of whatever sort, and the facts that are relevant to confirmation can be taken, in principle, from anywhere in this field. So the picture expressed here is that scientific confirmation has a range of truths to draw from, and the hypotheses fitting well with the relevant bits are the ones that are chosen (and believed). Since this process can in principle appeal to any part of the field of background information, scientific confirmation is holistic in that respect.

The second respect in which scientific confirmation is holistic has to do with the closely related claim that it is Quinean—that "the degree of confirmation assigned to any given hypothesis is sensitive to the properties of the entire belief system" (p.107). According to Fodor, this point is illustrated in Goodman's discussion of projectibility. Suppose that there are two hypotheses which are thought to have degrees of projectibility. If both have the same available data and are similarly consistent with that data, which hypothesis ought to be conferred the status of being confirmed? What can settle the dispute is to look at the past histories of each of these hypotheses. How projectible was each hypothesis in the past (assuming that each has a past history)? Take the hypothesis having the greater degree of past successes and claim that it is the one which ought to be confirmed given the current data. In order to do this, what is presupposed is that there is a background web of (scientific) belief not only keeping track of the rates of success for each projectible hypothesis, but also making it possible to compare the current situation and past performances. That is, the degree of confirmation is sensitive to the properties of the entire evolving web of belief.

The Analogy Between Confirmation and Central Processes

At this point an apparent problem arises. If we can account for the process of scientific confirmation and what hypotheses are accepted on the basis of the degree of projectibility, then how does this bear upon the general conclusion that there can be no genuinely scientific cognitive
science? Given that Fodor wants to argue by analogy (between confirmation and central systems), hasn’t he supported the opposite conclusion, namely that there can be a serious cognitive science by looking at a similar story about projectibility and how our beliefs generally get fixed? This confusion stems from the failure to distinguish between the results of the confirmation process—which are open to being evaluated on the criteria of projectibility, fit with past and current data, etc.—and the so-called “logic of discovery” (p.106). In the latter case, how discoveries are made is more or less a mystery, and all we have are loose after-the-fact appeals to holistic webs of belief and the like. More importantly, the point with discovery, as with confirmation, is that the form of the problem is pretty much the same, at least with respect to the analogy that Fodor wants to make. The proper form of the problem raised by scientific confirmation and discovery is about looking at the web of belief and noting that the bits that underlie a confirmation (or a discovery) are a function of the web as a whole—that is, in principle there is full access to any bit in the web, relative to what hypothesis is being confirmed (or formed). What occurs after-the-fact may be projectible, but an account of how this occurs is not; again, given the sensitivity of the bits of information to the entire evolving web, there are too many ways to carve up these processes, and so with, say, a discovery, there is no definitive account of how an “ah-ha!” moment occurs.

The form of the problem raised by confirmation analogizes to central processes, which are paradigmatically about the fixation of belief. Fodor hedges the analogy by noting that to the extent that the fixation of belief is holistic, it is non-modular (or “unencapsulated”). Still, the argument looks circular since the lesson for central processes appears hinged to the case made for scientific confirmation. But scientific confirmation need not be the same as what occurs in central processes (cf. pp.110-1). This point is made more clearly in Psychosemantics: “Confirmation Holism doesn’t, after all, imply holism about meaning” (p.66). One can be
committed to the former, which primarily requires a rejection of a splitter's account of piece-meal confirmation (between, say, observation sentences and theoretical sentences), while at the same time adopting a neutral attitude toward the latter more ambitious claim about semantic holism (which is what central processes are about). The implication is that Fodor is not begging the question in his use of the analogy; the main force of his argument is that the form of the problem is the same in scientific confirmation and central processes.\textsuperscript{17}

**Natural Kinds and Folk “Kinds”**

I should note that Fodor’s argument changes slightly in some of his later works, although the general tenor or the critique remains the same. For example, in *Psychosemantics* Fodor argues that insofar as there can be a scientific psychology, it probably needs to be along the lines of what folk psychology gives us. This is, contrary to appearances, not substantially different from the above argument that occurs in *The Modularity of Mind*. The caveat for the above argument is that from the viewpoint of the cognitive science research program (circa 1983) there cannot be a genuine science of psychology due to the failure of the projectibility requirement. This requirement is situated relative to how peripheral systems interrelate to central systems.

Central systems deal with folk states, and the objection is that while we can give a projectible account of peripheral systems, we cannot carry that approach over to folk states, contrary to what “the” cognitive science research program would like to do. In a slightly different vein, in *Psychosemantics* Fodor claims that if a scientific psychology were ontologically committed to beliefs and desires (i.e., to folk psychology), then that would vindicate commonsense belief/desire explanations (pp.25-6). As he argues, folk psychology just works, and so to the extent that it gets

\textsuperscript{17} Note that Quine’s web of belief trades on an ambiguity between the web as an individual’s belief network, and the web as a community of (individuals’) beliefs. However, Fodor’s emphasis on the form of the problem is not about a dispute over Quine scholarship—as Fodor explicitly notes (1998, p.66)—but rather about the “geographical moral” (p.66) concerning semantic holism and the analogy he (Fodor) wants to make.
things right (especially given that we so heavily rely on it), it is "projectible"—for example, the ways in which we rely on other people's schedules depend on various belief-desire attributions, and such attributions are in some sense reliable in new/future instances.

The above argument from *The Modularity of Mind* appears to be in conflict with the view offered in *Psychosemantics*. The former claims that folk states are not projectible, while the latter speaks of a scientific folk psychology that looks "projectible." The way in which Fodor avoids this apparent inconsistency is by claiming that science generally is still about finding natural kinds. But in the case of folk psychology, it is not "required that the folk-psychological inventory of propositional attitudes should turn out to exhaust a natural kind" (p.26). Rather, the way in which Fodor's scientific psychology—his "representational theory of mind" (which I will not delve into)—deals with this issue is that there is partial co-extension between propositional attitudes (which express folk states) and kinds (or a kind). Fodor simply shifts his attention elsewhere by focusing on what his representational theory provides—namely how intentional states have causal powers.18

So what about the objection from the *Modularity of Mind* that central processes don't form natural kinds? Fodor appears to merely sidestep the problem. However, I think a distinction made above can be applied here. The distinction, again, is that what occurs "after-the-fact" may be projectible, but an account of how this occurs (in "real-time") is not. What occurs after-the-fact is expressible through the vehicle of propositional attitudes (e.g., "I guess she went to the store because she was hungry"), and since folk psychology tends to work, it has a claim to projectibility. But how this occurs is not projectible, because for central processes "it may be unstable, instantaneous connectivity that counts" (1983, p.118), which highlights the idea that

---

18 It is important to note that such causal powers are not to be read dispositionally, for according to Fodor, dispositions are not really causes, since only episodic or occasional sorts of events can be classified as causal. Dispositions and causes will be discussed in the following chapter in the context of understanding "science" more generally.
"unique events" aren't reliable explanations for future instances. This idea can be illustrated by observing that particular acts of spontaneity aren't predictable, although someone with a "spontaneous character" can be relied upon to do spontaneous (stupid?) things; but predicting that a spontaneous act will occur is not the same as predicting what particular exhibition of, say, stupidity will ensue. Spontaneous characters, let us suppose, form a kind. But how particular spontaneous acts get sparked is, well, spontaneous. So if a scientific psychology starts with a folk ontology, even though it cannot account for unique events (like particular acts of spontaneity), it can still remain consistent with the project to find natural kinds.

There are problems with this proposal, as I think Fodor is aware. Firstly, what would prevent the cognitive science research program from adopting a folk ontology in addition to looking for connections between modular and non-modular (e.g., connectionist) processes? Another problem is that even if Fodor's picture of scientific psychology can remain consistent with the project to find natural kinds, the question remains how this really squares with his view that folk states are holistic. For on the one hand, whatever plausibility is lent to the projectibility of folk states stems first from the intuition that folk psychology just works. If it works as well as Fodor seems to think it does, then why not be an optimist in offering the counter-intuition that there can be a serious cognitive science that will map the reliabilities (kinds?) of folk psychology using an extended modular approach? In this case, it looks like the best of both worlds—the projectibility of modules and the reliable ontology of folk psychology. On the other hand, Fodor's pessimism about the prospects for cognitive science ("Think it can work? Show me—

---

19 The phrase "after-the-fact" is shorthand for processes whose "mechanics" can be worked out (e.g., at the neural level) to yield a projectible account, or for folk processes where success is explained at a reflective level—that is, although we may anticipate the behaviors of others, technically speaking, the projectibility of folk psychology is still a matter of figuring out reflectively that anticipation of the behaviors of others worked in the past (and will probably work in the future); this reflective explanation occurs after-the-fact. These sorts of after-the-fact phenomena are open to public, say, reconfirmation (e.g., a putative theorem by a math genius can be checked, but how the genius came up with the solution, not even she may exactly know).
I’ve yet to be convinced given the past rather lackluster record\textsuperscript{20} rides the holism intuition when it needs to, yet a different attitude is adopted when it comes to natural kinds and propositional attitudes. Why?

**Different Sciences, Different “Kinds”**

Here is a sympathetic reconstruction of Fodor to answer the above concerns. My response to the first concern is simple: I think Fodor would be happy if research programs adopt a folk psychology ontology and then proceed to look for systematic interconnections between this ontology and modular and non-modular (particularly connectionist) processes. Probably any elimination, reduction, or absorption of these three parameters would be met with some skepticism from Fodor. All three of these things seem to have decent traction in certain domains; why not respect such traction in pursuing a cognitive science? As for the second concern, the first point I want to make is that Fodor appears to be skeptical of any research program which attempts to use a modular natural kinds approach to build an account of the mind more generally. Indeed, Fodor thinks that this optimism as found in evolutionary psychology, which roughly attempts to bridge certain folk states with an expanded “modular” account, is misplaced. I will discuss this in further detail in chapter four, but for now it is enough to note that Fodor thinks that folk psychology, insofar as it gives us some prospect for finding natural kinds, gives us our best chance of finding such kinds in spite of the holistic nature of folk states. Informally, Fodor sides more with the groupers in emphasizing holism, whereas evolutionary psychologists tend to be “more conservative” in placing less emphasis on holism and more emphasis on extending modular ideas.

\textsuperscript{20} As previously mentioned, an especially heavy burden is placed on Fodor’s pessimistic argument given philosophy’s poor performance in predicting the (past) course of science. However, a charitable reading of his critique is that it means to function as a “skeptical check” on overly optimistic predictions about the “new sciences of the mind.” For science, too, has been guilty of promising more than it has delivered.
As for the last objection above, I think Fodor adopts differing attitudes relative to his view of science generally, and to his view of what he calls the “special sciences”—roughly, the social sciences (or more weakly, the non-physical sciences). Science generally looks for natural kinds; what the special sciences look for are functional regularities that have some claim to being (putative) kinds. Biology, for example, is rife with talk of function, and it is a bona fide science. But unlike kinds in the physical sciences, which are supposed to exhibit a higher degree of projectibility, kinds in the special sciences are “fuzzier.” Psychology is a special science, as cognitive science also appears to be. (In chapter eight I will more adequately cover the topic of the special sciences as it pertains to a debate between reductionism and antireductionism. My concern here is to situate how Fodor views folk psychology in relation to natural kinds and the so-called special sciences.) Fodor’s basic attitude is that cognitive science probably cannot do better than what a robust folk psychology can give us (for example, his previously mentioned “representational theory of mind”). So on the one hand he is pessimistic about the prospects for cognitive science because he thinks that it has generally looked for natural kinds in a manner that imitates the way the non-special sciences look for natural kinds—recall that the “condition for successful science (in physics, by the way, as well as psychology) [emphases mine] is that nature should have joints to carve at: relatively simple subsystems which can be artificially isolated and which behave, in isolation, in something like the way that they behave in situ” (1983, p. 128). Looking at these “encapsulated” systems and then trying to piece them together is not the way to get at holistic central processes—processes which are fundamentally unencapsulated. But on the other hand, if we respect that psychology and cognitive science are special sciences, the implication flowing from this is that we ought not to expect similar success using the above method. Rather we should start by acknowledging the holistic nature of folk states, and then build a picture of the mind from there. This picture suggests searching for natural kinds as well,
although the expectation is that things will be "fuzzier." The upshot is that while folk psychology has some claim to projectibility since it "just works," such projectibility is more along the lines of functional regularity of the sort that is found predominantly in the special sciences.\footnote{It is important to note that it would be incorrect to claim that folk psychology is just a matter of "functional" regularities—the former is "thicker" than the latter.}

Projectibility as found in the non-special sciences is much tighter, but this type of projectibility is inappropriate for assessing folk states. Thus Fodor's differing attitudes. Charitably speaking, he isn't waffling; he is respecting the differing natures of different kinds of sciences while at the same time wanting to preserve a core intuition (i.e., about kinds) that still makes them sciences.

1.3: Summary

The above discussion might seem a bit out of place, since this first chapter is supposed to be about the question of whether there can be a science of emotion. In fact, the way in which I've sketched the general debate is important for understanding a variety of approaches to emotions research programs. I am serious when I call this project "Framing Emotion"—the "frames" that I intend to explore are meta-level conceptual frameworks that guide actual research programs. And it turns out that the "splitters-groupers" continuum (read in the light of the previous discussion of holism) helps to frame many of these programs and their assumptions. More importantly, the theoretical question not normally addressed by particular research programs that investigate certain emotions is whether there can be a science of emotion. This question is primarily a "philosophical" one, at least in the sense that it is a very abstract question exploring the possible "bounds of (scientific) experience." The question as raised in Griffiths' work is not substantially different from the question of whether cognitive science can truly be a science. I hope that the general "homology" between these two questions has been made evident by the pictures given in
the two previous sections. In particular, the homology centers around the employment of similar notions—the primary notions being natural kinds and projectibility.

There are slight differences, to be sure. Griffiths’ notion of natural kinds gets a soft-realist reading (recall, an as-if attitude toward the reality of kinds as the best explanation for why an account works): the relevant categories utilize projectibility through the mechanism of causal homeostasis. Fodor doesn’t comment on such metaphysical issues; rather he focuses on projectibility and its relation to holism. In particular, there is a tension between his appeal to natural kinds as used in the non-special sciences, and the “kinds” employed in the special sciences (although recall that there is partial co-extension between propositional attitudes, which express folk states, and natural kinds). As a result, it seems to me that Fodor places greater emphasis on projectibility-related issues from the implicit standpoint of how the special sciences operate. (Informally: if cognitive science wants to be a science by imitating physics, then it won’t succeed—folk states aren’t projectible in that sense.) By contrast, Griffiths’ employment of causal homeostasis emphasizes the background theory of a science that then brings into relief issues relating to projectibility. (Informally: look at the background theories of various emotions research programs—most of these programs aren’t really scientific because they offer too many disputable kinds.) These pictures are generally consistent with one another, but the differences lie in what gets emphasized.

There is a deep implicit assumption in the two arguments that has to do with the sort of subject matters under investigation. Both the holism appeal and the lack-of-causal-homeostasis appeal are hinged to the idea that higher-order processes are underdetermined by the various theories that attempt to account for them. The implicit assumption is that such subject matters are heterogeneous. There are at least two senses of this term (which I discuss in the beginning of Chapter 2); the one that I will focus on throughout the next chapter is just the commonsense
notion that sciences like psychology deal with very diverse subject matters (i.e., human cognition is intuitively thought to be complex and varied). Because Griffiths and Fodor also appeal to a certain view of how science operates in generating their negative conclusions, in order for us to explore the possibility of a science of emotion, we must first explore the issue of thinking about sciences that deal with heterogeneity. The next chapter does precisely this in providing the framing tools needed to understand science from the perspective of heterogeneity.
Chapter 2: Science and Heterogeneity

Recapitulation

Chapter 1 examined the reasons why there probably cannot be a science of emotion, nor a genuine cognitive science. These conclusions were premised on the claims that science searches for natural kinds, and that evidence for natural kinds stems from categories exhibiting projectibility. Since apparently emotion and mind do not refer to kinds, each is not a single thing by which one could claim to establish a “science of” emotion (or mind).

Prospectus

“Emotion” is a term of art which research programs use in a variety of ways, not to mention the numerous philosophical senses of the term. However one understands “the” category of emotion, it is clear that intuitively it exemplifies “heterogeneity.” In the previous chapter the implicit appeal to the heterogeneity of emotions and central systems generated the negative conclusions that there cannot be a science of emotion, nor can there be a genuine cognitive science. In a related context, there are debates in the philosophy of mind and in cognitive science over the relative autonomy of the “special sciences.”¹ The indication is that a fair number of special-science researchers subscribe to such autonomy, where part of the basis for this judgment lies in appealing to intuitions about “heterogeneity” which dismiss any reduction of such sciences to the harder sciences. In particular, there are two intuitions about heterogeneity that need distinguishing: the diversity of subject matters within the special sciences (heterogeneity₁), and the diversity of physical implementers (heterogeneity₂) arguably implementing a higher-order

¹ It is not clear exactly what range of sciences counts as “special” for Fodor (1974, 1997), although the general divide seems to be between the physical sciences and the non-physical sciences.
property of a special science. I focus on the former sense in this chapter; in Chapter 8 I discuss the latter sense.

The plan in this chapter is to explore first the issue of autonomy as it relates to biology. In the subsequent section I argue that questions regarding autonomy can be organized via a conceptual space which takes into account certain key "parameters." This section deals with meta-level parameters that apply primarily to issues dealing with scientific theories. Section 2.3 then takes a similarly higher-level perspective regarding the synchronic and diachronic aspects of scientific practices and how these practices change. Both sections 2.2 and 2.3 intend to provide a conceptual map for organizing questions about the autonomy of various sciences from the standpoint of heterogeneity. The overarching reason for doing this is to provide a framework that addresses the previous chapter's view of science; specifically, the framework accommodates the rather restricted conception of science that Griffiths and Fodor offer. I argue that we need to account for more parameters to properly frame scientific concerns—the required range of tools for investigating heterogeneous subject matters cannot be restricted just to projectibility. The basic idea is that projectibility should be "widened" to accommodate these new tools. In conformity with this goal, I extract a new picture of what "ought to count as science"—which includes a space for debates over conceptions of just what constitutes "science"—from an examination of actual scientific concerns and practices (again, relative to the standpoint of heterogeneity). Sections 2.3 and 2.4 explicate this new picture by way of outlining a four-dimensional conceptual space for thinking about science.

Biology and biology-related concerns occupy center-stage in this chapter. The reason for this is twofold. Biology is on the border between the hard and soft sciences and is also a paradigm for thinking about heterogeneity—indeed, biology is often informally characterized as the science of exceptions. And secondly, Griffiths draws important parallels between the
methodological (and theoretical) architectures of biology and cognitive psychology. I explicitly discuss this parallel in Chapter 7, so for now I provide only a "foretaste" of this parallel, via a diagram. The diagram indicates why we need a brief explication of the structure of the biological sciences. (More accurately, since nothing in biology makes sense except in the light of evolution—a claim I exposit in section 2.1 and Chapter 7—I discuss the received view of evolution as it relates to biology generally, and to ecology in particular.)

The following (modified) diagram from Griffiths outlines the parallel (1997, p.221):

<table>
<thead>
<tr>
<th>Biological Explanations</th>
<th>(Cognitive) Psychological Explanations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population Dynamic Level</strong> (e.g., a statistical model mapping relative fitness functions for traits)</td>
<td></td>
</tr>
<tr>
<td><strong>General Ecological Level</strong> (e.g., a model mapping particular sources for a trait)</td>
<td><strong>Ecological Level</strong> (e.g., the role of the psychological “trait”)</td>
</tr>
<tr>
<td><strong>Natural Historical Level</strong> (e.g., a particular historical narrative for that trait)</td>
<td><strong>Computational Level</strong> (e.g., the black-box processes accounting for that trait)</td>
</tr>
<tr>
<td><strong>Anatomical Level</strong> (e.g., physiological and anatomical “mechanics” of that trait)</td>
<td><strong>Implementational Level</strong> (e.g., the brain regions involved)</td>
</tr>
</tbody>
</table>

Figure 1. The (Modified) Parallel Between Biology and Cognitive Psychology

Biological explanations and cognitive psychological explanations "parallel" each other through the levels they aim at, and through the methodological orientation each level adopts. There are two dissimilar things to notice in the diagram. First, the top level is missing from the right side. Unlike evolutionary biology—which has a "unified" (or more appropriately, a "consilient") framework as the top level partly expresses—cognitive psychology is to a greater degree a ragbag of techniques and levels of investigation. Because of this lack, further support could be adduced for the conclusion that there can be no science of emotion, at least with respect to those emotions research programs that fall under cognitive psychology. The second thing to notice is that at the
lowest levels, the lines don’t match up, for while the anatomical level and the implementational level overlap in some respects, there are ways in which the anatomical level in biology mixes the computational and implementational levels. (This last point usually falls under the rubric of “proximate mechanisms/causations.”) As mentioned, I will explain both the above diagram and its implications in greater detail in Chapter 7. What needs to be put in place in this chapter is an overview of biology and biology-related concerns as they pertain to autonomy.

2.1: Autonomy and Biology

Heterogeneity and Ecology

A rich area where considerations about heterogeneous systems arise in relation to evolution is ecology. This is not surprising given that ecology must deal with numerous interactions between “biotic” and “abiotic” factors. The intertwining of the manifest complexity of ecological systems and the different types of questions generating diverse research strategies poses the following problem: what overview should we give of the diverse strategies? The overview that Gregory Cooper proposes, in *The Science of the Struggle for Existence*, involves the broad assumption that “any satisfactory explication of the disciplinary structure of the background knowledge [which generates a research strategy], relative to some particular field of science [ecology in particular], will have the consequence that some kinds of considerations are much ‘closer to home’ than others” (p.118). Cooper continues: “The basic idea is this. We can think of a scientific field as being defined, in part, by a set of conceptual resources and a set of fundamental commitments” (*ibid*). The *conceptual resources* are the field’s “natural kinds,” and the *fundamental commitments* are the shared, usually implicit beliefs characteristic of a practice which determine the salience of certain research strategies over others. These fundamental commitments strive towards “nomic generalizations” (see fn.2 below), using the field’s existing
natural kinds. The ideal aim is then to find "genuine" natural kinds that form an invariant partitioning over a reference class. The general view that Cooper gives is in essence no different from the earlier remarks on kinds, with the exception that kinds now have a statistical interpretation with respect to homogeneous classes: "The pursuit of this ideal involves locating the phenomenon to be explained in an ever-narrower reference class until, in the limit of this process, one reaches a reference class that is homogeneous for the attribute in question. The reference class is homogeneous when no further partitioning in terms of statistically relevant factors is possible" (p.105).2 At this point the temptation may arise to reduce ecology to physics, since ultimately, homogeneous kinds are just physico-chemical processes "in the limit." While Cooper doesn't draw the implication that a reductive view is appropriate to ecology (indeed, he recognizes that such an image is not appropriate3 given his exposure to contextualized ecological issues), two questions get raised: (Q1) what would prevent someone from taking that next reductive step? and (Q2) are nomic generalizations appropriate to ecology?

The Qualified Autonomy of Biology

Ernst Mayr indirectly provides an answer to both questions, insofar as ecology is a part of evolutionary biology, broadly construed.4 According to Mayr (1997), three questions structure the biological sciences: what? how? and why? What-questions appeal to descriptive projects that usually involve comparisons (e.g., between taxa in issues concerning systematics). How-questions ask how a process works, invoking proximate mechanisms to explain how, say, a

---

2 In slightly different, but more precise, terms, homogeneity can be defined according to a resiliency function: given a sentence q, a set of sentences S={p_{i}}, and Pr(q)=n, Res(Pr(q)=n) = 1-Max_{l<s<Pr(q/p_{i})} for p, ranging over S. The best-case scenario is when q is independent from S; the more that q depends on elements of S, the lower the resiliency. The degree of resiliency then determines the degree of "nomic generality." See Cooper (2003), p.114. Note that resiliency takes the ("widened") place of what narrower work projectibility does for Griffiths and Fodor.

3 Cooper (2003), p.249.

4 The extent to which ecology is a part of the general evolutionary framework is a controversial issue, as Mayr recognizes. But to the extent that biology isn't merely physics cum chemistry, Mayr's insights are crucial in understanding what makes biology a different kind of science.
heart develops and works to maintain homeostasis. Lastly, why-questions appeal to "ultimate causations" for why a certain trait might have evolved. These three questions guide a proper understanding of what it means to talk about the framework of evolution. It is within this framework that Mayr defends his claim that biology is autonomous from the other sciences, and cannot be reduced to the laws of physics and chemistry. There are three senses of "reduction" that Mayr discusses (1988, pp.8-24): constitutive reduction, explanatory reduction, and theory reduction. The first holds that all organic processes are ultimately physico-chemical processes, or perhaps more weakly, are consistent with such processes. Mayr claims that biologists in general have no qualms with this since appeals to proximate mechanisms and answers to certain what-questions may involve such lower levels. It is the latter two reductions that are in dispute. So the claim for autonomy is not based on an "ontological" appeal, but rather an explanatory and methodological-theoretical appeal. Explanatory reduction fails for the simple reason that emergence prevents lower-level explanations from having the proper kind of traction required for adequate biological explanation. And theoretical reduction fails just because past attempts at such reduction aren't practically feasible—when the resources normally used by biologists are replaced by lower-level theories and laws, the latter produce virtually no results of interest.

In answer to Q1 above, Mayr appeals to empirical complexity: "What counts in the study of a complex system is its organization. Descending to a lower level of analysis often decreases the explanatory power. ...Actually in the course of downward analysis invariably a level is reached sooner or later where the whole meaning of the system is destroyed when the analysis is carried down any further" (2004, pp.72-3). Explanatory reduction fails because the salience of

---

5 I present a fuller picture of evolution in Chapter 7. Note that the above three questions do not imply a "reduction" of various biological fields to the particular field of evolutionary biology; rather most fields operate in semi-autonomous fashion (see section 2.3, "Biology and Research Programs"). The primary sense in which "nothing in biology makes sense except in the light of evolution" concerns the general evolutionary process by which various fields, operating at differing levels of investigation, attempt to further understand aspects of this process.
the system—the locus of explanatory relevance—gets lost in the details of the reduction. Mayr continues: "It is now abundantly clear that evolutionary emergence is an empirical phenomenon without any metaphysical foundations. Acceptance of this principle is important because it helps to explain phenomena that previously had seemed to be in conflict with a mechanistic explanation of the evolutionary process" (p.77). For example, emergent properties of water are not due to any intrinsic metaphysical foundations, but rather are due to new, multiple interactions among water molecules (p.76). Seeing how such interactions are organized requires a certain level of analysis, especially so when dealing with the complex interactions that biological systems manifest.

**Autonomy and Nomic Generalizations**

What about Q2? The answer to this links to answering in further detail how theoretical reduction fails. The first point is that insofar as there are "laws" in biology, they don't operate like the laws at lower levels. "Generalizations in modern biology tend to be statistical and probabilistic and often have numerous exceptions...these generalizations have so limited an application that the use of the word law, in the sense of the laws of physics, is questionable" (1988, p.19). Secondly, there is a plurality of theories in biology, making it difficult to see how any feasible reduction would take place. Appeals to in-principle reduction also miss the mark, since finding appropriate translations between theories (e.g., "bridge laws"—see Chapter 8) places the burden directly on the reductionist to show just how such in-principle reduction would work:

Going through the glossaries of books in various branches of biology one encounters hundreds if not thousands of such untranslatable biological terms. Examples are territory, speciation, female choice, founder principle, imprinting, parental investment, meiosis, competition, courtship, and struggle for existence, to give only a few examples. ...It is only in the biology of proximate causations that theory reduction is occasionally feasible. On the other hand, no principle of historical evolutionary theory can ever be reduced to the laws of physics or chemistry. Contrary to the claims of some reductionists, this has

---

6 For example, Mayr argues that Darwin's evolutionary theory is really a whole package of theories, five of which require special attention (2004, Ch.9).
nothing to do with any alleged immaturity of biology. Indeed, the new insights gained by molecular genetics during the last forty years have made the impossibility of reduction even clearer than it was before (2004, pp.78-9).

Even if a revised picture were given of how such translatability could be effected, Mayr's point is that this view of how science is ideally supposed to function does little regulatory work for biologists—a different picture is needed.

Are there, then, no genuine laws in biology? Earlier I made the caveat that Mayr provides an indirect answer to this question insofar as ecology is a part of evolutionary biology. Just how far is that? Again, Mayr provides a succinct answer: "Ecology differs from most other biological disciplines in that it does not squarely fit into the biology of either proximate or evolutionary causations. Moreover, parts of ecology, like evolutionary ecology, are dominated by an intricate synergism of proximate and ultimate causations." (1997, p.225). So it seems that biology is semi-autonomous in that there is permeability with respect to some proximate causations and with respect to certain ecological investigations; but biology is still autonomous insofar as it is a historical discipline (that searches for ultimate causations) with conceptual resources that do not effectively admit of explanatory and theoretical reduction. And if biology (as well as ecology) has not "laws," but more appropriately certain kinds of nomic generalizations, they are probably located in regions similar to those involving "intricate synergism." What requires further exploration is the range of kinds of nomic generalizations and how such a range relates to questions about autonomy.

7 See Kitcher (1984).
8 For example, Schaffner (1993), Ch.9. A particular type of reductionism will be discussed in section 8.1.
9 Indeed, this is part of Mayr's motivation for writing Toward a New Philosophy of Biology.
2.2: A Conceptual Space for Questions of Autonomy

Assuming that nomic generalizations occur in the light of background assumptions/fundamental commitments which, according to Cooper, are difficult to tease out (p.119), the issue arises: what sorts of considerations help us to think about how fundamental commitments are generated? Even if we cannot explicitly formulate what those commitments are, it would still be of service to provide a higher-level analysis of how such commitments are structured.

The reason why talk of "nomic generalizations" is appropriate has to do with the fact that concepts—even in evolutionary biology—are implicitly general to some extent. The degree to which such generalizations are stable (measured for Cooper by the degree of "resiliency"—see fn.2) determines how nomic those generalizations are. (Note that this suggests that there are generalizations which aren't appropriately classified as nomic.)

Cooper (Ch.7, 8) provides a conceptual space for mapping the different types of generalizations that one ought to countenance in portraying a broad view of science. The space is organized around three "orthogonal" axes: concrete-abstract, causal-phenomenological, and low-high fidelity.

The Causal-Phenomenological Axis

To explain this scheme let us start with the two-dimensional case, with the "x-axis" as the causal-phenomenological range and the "y-axis" as the concrete-abstract range. For the former range, causal generalizations are capturable by a homogeneous partitioning of the statistically relevant factors for an attribute (the ideal being maximal resiliency).\textsuperscript{11} Phenomenological

\textsuperscript{10} For example, narratives in evolutionary theory. See Chapter 7.

\textsuperscript{11} To add to the picture involving resiliency, here is a more formal presentation (Pearl 2000): Define a causal model as an ordered pair \(<M, O_M>\), where M is a causal structure—not merely a correlation set—relative to \(O_M\) (which is a set of parameters compatible with M). M is defined as a Bayesian network with transition probabilities; just think of it as a (messy) graph with conditional probability transitions between each parent-node and child-node. The elements of \(O_M\) are structural functions that describe the transition "morphism" defined by \(x_r=f(pa_r, u_r)\) for parent function \(p\), and for a random independent disturbance \(u_r).
generalizations, by contrast, track correlations between events (e.g., the correlation between weather and a barometer). The difference between causal and phenomenological generalizations is that one can have the latter without the former—the familiar claim that correlation is not necessarily cause.\textsuperscript{12}

The Concrete-Abstract Axis

For the y-axis range, concrete generalizations tend to be localized and specific in contrast to abstract generalizations.\textsuperscript{13} In order to understand this range, first let us focus on the concrete pole of this dimension. A distinction is made between “active invariant counterfactuals” and “less active” invariant counterfactuals, which serves to explicate the sense of “concreteness” that Cooper has in mind. Cooper uses “counterfactual talk” to cash-out intuitions about various (usually causal) regularities—i.e., “under certain conditions, such-and-such events \textit{would} follow,” where such talk intends to capture the broad logical structure of a regularity. Accordingly, the active invariant counterfactuals are active in that they are accepted as \textit{causally} ranging over particular domains. The context in which this occurs is “the contemporary epistemic community that the current science represents. The nomic judgments of this community can be thought of as a function that associates, for each generalization the community accepts and for each potential domain of resiliency, a degree of resiliency for that generalization over that domain” (Cooper, p.116). The “intended domain” (of resiliency) is the hoped for range of the generalization; the

\textsuperscript{12} Another way to express the difference is that phenomenological generalizations usually occur within a single level of analysis, whereas causal generalizations move between levels (especially with respect to finding \textit{underlying} causes).

\textsuperscript{13} Note that (concrete) phenomenological generalizations would only be “localized and specific” with respect to saving the phenomena.
"actual domain" is what is empirically found. So active invariant counterfactuals are measures of how good a generalization is, causally speaking, relative to the actual domain. On the other hand, the "less active" invariant counterfactuals—"less active" since they aren't really doing any "ontological work," although they are also viewed as expressing empirical regularities—are the phenomenological generalizations. Both causal and phenomenological generalizations are concrete in that they express invariant counterfactuals with respect to some actual domain.14

Now let us focus on the abstract pole of this dimension. One apparent virtue of a scientific explanation is having a causal account, one that gives the concrete "mechanisms" of how something works; another arguable virtue is unification (and systematization)—bringing together a wide range of cases under a few explanatory principles. Since unification often abstracts away from the sorts of details that a causal account seems to demand,15 there is a tension between these two virtues. Another way to express this tension is to observe that causal accounts require an appeal to homogeneous reference classes, and unification accounts range over heterogeneous cases.16 To acknowledge some of the demands of the two virtues, Cooper, using Nancy Cartwright's Nature's Capacities and their Measurement, draws a distinction, respectively, between "capacities" and "tendencies"—two foci in the abstract range. The difference between them is that capacities are "about causal structure and [tendencies] are about

14 The discussion of counterfactuals looks rather abstruse since Cooper, I think, wants to avoid committing to any particular interpretation of counterfactuals (e.g., a possible-worlds analysis). This "minimalism" remains consistent with his bare employment of resiliency to cash-out the broad features of nomic force.
15 For example, even in Newtonian mechanics, the "unified" family of equations governing motion—while expressing a causal basis—requires that exact solutions to particular equations be given in order to qualify as a genuine "causal account." These exact solutions are difficult to come by relative to the overall solution space of the unified family, so the point still remains that unification often abstracts away from the details that a causal account seems to require.
16 Note that this tension is sometimes expressed by a distinction between predictive goals that tend to be associated with causal accounts (though not necessarily), and explanatory goals associated with unification accounts. So the sense in which both are still explanatory has to do with what kind of explanation is taken to be a proper explanation—predictive-causal? or systematic-unificational? But since prediction can be unhinged from causal accounts (e.g., Daniel Dennett's Intentional stance as a predictive strategy), a distinction ought to be made between predictive goals and explanatory goals.
how the phenomena behave. Where capacities are the ontological ground for causal laws, tendencies are the ontological ground for laws of association" (p.228).\textsuperscript{17} The way Cooper informally interprets talk of capacities and tendencies is by conceptualizing them as “referring to nomically strong generalizations with vague \textit{ceteris paribus} clauses” (p.230), where \textit{ceteris paribus} clauses function to make adjustments between the actual domain of generalization and the intended domain.\textsuperscript{18} Here is the 2-D picture capturing the discussion thus far (Cooper, Figure 7.1, p.231):

\begin{figure}[h]
\begin{center}
\begin{tikzpicture}
\draw[->] (0,0) -- (6,0) node[midway, below] {Causal};
\draw[->] (0,-1) -- (6,-1) node[midway, below] {Phenomenological};
\draw[->] (0,0) -- (0,-1) node[midway, left] {Concrete};
\draw[->] (6,0) -- (6,-1) node[midway, left] {Abstract};
\draw[->] (1,0) -- (1,-1) node[midway, left] {causal laws};
\draw[->] (2,0) -- (2,-1) node[midway, left] {capacities};
\draw[->] (3,0) -- (3,-1) node[midway, left] {tendencies};
\draw[->] (4,0) -- (4,-1) node[midway, left] {laws of association};
\end{tikzpicture}
\end{center}
\caption{The Causal-Phenomenological Axis and the Concrete-Abstract Axis}
\end{figure}

\textbf{Capacities and Tendencies}

But what does it mean to say that capacities and tendencies refer to strong nomic generalizations with vague \textit{ceteris paribus} clauses? Are we smuggling in the remnants of an inappropriate reductive view from the philosophy of science? Cooper suggests that the “right metric for abstraction is disciplinary uncertainty about the interpretive principles needed to render the abstraction complete” (p.244). It may seem that interpretive principles are just promissory notes expressing more sophisticated “translation principles” (to reduce, say, biological terms to physical terms) that are to be cashed-in at some later (perhaps ideal) time. This would not be a wholly unjustified charge, as part of Cooper’s project is to create a space for the two previously

\textsuperscript{17} Note that by claiming that tendencies are the ontological ground for phenomenological “laws,” the laws of physics don’t lie insofar as they are phenomenological laws.

\textsuperscript{18} Thus as a nomically strong generalization, a tendency is intended to capture some of the demands of unification concerning a range of phenomena.
mentioned explanatory virtues. However, the above diagram of Cooper's conceptual space for general knowledge claims is more nuanced than that. Significantly, he writes:

What is fundamental, from the standpoint of nomic force [degree of invariance], are the local and restricted generalizations that describe the operation of actual situations. The midlevel laws are grounded in facts about these more concrete nomological accounts; in effect, they represent second-order claims about what sorts of concrete nomological descriptions we might expect to find. The tendencies and capacities, in a similar way, express features of the midlevel laws. Put in epistemological terms, both the phenomenal and the nomological content of abstract dispositional claims [tendencies or capacities] is given by the midlevel laws that can be inferred from the dispositions, given the interpretive principles accessible to the discipline. In that sense, the abstract dispositional claims are grounded in the midlevel laws. The same holds for the relationship between the midlevel laws and the concrete nomological generalizations. I leave it an open question whether this most concrete level of laws ever actually bottoms out, or whether the list of factors to be included is open-ended, even in principle [emphases mine] (pp.250-1).

"Midlevel laws" serve to constrain the possible range of application of "dispositional" claims; the "interpretational principles" are either tacitly or explicitly contained in each discipline; and while "concrete generalizations" provide the initial constraints on midlevel laws, Cooper leaves it open as to whether it is appropriate to speak of ceteris paribus clauses and translation principles that would effect the conditions under which one level is related to another. Thus far what

19 Cooper (Ch.6) does make amendments to the causalist view represented by Salmon, and the unificationist view represented by Kitcher.
20 Observe that tendencies and capacities are epistemically parasitic on midlevel generalizations, which in turn depend on concrete generalizations. But while this is accurate from the standpoint of the quest for nomic force types and from the standpoint of how ceteris paribus clauses function, it is distinct from the claim to ontological grounding. The latter holds that judgments about estimations of nomic force and using ceteris paribus clauses tacitly presuppose that (the category of) Nature requires the concept of "power" as the condition for revealing certain abilities under appropriate conditions (see Cartwright 1989, 1999). Cooper employs both senses—"grounded in" as epistemological dependence (abstract dispositional claims are grounded in midlevel laws), and "ontological ground" as "transcendental" condition (capacities are the ontological ground for causal laws). Note also that Cooper's use of "dispositions" covers talk of tendencies and capacities without explicitly appealing to modal concerns (usually dispositions are associated with necessity and capacities with possibility); rather Cooper favors, again, talk of nomic force and invariance.
21 Clearly reduction is not the issue for Cooper, as both causal and phenomenological generalizations are concrete. Also, midlevel laws are midlevel relative to concrete generalizations, which means that physical scale is not the way in which to conceptualize levels; rather the relation is between concrete and abstract, which has to do with degrees of heterogeneity and homogeneity that can be exhibited at "small" as well as "large" scales.
Cooper's conceptual space affords us a potential way to organize disciplinary claims to "autonomy," at least relative to the degree to which different kinds of generalizations can be made. It ought to be noted, though, that the notion of interpretational principles still remains black-boxed. Before addressing this point, a last axis dealing with degrees of fidelity needs to be added to the picture.

**The Fidelity Axis**

The two-dimensional conceptual space, with the third fidelity axis, intends to show how theories can function as tools, which recognizes that theoretical models come in many types and operate in different ways depending on the variegated demands of practicing scientists. What this third axis acknowledges is that some representations "lie" more than others, in that the initial conditions and assumptions of some representations are more artificial than others; fidelity tracks the degree to which models explicate relevant theoretical ideas. Why is this a third independent axis? It would seem that high/low fidelity maps on to the concrete/abstract axis. But while it is somewhat accurate to claim that concrete representations are also of high fidelity (or at least are intended to be, especially with respect to acknowledged phenomena within a single level of analysis), it would not be correct to claim that (abstract) dispositions in general are always of low fidelity, even though the way in which dispositions are expressed abstract away from the details that would appear to be required for claims to high fidelity. What would be missing are the aims and purposes to which representations are used. If a high level theory aims at capturing certain general processes, and the assumptions and initial conditions aren't properly viewed as merely artificial but rather as the right sort of conceptual starting point, then in such cases dispositions
have high fidelity. Hence this third axis is "independent" of the other two, as it tracks the capacity in which models intend to explicate general theoretical ideas and intuitions.

Implications of the Three-Dimensional Space

So with this three-dimensional space for differing types of generalizations, what implications follow? For Cooper, who sees theories as tools, the conceptual space can help to disclose when theoretical debates are at "cross-purposes. Debates that are ostensibly about the explanatory and/or predictive success of a theoretical treatment can actually be about the appropriateness of a particular modeling strategy, and debates about whether a particular theoretical approach shows promise are often encumbered by misguided expectations about what would constitute success for that theoretical treatment" (p.265). Additionally, locating a theory within some region of the conceptual space helps to see where a theory might over-extend its bounds: a theory could intend to be causal, yet at the abstract level of capturing dispositions, that theory would be better classified as capturing a tendency; it could also happen that a theory makes claims for the existence of a disposition that doesn’t in fact exist (e.g. a false correlation or a false causal link); or it could happen that a theory which holds for a concrete range of cases might not hold over a more general range. These implications, as disclosed by the three-dimensional conceptual space, assist Cooper in understanding ecology. That is, Cooper uses this general framework to understand ecology as a science; in particular, the framework intends to

22 A putative example is (a dispositional approach to) evolution, which I will cover in Ch.7. I give another example of a high-fidelity disposition involving rainbows in Chapter 8.

23 "Theories" can range from highly localized models to abstract "models" that normally are classified as theories. Thus the midlevel laws discussed previously play the role of models as schematic intermediaries between particular empirical application and higher-level theoretical concepts. Cooper recognizes the distinction between models and theories: "Theoretical ideas about abstract tendencies and capacities seek straightforward descriptions of these dispositions. Models seek to represent this same descriptive content, but indirectly. ...Because the phenomenal content of abstract dispositions is typically difficult to pick out on its own, we build alternative systems where the consequences of the abstract dispositions might be played out in a more perspicuous way. These alternative systems are models" (p.255). By emphasizing theories-as-tools, Cooper is focusing on the "continuum" between models and theories.

24 For further discussion of these points, see pp.265-6.
apply to a broad conception of ecology as an "interfield" theory (pp.282-3), as a borderland account between evolutionary population biology and ecosystems ecology (or "biogeochemistry"). And although this three-dimensional space applies specifically to ecology as a borderland science, upon reflection its import clearly is far more general.

Lastly, with respect to background assumptions and interpretational principles, this conceptual space also assists in bringing to the fore the assumptions and principles being used relative to a particular context of inquiry. It does so by focusing on what types of generalizations are being aimed at, in addition to any explanatory or predictive goals that might be at stake. However, this space as it stands has the following potential difficulty in disentangling issues relating to autonomy. Suppose that two inquiries are carried out in disciplines thought to be autonomous. Suppose further that the three-dimensional framework locates the types of generalizations for each inquiry in the region concerning, say, tendencies, and let us assume that the particular tendencies for both disciplines are “the same.” Is this mapping plus explanatory or predictive goals enough to distinguish how these disciplines might be autonomous? It is here that the issue raised at the beginning of this section must be pursued in further depth, namely, how the black-boxed fundamental commitments are structured. For Cooper’s conceptual space assumes (correctly) that such commitments are matters internal to each practice; but while the framework can help to articulate some of the background assumptions at work by virtue of what generalizations are aimed at, it is not immediately clear how the trio of explanatory goals, predictive goals, and the conceptual space further assists us with the above difficulty. In the next section I will argue for the claim that three additional concepts must be kept track of that will help to articulate some of the features regarding how fundamental commitments are structured.

23 Recall that for Cooper, a scientific field is defined, in part, by a set of conceptual resources and a set of fundamental commitments. His interfield approach occurs relative to fields that are “autonomous enough” to enable a fruitful exploration of these connections.
Keeping in mind that the overall project is to explore the possibility of a science of emotion, Cooper’s framework must be widened to accommodate the shifting borders of various sciences and their potentially shifting claims to autonomy.

2.3: Paradigms, Research Programs, and Philosophical Frameworks

A major reason why we need a higher-level analysis of background assumptions and interpretational principles has to do with the ways in which studies of science generate different pictures of scientific change, relative to the type of science investigated. Thomas Kuhn, for example, focuses primarily on physics, and his emphasis on problem solving has a distinctive applied-mathematical flavor. Others (such as David Hull) focus on biology and emphasize an account of scientific change (especially with respect to internal shifts in theory) from an “evolutionary” perspective. And then there is the sociological query as to whether certain (e.g., social) sciences are really sciences, the question perhaps motivated by the charge that change is just the substitution of one pet theory with another. I suggest that a host of confusions can be avoided by distinguishing (at a minimum) between paradigms, research programs, and philosophical frameworks, which nevertheless form a “continuum”—a fourth axis to add to our conceptual space. I will extract the tools associated with each of these three notions from the works of Thomas Kuhn, Imre Lakatos, and Larry Laudan.

Incommensurability is often thought to be the crux of the differences between Kuhn, Lakatos, and Laudan. I think this is in part correct. I will defend the position that while theoretical incommensurability is very important, there are the associated concepts of progress, rationality, and logic\textsuperscript{26}; it is the differing characterizations of these concepts and their interrelations (with each other and with incommensurability) that are responsible for much of the

\textsuperscript{26} Each of the three authors takes some sense of “logic” for granted, but they do not explicitly cash-out what that means.
dispute. Closely connected with this is another muddle that I expose—namely the status of wholes. In the process of unfolding these issues, I hope to make it clear that paradigms, research programs, and philosophical frameworks form a “continuum” for keeping track of background commitments, interpretational principles, and disciplinary change.

**Kuhn and Paradigms**

Kuhn distinguishes between Kuhn₁—as he sees himself—and Kuhn₂—as he thinks others see him. But even with a charitable reading neither of the two definitively exists. Roughly there is the Kuhn of *The Structure of Scientific Revolutions*, of *The Essential Tension*, and of (his reflective remarks in) *Criticism and the Growth of Knowledge*. Here is perhaps yet another.

In order to understand science we need to look at research practices to grasp the structure of paradigms as they function historically. Kuhn sees himself as investigating the “psychology of research”: the history of a disciplinary matrix involves a period of normal science punctuated by revolution and its replacement by another matrix, and what requires examination are the ways in which scientists behave and the functions those behaviors serve. The expectation is that “in the absence of an alternate mode that would serve similar functions, scientists should behave essentially as they do if their concern is to improve scientific knowledge” (1970, p.237). During periods of normal science, scientists engage primarily in puzzle-solving. The sense of “paradigm” as exemplar is the primary vehicle for training prospective scientists; the exemplars are recognized achievements providing model problems and solutions for a community of practitioners (1962, p.x). Kuhn distinguishes between exemplar and disciplinary matrix, where the latter requires a community of practitioners embedded in a “form of life”—a form including ordered elements such as symbolic generalizations, models, values, heuristics, etc. These are the
shared global commitments,\(^{27}\) while the exemplars are localized problems that are usually standardized (e.g., textbook solutions of a set of differential equations that model projectile motion) but can also be recognized problems in need of further articulation (e.g., different interpretations of the Schrödinger equation as cashed-out in branches of physics such as solid state physics and field theory (1977, p.307 fn.17)). Thus puzzle-solving has the dimension of “filling the interstices” of, and expanding upon, the discipline’s framework—including the procurement of new modes of behavior to satisfy certain functions (for instance, inventing new experimental techniques).

Kuhn critiques the idea that progress is linear, directed accumulation; yet he (sometimes) endorses local, paradigm-relative “progress” that 1) occurs through improvements in the ability to better account for outstanding previous anomalies, and 2) must be able to preserve a previous body of problem solving techniques to some decent extent (1962, p.169). The former depends on the testimony of a proven elite (i.e., professional authority), and the latter is a promissory note, a task to be fulfilled. Kuhn is a realist of sorts, since a paradigm\(^{28}\) incorporates a way of posing questions to “nature” (p.103) that redound to the efficacy of the practice.\(^{29}\) However, this does not mean that progress is either linear or hierarchical, since sometimes standards neither rise nor decline, but are simply different (pp.108-9). The sort of progress that Kuhn has in mind is, I suspect, “pragmatic”—that is, progress is judged relative to the estimated (and fallible) puzzle-solving abilities of paradigms. But how does this occur? For while there might be techniques, data, etc. in common between paradigms, there still is theoretical incommensurability between disclosed worldviews. Kuhn claims that shifts cannot stem from the “logical” structure of

\(^{27}\) While in a sense global commitments are also an aspect of philosophical frameworks, what distinguishes in part a disciplinary matrix from a philosophical framework is the situated nature of such a matrix—i.e., the recognition of what exemplars are contained in the matrix, what values are shared, etc. The distinction is not of kind but of degree.

\(^{28}\) I use “paradigm” as including both disciplinary matrix and exemplar.

\(^{29}\) See also 1970, pp.234-5. The issue for Kuhn is over finding more apt languages for “handling” nature.
scientific knowledge (p.95). And theoretical incommensurability is not merely logical incompatibility, but includes the stronger element of a gestalt shift.\textsuperscript{30} Paradigms provide "maps" of nature (ways of using representations) that only reveal so much, and when they clash, the debate over the significance of standards and values within practices in part is what is at stake (pp.109-10). But these practices in turn "reflect nature" via the view they offer since it is "hard to make nature fit a paradigm" (p.135)—i.e., we must beat things into line with tools like renormalization, artificial initial and boundary conditions, etc.\textsuperscript{31} It is when we grasp various structures that the new view comes into focus, where the shift is abrupt and qualitatively different. The leap, as Kuhn puts it, is akin to a religious conversion. Thus theoretical incommensurability between worldviews differs not just in techniques, data, etc., nor in praxis (since there can be not insignificant overlap between two paradigms) but crucially in an "emergent organizational property." Is the shift, like a religious conversion, an irrational leap of faith? Kuhn at times (1962) suggests that it is like this, but commentators have "mistakenly" focused on the wrong aspects of the analogy (Kuhn\textsubscript{2} is "mistaken" for Kuhn\textsubscript{1}); and in addition the focus on conceptual schemes, non-translatability theses, and languages has been consequential, but off the mark, I suspect, with respect to Kuhn’s concerns (at least relative to his later self-reflective remarks).

Kuhn (1977) claims that theory transition is not merely arbitrary, and must satisfy at least five criteria: 1) accuracy; 2) consistency (internally and with other accepted theories); 3) broad scope; 4) simplicity; and 5) fruitfulness (pp.321-2). Is theory transition then rational? Not quite, since Kuhn spends a great deal of time arguing that "rationality" as narrowly construed by various logical positivists (and their kin) is not what he has in mind, for there are always

\textsuperscript{30} Apparently the reason why "logic" isn’t enough is that, by the Quine-Duhem thesis, a new phenomenon or a new theory might accommodate a new scope, or carve out a new niche (see p.95 for examples).

\textsuperscript{31} These structures don’t literally "reflect"; to borrow from Hacking, we represent by intervening. Our representations are the product of—and serve to guide—our interventions.
judgments at the individual and institutional levels that infuse values into any conception of what is rational. Tentatively, Kuhn submits that "rational" progress might be judged by looking at groups that are large enough in well-established scientific communities and, despite statistical fluctuations due to individual differences, on the whole the "mean" choice should count prima facie as rational (p.333). If, however, the group is too small, then judging progress becomes more problematic using this method. Whatever Kuhn might mean by "rationality" it probably cannot be fully cashed out, but it must involve at the very least sketches of individual appraisal (e.g., 1962, p.44 on tacit knowing) and institutional habits that in some sense succeed.

Part of the upshot of Kuhn's focus on the historicity of science and the context of practice is that progress as teleological is ruled out; rather there is a modest bootstrapping process that moves "forward" by moving away from the past (1962, pp.170-1), in the sense that we judge progress by relatively better puzzle-solving ability. What isn't clear is how much forward we go by moving away, since not only does knowledge get lost in transitions, but additionally there seems to be an inverse relation to the pursuit of knowledge, namely the more we find out, the faster the horizons expand. As Kuhn ruminates, "are the interstices between these points of attachment [for current scientific theories] perhaps now larger and more numerous than ever before" (1977, p.290)? Neither naive continuity nor naive discontinuity can be attributed to Kuhn, since he emphasizes that there is within practices an "essential tension"—in a sense, that which gives birth also sows the seeds of its own demise. Aperetically, Kuhn observes that while

---

32 Presumably Kuhn has in mind the natural sciences, and maybe certain social sciences. Most certainly he can't be thinking of modern mathematics.
33 Note that even Kuhn's "faith" (1962, pp.157-8) is a leap based on skill and aesthetic choice.
34 Yet Kuhn intriguingly ends his 1962 work with the metaphysical question, what must the world be like in order for science, and progress, to be possible?
scientists value unity, the future is left open as to whether in the face of the above inverse relation scientists might change their values (p.289). 35

An Overview of Kuhn

Thus far we have progress as “moving-away,” rationality as institutional and dependent upon professional (scientific) authority (see 1962, p.167-8), and incommensurability as a gestalt shift. These concepts are closely intertwined: young scientists train in a paradigm and must pass through the “exemplar-phase” set up by a workable tradition (assuming normal science is already in place). They develop the skills to become embodied certifications of their institution’s values; they push the field until there is enough cognitive weight provided by repeatedly failed attempts to crack anomalies (assuming they arise); certain members entertain and potentially develop other frameworks at the fringes of the paradigm; usually the young guns develop the new theory that eventually becomes the successor, and the break marks a new emergent whole such that the theory transition, while not arbitrary, is still an incommensurable leap. 36 Yet at lower levels (techniques, worked problems, etc.) within the transition there is some overlap between old and new paradigms. How much reconstitution there is and what this might mean is an open question that depends on particular historical cases, but if we are to make sense of Kuhn’s evolutionary

35 Kuhn (1970, p.264) claims that his view of scientific development is “evolutionary”—he uses a tree of descent metaphor where science is unidirectional and irreversible. What is left unclear in the metaphor is whether science is evolutionary with respect to “common descent”—in which case theories in some sense build upon one another, though without a global telos—or with respect to progress as “busby” (the interstices between points of attachment for paradigms?), which leaves open the question of what is meant by unidirectional.

36 Note that this reconstruction of Kuhn starts in the middle—i.e., what is not emphasized are the cases when modern science gets started and (preparadigmatically) there are many programs that might not fit Kuhn’s criteria (at least 5, recall) for theory transition. Perhaps in this case there cannot be theory transition since preparadigmatically we have pre-theories. However, a potential problem is that Aristotelian scientia—which, according to Kuhn, is scientific—presumably had theories, and it may be disputable that the five criteria apply. This issue would depend on judging theory change from the standpoint of the new science versus focusing on the transition between (“worthy”) theories. (Arguably, this is where a transition in philosophical frameworks is involved, which includes a change in the conception of the kind of knowledge sought by “science,” the reasons for seeking such knowledge, and the methods appropriate for seeking it.)
view of progress as moving-away, some kind of fudging needs to be given as to what the
mechanics are that make revolutions a "discontinuous continuity." The two notions that Kuhn
and his critics (as well as sympathizers) implicitly use are "logic"—sometimes used
coextensively with "rationality," sometimes used more narrowly (e.g., 1962, p.95), sometimes in
relation to Popper and others (1977, pp.283-92), and sometimes in the colloquial sense of offering
arguments (1970, p.261)—and organizational wholes. In the responses to Kuhn in Criticism and
the Growth of Knowledge it seems to me that most of the dispute is generated by different
sensibilities about these two notions. The differing intuitions over these two things form, as it
were, the canvas upon which rationality, progress, and incommensurability are painted. Such a
claim is nowhere more apparent than in Lakatos's "Falsification and the Methodology of
Scientific Research Programmes."

Lakatos and Research Programs

Lakatos accuses Kuhn of irrationalism and opts for rational progress through his
sophisticated methodological falsificationism, whose slogan is that we should endorse a
proliferation of theories. This is a pluralism where theories progress by a process in which 1)
theory T* has excess empirical content over T (i.e., T* can predict novel facts not easily
accessible by T), 2) T* captures and explains T, and 3) some of the new excess is corroborated
(1970, p.116). Refutation of a theory is retrospective, since there is no crucial experiment that
rejects a theory except when judged by hindsight after a better theory has been put into place. As
an analogue to Kuhn's normal science, Lakatos proposes a methodology of research programs
that appraises a succession of theories as scientific or not through "negative" and "positive
heuristics" (i.e., those paths not to pursue and those to pursue). The metaphor Lakatos uses is that
of a hard core and a protective belt. Negative heuristics tell us not to upset the core tenets of a

37 Kuhn's sense of logic ought to be Deweyan, as far as I can tell—namely logic construed widely as
inquiry. For example, see his psychology as research (1977, pp.308-19). See also 1962, p.196.
theory, whereas the protective belt can be toyed with using auxiliary hypotheses, should apparently anomalous results make their appearance. The positive heuristics suggest to us how to refine the protective belt by modifying (mostly mathematical) models that better simulate reality (p.135).

According to Lakatos, Kuhn's normal science is a monopoly that prevents other theories from having their say. But Kuhn might respond to this charge by claiming that Lakatos hasn't appreciated his distinction between paradigmatic and preparadigmatic science. Plurality is compatible with the preparadigmatic stage. Moreover, Kuhn's progress as "moving-away" appears to be compatible with Lakatos's scheme. So what is the dispute, besides some mudslinging over "irrationality"? Lakatos's counter is that Kuhn's psychology of research "describes" the sociology of scientific change (the old guard dies, young guns brandish their new theory), whereas his methodology is normative—there is the active determination on the part of rational agents to cut away programs that aren't working over some "reasonable" span of time. For Kuhn this makes no sense since the descriptive-prescriptive divide is a non-starter (p.237); his description of functional behavior that works is already prescriptive. Where's the difference?

The major difference comes at the end of the essay where the growth of knowledge occurs within a "Platonic"/"Hegelian" scaffolding.\(^{36}\) First, growth occurs "in the world of articulated knowledge which is independent of knowing subjects" (p.180). And second, Lakatos's rationality is a reification of progress that "dialectically" unfolds itself within an independent scaffolding—what appears like the process of competing research programs is really the underlying expression of the third world of "propositions, truth, [and] standards" (p.180, fn.1). Whether or not we buy into this picture is an issue that I'll place on the side. There is

\(^{36}\)Lakatos somewhat skews Popper's three worlds in claiming that the third world is the world of objective knowledge, which is not quite that of minds (p.180, fn.1).
another important difference that is not quite captured by the claim that plurality is (merely) compatible with the preparadigmatic stage. The history of biology proves to be instructive.

**Biology and Research Programs**

Biology's history appears far patchier than Kuhn's distinctions between preparadigm, paradigm, and revolution seem to suggest. In discussions of several examples of major and minor "revolutions" in the history of biology, Mayr (2004) cautiously concludes (p.168) that (1) while there are both large and small upheavals, even when large ones occur they aren't quite sudden, and earlier and later paradigms (not necessarily incommensurable) can coexist for some time; (2) there don't seem to be any periods of normal science (even though it is convenient to speak of "paradigms"), but rather for active areas of biology there appears to be a series of minor revolutions between major ones; (3) the model of evolutionary epistemology better fits theory change in biology, where there is a "steady proposal of new conjectures (Darwinian variation) and some of them are more successful than others. One can say that these are 'selected' until replaced by better ones"; and (4) a "prevailing paradigm is likely to be more strongly affected by a new concept than by a new discovery."39 To the extent that we need a conceptual placeholder for something like "paradigms," I suggest that "research programs" is a better (if still not wholly adequate) candidate for the job.

In accordance with (1) and (3), we reject research programs (whether competing or perhaps complementary) only on hindsight by virtue of their comparative fruits, and so they can coexist (in semi-autonomous fashion) and to an extent can tolerate small upheavals. For (2), biology is not driven by puzzle-solving and the emphasis on exemplars to the same degree as physics seems to be. In part due to the heterogeneity of its subject matter, and the difficulty of finding standardized representations that allow for efficacious and projectible methods of

---

39 For a full discussion of the historical examples, see Mayr (2004), Ch.9.
intervening, it is hard to find “normal” modes of widespread operation in biology.⁴⁰ And for (4), the notion that concepts play a greater role in comparison to new discoveries has to do with how one views the tangled web of biological inquiry—often a new way of viewing things proves to have greater impact than any new discoveries per se. Such a tangled web fits better with the tenor of research programs.

Although the use of research programs is more helpful than Kuhn’s distinctions (preparadigm, paradigm, revolution) as applied to theory change in biology, the concepts that such usage affords depend heavily on the issues raised by Kuhn (i.e., the overall similarity of Lakatos to Kuhn that the former implicitly exploits). To further clarify just how close Lakatos is to Kuhn, and where the subtle differences lie, the following defenses of Kuhn should be of service.

**Appropriating Research Programs**

First, Lakatos caricatures Kuhn by claiming that gestalt shifts are a bandwagon effect, or more infamously, a “mob psychology.” But recall that Kuhn appeals to skilled authorities. Also, even older members of a paradigm can “get” a new framework but cannot persuade themselves to believe it. Secondly, there is a conflation between “logic” and “rationality” when Lakatos states that “in Kuhn’s view there can be no logic but only psychology of discovery” (p.178). The logic that Kuhn criticizes is a narrow, roughly positivist conception that is not to be equated with rationality. Lakatos’s claim that there is no rational cause for a crisis (p.178) is based on this conflation and a standard of rationality that is overly strong. These points highlight that Lakatos is still beholden to the spirit of positivism (and perhaps “Hegelianism”), whereas Kuhn emphasizes the historical, sociological, and pragmatic dimensions of scientific practice. A proper

⁴⁰ There are model organisms like *C. elegans* and *Drosophila melanogaster*, which have a “puzzle-solving feel” to them, but the function of model organisms differs from exemplars in physics, primarily due to the restricted scopes of generalizations made in biology (and how they are interpreted); see section 2.1.
amendment to the notion of research programs ought to side with Kuhn on this matter (as I do in my *appropriation* of the general conceptual features of research programs).

Lastly, while Lakatos largely avoids the issue of incommensurability, he claims “we can *make*” incommensurable theories—“by a dictionary [i.e., a “translation” manual allowing for the possibility of comparison]”—“inconsistent and their content comparable [i.e., comparable only as the translations are seen to be inconsistent with *one another*]” (p.179 fn.1). Dictionaries have nothing to do with it; the issue is over the status of emergent wholes and whether such new levels are breaks, and to what extent the breaks are “broken.” Lakatos’s dictionary begs the question by presuming some standard of translatable by which comparison we can (in principle) always effect. *A fortiori*, this misses the mark since the organization is structural, which requires more than merely a focus on language-related issues.

**Laudan’s Reticulation Model**

The issue of wholes becomes more salient in Laudan’s critique of Kuhn. According to Laudan, Kuhn commits the “covariance fallacy” by confusing the relations between the three levels of methods, aims, and theories in science (1977, p.43). For example, Laudan claims that for Kuhn a disagreement about facts is also to some extent a disagreement about aims, since relative to organizational wholes, facts and aims mutually inform one another such that disagreement at one level reverberates to other (hierarchical) levels.41 In place of Kuhn’s hierarchical model, Laudan proposes his reticulation model where aims, methods, and theories have integral relations to one another, but are also independent enough to behave in certain functional ways (p.62-3). Change is not an “all or nothing affair” but rather usually occurs piecemeal—i.e., in “unitraditional change” some debate might occur at the theory level; existing

---

41 According to Laudan, the Kuhnian hierarchy has, from top to bottom, aims, theories, and methods. Experimental results can be challenged, but the last to give way are the “core beliefs” embodied in aims and values. Such aims and values are a key aspect to philosophical frameworks.
methods indicate that a new theory could be more warranted; scientists try the new theory, and after time pressure arises to amend the old methods with new ones (e.g., new computer techniques using simulation models); this pressure results in debate over aims and current theory, and how they mesh with old methods; practitioners adopt the new methods because they mesh better with existing theory, but in turn this reveals that the old aims need to be revised (e.g., cutting-edge computer simulations can be legitimate forms of experimentation/proof, with caveats) (see pp.76-7). This gradualist picture of scientific change doesn’t exclude revolutions, since sometimes there can be radical shifts where theory, aims, and methods all change together in a substantially different way, but Laudan makes the point that such change is rare (p.86). Change has more variety than Kuhn seems to realize: it can be piecemeal, it can involve shifts of more than one piece at a time, it can reverberate with varying degrees of strength through the disciplinary network, and so on. The point is that continuity is entwined with degrees of discontinuity, and the project is to figure out the mechanisms and dynamics at various levels of operation to see where we ought to give some form of discontinuity greater weight, or whether seeing certain pieces as continuous is a more fruitful way of understanding change. Rather than the blanket category of shifts as incommensurabilities, there are wholes within wholes—multiple “variables” and their interacting relations—and we ought to understand shifts as grades within this “discontinuous continuity.” For some shifts are inclines, other steps, and on occasion there are leaps.

Appropriating the Reticulational Model

Again, the status of wholes lies in the background of this debate. Intimately tied to the gradualist view are conceptions of progress and rationality. Progress is much akin to Kuhn’s moving-away view. As Laudan asks, in judging progress does a displacement yield a more optimal situation relative to our “whiggish” aims (p.65-6)? We should not gauge progress against
some objectivist conception of reality (see Laudan 1977, Ch. 5, for his critique of realism), but not just any aims will do either. Rather progress is moving away for communities of professionals, and in addition aims and their satisfaction must be pragmatic. As Laudan states, “the engine driving axiological change is grounded in a theory of rationality, acting to overcome a state of disequilibrium” (p. 55). While Laudan doesn’t specifically mention Dewey’s logic-as-inquiry, consistent with much of Laudan’s work is the idea that rationality and logic are construed widely as the labor of settling problematic situations. The working through of such situations within the institutional framework of practicing scientists allows for the disclosure of further aims relative to degrees of satisfaction of current aims. It would appear, then, that Laudan’s main difference with Kuhn lies not so much in his conception of logic and rationality, but rather in his reticulational model of wholes that looks for the further dimensions of structure within the many ways of construing discontinuous continuity. (It is important to observe that my pragmatic appropriation of Laudan leaves open the extent to which (pragmatic) logic includes various sociological dimensions.)

**Philosophical Frameworks**

What the reticulational model adds to the picture thus far is a way to keep track of paradigms and research programs by tracking the degree of integrity between methods, theories, and aims. Paradigms have a greater degree of integrity compared to research programs, which helps to partially explain why biology, as the science of exceptions, tolerates many (semi-autonomous) research programs (again, “preparadigm” suggests the wrong picture). To infer that biology isn’t a proper science would be unsafe. There is a different kind of integrity that is revealed from the standpoint of a philosophical framework. So what are philosophical

---

42 Deweyan, in particular. See p. 40.
43 See Burks (1994) for a sympathetic reconstruction of Dewey’s logic.
44 Note that this includes estimations of the satisfiability of future aims in relation to displaced ones.
frameworks? As a first approximation, a philosophical framework expresses the worldview that a paradigm discloses; the global commitments of a disciplinary matrix; the hard core of a research program; or the core beliefs that the aims and values of a science embody. But neither paradigms nor research programs are equivalent to philosophical frameworks. How then are they to be distinguished?

The view I offer is that philosophical frameworks are orientations or "stances" that bring into relief the different kinds of integrity that can hold between methods, theories, and aims. At one end of the spectrum where integrity is high, a paradigm offers a worldview that yields a well-substantiated philosophical framework. With a weaker degree of integrity, research programs disclose a vaguer picture (or pictures) if only because the various consequences of the methods, theories, and aims are not as well worked out in comparison to paradigms, perhaps due to the heterogeneity of the subject matters being investigated. Lastly, when integrity is weakest, philosophical frameworks might offer mere worldviews without any mention of methods and theories, ideals that tap into some existing senses of methods and theories, or perhaps a new concept that will help to reorganize existing methods/theories or to create new methods/theories. These are just a few possibilities. Importantly, the parameter of philosophical frameworks brings into relief disputes that may be over the very nature of the integrity between methods, theories, and aims. For example, when integrity is called into question concerning whether certain social sciences are really sciences, philosophical frameworks open a space for genuine

45 The previously mentioned fourth point made by Mayr highlights how concepts are of greater importance in biology. The salience of philosophical frameworks as "stances" is crucial to note, since biology's "different kind of integrity" is revealed, in part, by the stance taken toward biology's use of concepts. (Recall section 2.1; see also the discussion of narratives in Chapter 7.)

46 Another example is the role of speculative metaphysics in modern theoretical physics. Some argue that such speculation (e.g., string theory) is not scientific and is just "philosophical," while others argue that the role of speculation is more profitably viewed as a part of competing research programs.
dispute over differing conceptions of "science"—philosophical frameworks assist in disclosing where the very category of "science" is problematized (cf. fn.36).

The upshot is that philosophical frameworks, research programs, and paradigms form a "continuum" which is organized by questions relating to wholes and the sorts of integrity that frame those questions. As it turns out, Cooper's black-boxing of fundamental commitments and interpretational principles is not unproblematic. What I hope to have achieved, in line with Cooper's conceptual space, is the further articulation of the tools needed to account for not just the types of generalizations that sciences can make, but also the additional ways in which "sciences" can be distinguished. The fourth axis is intended to assist in that respect.

2.4: Summary

This chapter provides a pluralistic picture of science, as it provides a greater range of tools to accommodate a greater range of sciences (even those "sciences" whose claim to being scientific is disputed). The pluralistic picture also encompasses the restricted view of science given in Chapter 1; it does so through the general (statistical) tool of resiliency—when resiliency is high, what results is a partitioning of X into homogeneous reference classes, which can then be interpreted as referring to natural kinds. However, we ought to consider other scientific virtues in order to portray a less restricted view of "science." Thus by adopting a heterogeneous perspective, the four-dimensional framework purports to situate an indefinite array of research agendas. The framework is largely drawn from Cooper, who utilizes his three-dimensional space to situate differing research agendas (and potential borderland relations between agendas) within ecology; Cooper also uses that space to situate ecology as a highly heterogeneous science. I've extended the framework so as to be potentially applicable to the variegated relations between the natural sciences (specifically the biological sciences) and the "human sciences." Even stronger,
the fourth dimension includes an explicit parameter that creates a space for disputes over the very nature of conceptions of "science"—for example, whether certain research agendas which utilize highly "humanistic" assumptions are truly "scientific." To be clear: my four-dimensional space does not list "criteria" for assessing whether X is a genuine "science"; rather the framework is more liberally oriented. Since 1) research programs have contextualized origins—contextualized by various historical, scientific, sociological, technological, etc. considerations; 2) we assess research programs (relative to the previous sorts of considerations) as "properly scientific" only retrospectively; 3) research programs may employ philosophical frameworks whose claims to being "scientific" are highly contentious; and 4) these philosophical frameworks could issue from research agendas whose past history in purporting to be "scientific" might still be hotly contested, I think it would be imprudent to list criteria for a field's qualification as "scientific" when part of the problematic I am raising concerns pluralistic claims to being "a science of X." The explicit reflective element regarding disputes over the very nature of science is one of the important elements that my fourth dimension adds (specifically, the way in which the philosophical framework parameter bears on research programs, and whatever other tools may be used from the remaining three-dimensional space).

Hence the main upshot of this chapter is the four-dimensional framework situating "science" from a heterogeneous standpoint, as illustrated below:

Figure 3. The Four-Dimensional Framework
2.5: Looking Ahead

To reiterate: our four-dimensional conceptual space (along with any explanatory or predictive goals) for organizing questions relating to autonomy is a meta-level conceptual framework for thinking about science from the standpoint of heterogeneity. This space will help to align the topics in the next four chapters. The following brief overview serves as a guide to chapters three through six.

First, if any research program has become a (putative) paradigm of emotions research (or at least more paradigm-like), it is the psychoevolutionary approach. Griffiths argues (and I agree) that aspects of this approach exhibit causal homeostasis. Specifically, in Chapter 3 I cover Paul Ekman’s work on affect programs, which exhibit causal homeostasis. In Chapter 4 I discuss an “extension” to this approach, namely evolutionary psychology. As both Griffiths and Fodor argue, evolutionary psychology does not exhibit emotion-concepts that are causally homeostatic. Evolutionary psychology moves away from the concrete orientation of the psychoevolutionary approach, where the models that it uses blend phenomenological and causal elements in a “semi-concrete” manner. Evolutionary psychology is also better classified as a research program; what is especially important is its philosophical framework, which I identify as one source of conflict with other research programs.

More generally, the psychoevolutionary approach falls under the “Darwinian Program,” and the extended approach falls under the “Jamesian Program.” In chapters five and six I discuss two other general research programs—the “Cognitivist Program” and the “Social Constructivist Program.” In relation to our fourth axis, their philosophical frameworks are more prominent in comparison to the previous two programs (in terms of the splitters-groupers continuum, these programs are on the groupers’ side). Their models are, strangely enough, causal and
phenomenological, as well as abstract; the dispositions these models express are heavily oriented by their broad philosophical frameworks.

Many of the disputes between research programs occur relative to fidelity-related issues, but such discourses “talk past one another” because fidelity is not organized in light of the other three axes. (Recall the discussion in 2.2 of how Cooper’s conceptual space can assist in pinpointing where theoretical debates are at cross-purposes.) Deploying our four-dimensional framework to untangle the ways in which these discourses misalign with one another is a major task of the next four chapters.

Lastly, I use an additional tool—namely an “informational stance”—to further disentangle these disputes. In particular, I introduce three informational stances organizing the various claims made by research programs in chapters three through six; I give an exposition of, and motivation for, each of these stances in the course of the presentation. (Clarification of the notions of “information” and “stance” is saved for Chapter 9.) Putting Chapter 2 into perspective, the overall picture that chapters two through six are striving towards is this: the three informational stances applied to the four-dimensional framework will issue in three four-dimensional subspace “shapes” that conceptually situate the different emotions research programs. Only with such a picture can we then revisit the question of whether a science of emotion is possible. The picture I shall argue for (sketched in 2-dimensions) is as follows:

Figure 4. The Three Informational Stances and the Four-Dimensional Space
Chapter 3: Affect Programs and Modularity

Recapitulation

The pluralistic picture of science I provided in the previous chapter situates “science” from a heterogeneous perspective. The four-dimensional conceptual space is a meta-level space for organizing various sciences, and their claims to being “scientific.” Biology was discussed as a model heterogeneous science; biology’s interrelations with ecology were then used as a means to introduce the four-dimensional framework. In chapters three through six I apply the framework to four prominent emotions research programs (and their claims to being “scientific”). All four programs employ a variety of biological resources; tracking these resources and their relations the human sciences is a feature of the four chapters.

Prospectus

In this chapter I consider a particular modular approach to emotions, which falls under the broader program called the “Darwinian Program” (the reason for the name is simple: the program can be traced to Charles Darwin). While the modular approach has a general evolutionary orientation, evolutionary considerations operate in the background. Additionally, although there are a variety of research agendas/programs within the Darwinian Program, I focus on a specific set of issues illustrating the general tenor of the Darwinian Program. The particular agenda I shall discuss is Paul Ekman et al.’s work on emotions-as-“affect programs.”

Section 3.1 discusses affect programs. Applying our four-dimensional space, Ekman’s work is situated along the causal-phenomenological axis with an emphasis on concreteness; that is, his research program occupies the “region” outlined by emphasizing the concrete pole and the causal-phenomenological axis. Griffiths’ argument for the causal homeostasis of affect
programs can be construed as the manner in which this research program’s concrete orientation yields empirical results delineating natural kinds. The marks for such kinds are multiple: extensive cross-cultural studies; non-verbal studies; correlations between verbal reports and physiological measurements; consistently distinguished autonomic nervous system signatures; etc. All of these studies employ techniques that are classifiable as concrete, though they differ in the weight assigned to the causal and phenomenological dimensions. Also, while the fidelity of the employed techniques (explicating how affect programs are “pan-cultural”—which I discuss in section 3.1) certainly varies, the multiple techniques above “triangulate” to overall higher-fidelity results bolstering the projectibility of affect programs. If any research program has become a (putative) paradigm of emotions research, this is it.

Section 3.2 organizes the conceptual features of affect programs via a “reverse engineering” perspective. I exposit how affect programs are particular kinds of systems; I also exposit how these systems are analyzed in cognitive psychology—the general field of study consistent with the study of affect programs. Specifically, there are three levels of analysis pertaining to reverse engineering that frame affect programs as particular sorts of (functional) systems.

3.1: Affect Programs

In Chapter 1 it was mentioned that Griffiths argues for the projectibility of (concepts provided by) the psychoevolutionary approach; however this is not wholly correct since a variety of research agendas fall under the psychoevolutionary approach, and not all of their concepts manifest projectibility in Griffiths’ estimation (1997, Ch.3). What remains true is that the view of emotions as affect programs exhibits projectible concepts, and this view does fall under the psychoevolutionary approach—an approach characterizable as the modern extensions of
Darwin’s ideas put forth in *The Expression of Emotions in Man and Animals*. Ekman et al.’s (henceforth just “Ekman” for short) work on affect programs is perhaps the most prominent illustration of that approach, which is why Griffiths appropriately spends a great deal of time focusing on how affect programs exhibit causal homeostasis. In this section I will expound Griffiths’ argument, although the path I take deviates slightly from his presentation. In particular, I first outline Fodor’s notion of modularity, and then apply certain resulting concepts to show that affect programs exhibit causal homeostasis. For if affect programs have the same general features of modules, and if modules refer to natural kinds as Fodor argues they do, then affect programs also make a strong claim to being causally homeostatic.

**Fodorian Modularity**

Recall the view given in section 1.2 of modules as compiled transducers that take inputs and, in factory-like fashion, output certain determinate products. There are actually nine features of modules that Fodor discusses, although this is too much detail for our purposes. The reconstruction of “Fodorian” modularity I give stems from the three general features of modules that Griffiths cites: the relevant modular systems need to be 1) “mandatory”—they fire whether a person wants, say, the fear system to operate (they are involuntary); 2) “opaque,” as we are not aware of how our visual system, for example, processes information that gets passed to central systems; and 3) “informationally encapsulated”—e.g., the visual system cannot access all of the other systems since it is a peripheral system (1997, p.93).

With respect to the first feature, mandatoriness (Fodor 1983, p.52) is closely associated with several other features. One of these features is that input systems are “fast” (p.61); in contrast to processes like problem solving, which require forethought and planning, modular systems take a particular range of inputs and quickly and efficiently process that information into ready-made outputs. A good example of a fast input system is what Joseph LeDoux calls the
"low road" for processing certain emotional stimuli (1996). LeDoux's research on the fear reaction system illustrates some of the structural considerations regarding other emotions. The fear reaction system is actually (schematically) two roads: the low road, which takes certain stimuli straight to the amygdala and then ends in emotional behavior, and the "high road," which travels through more of the cortex before going to the amygdala. The high road is slower than the low road, and is involved with higher-order appraisals (and reappraisals) of stimuli (and also just what those "stimuli" are taken to be). The contrast between the high road and low road highlights that the low road is fast and mandatory—especially when we feel fear "in spite of ourselves" (see Ekman, 1999). But how does the amygdala first sort out stimuli that are subsequently categorized as emotional (since the amygdala is involved in multiple systems)? The answer is that the trajectory—of what gets processed through the low road leading to emotional behavior—qualifies those stimuli as emotional inputs. This expresses two other features associated with mandatoriness: the systems only compute specific inputs (Fodor, p.47), and the systems have shallow outputs (p.86). For the former, the range of stimuli processed directly by the amygdala is restricted to a greater extent compared to what is processed by the high road, and so is "domain specific" by virtue of narrow activation; this also goes hand-in-hand with the shallow scope of the system's resulting behavior.

The picture of a module thus far is that it is a "factory" taking only a specific range of inputs, since the factory activates only with the proper raw materials; once those materials activate the system, it processes the materials relatively quickly; how these peripheral processes operate is typically not known by central or conscious processes—all we know is that the peripheral products are produced, whether we indirectly know this by virtue of the experience of seeing what we see, or feeling fear in spite of ourselves (and so on); in addition to being opaque, these processes are also informationally encapsulated, meaning to the degree that the processes
don't have access to the holistic web of belief, they are systems which can be adequately
individuated; and lastly, the outputs are mandatory and ready-made.

One of Fodor's conditions for successful science (recall) is that "nature should have
joints to carve at: relatively simple subsystems which can be artificially isolated and which
behave, in isolation, in something like the way that they behave in situ." Modules are relatively
simple subsystems because they are "easy-in, easy-out"; they can be artificially isolated to
determine what range of inputs is appropriate, and what corresponding range of outputs gets
generated; and from various lesion studies, ethological studies, and brain-imaging studies, it
appears that modules behave something like the way they do in situ. So modules are projectible.

In order for Griffiths to support his claim that affect programs exhibit causal homeostasis, several
things need to be put in place: first, what exactly are affect programs? This relates to why
"innateness" and "universality" may be improperly associated with affect programs. Secondly,
the claim that affect programs are modular requires that we examine their fit with the general
classification of a module, and this in turn requires a brief examination of the experimental
evidence for such programs. Thirdly, I show that causal homeostasis squares with the modularity
of affect programs.

Characterizing Affect Programs

There is some ambiguity over what affect programs really are. As Griffiths accurately
notes, at times Ekman speaks of programs as literal neural "programs"; at other times the
programs merely have dedicated neural circuits; and still at other times the programs appear to be
confounded with being "innate" or "universal." (I discuss these issues in the following subsection.)
An earlier view put forth by Ekman claims that there are roughly six basic emotions: fear, disgust,
anger, and sadness have the most traction, with joy and surprise having stable, but more modest,
support. It is crucial to observe that the six terms should not be confused with "culturally
informed" notions; while there is *some overlap* with these "thicker" notions, the six terms are really *technical shorthands* for particular affect programs. So what is an *affect program*? Griffiths' broad and charitable characterization is that it is a "coordinated set of changes that constitute the emotional response" (p.77). Of course, not just any coordinated set will do. This characterization is intended both to avoid certain overextended interpretations of the data, and to properly represent Ekman's position. More accurately, the changes are complex, coordinated, and automatic. They are complex since the elements of an affect program include "(a) expressive facial changes, (b) musculoskeletal responses such as flinching and orienting, (c) expressive vocal changes, (d) endocrine system changes and consequent changes in the level of hormones, and (e) autonomic nervous system changes" (p.77); they are coordinated because the above elements occur together in regular, patterned ways that are distinguishable (e.g., the pattern of fear is distinguishable from the pattern of anger); and they are automatic because they unfold without conscious direction.

**Situating a Proper Understanding of Affect Programs**

One thing Griffiths wants to avoid through the above characterization is the talk into which Ekman sometimes incautiously slips when claiming that affect programs are literal neural programs. The weaker reading which Griffiths' characterization affords, and which coheres with Ekman's research, is that while a general neural base is crucial to the set of coordinated changes constituting the emotional response, there may be any number of physiological and developmental processes also undergirding these coordinations. It is better to attribute a conservative reading to Ekman's work since the details of the neural-physiological-developmental processes are still not fully known. In accordance with the conservative reading, a proper understanding of affect programs avoids a host of potential pitfalls. The first point requiring clarification is the (inappropriate) inference that affect programs are "innate" (a term
which Ekman has repeatedly qualified), which even a conservative reading may seem to suggest. Ekman's reasons for arguing against the "innateness" of affect programs are several: critics often confuse his technical use of terms like "fear," "anger," etc. with associated higher-order cultural notions; there are the related confusions that an "innate program" means it is "fixed," "unmodifiable," or it has "no cultural variation." Griffiths' slightly different critique of the claim that affect programs are "innate" is that innateness is "a fundamentally confused concept. It confounds under one term several independent properties. These include the properties of having an evolutionary explanation [e.g., evolutionary explanations appealing to the innateness of a trait as a placeholder for further proximate investigation], being insensitive to variation in 'extrinsic' factors in development [e.g., under "normal" environmental conditions, the phenotypes that regularly manifest themselves in development], being present at birth [like having eyes, ears, arms, etc.], and being, in various senses 'universal'" (p.59). It is important to be clear about how Griffiths' objections mesh with Ekman's concerns so as to avoid improper portrayals of affect programs. First, while Ekman holds that affect programs make sense from an evolutionary perspective, to correct the subsequent inference that the programs are literal neural programs, we should emphasize that affect programs are placeholders for what to expect upon further investigation of the neural-physiological-developmental (proximate) mechanisms involved in these "programs."

Secondly, there is a sense in which affect programs are insensitive to variation in extrinsic factors (in spite of the fact that the actual developmental pathways are not fully known), namely that the coordinated set of changes which constitute a basic emotional response (such as fear) develops reliably across cultures—both literate and preliterate—as well as for infants who are blind from birth. But this does not mean that there is no variation in the elements of these coordinated changes (they are not "fixed")—for example, facial expressions might show more
variance on average in one culture compared to another. As Ekman notes, his view is a very narrow one since only the basic emotions are truly emotions—the restrictive technical sense of “affect programs” yields the only currently viable taxonomy of emotions, and hence affect programs are the only entities deserving to be called “emotions,” “scientifically” speaking (see fn.1 below). Other folk uses of “emotion” are based on what Ekman calls “display rules”—the pluralistic cultural scripts (informed by the human sciences) that allow for folk-emotion terms which are unique to cultures, or which may lead to incommensurabilities in emotive language when translating between cultures. The point is that it is crucial to distinguish the variability of display rules with the constrained variability of affect programs; while there are significant differences between display rules, more importantly affect programs exhibit significant cross-cultural similarities.

Thirdly, affect programs have developmental windows for expression, so being “present at birth” is not the proper way to understand Ekman’s claims. Affect programs aren’t “fixed” in any helpful sense; rather they are developmental regularities that depend upon appropriate environmental affordances to exhibit themselves at the right time. For the same reason it is also misleading to claim that they are “unmodifiable.” And yet in spite of this, from a wider perspective there is a sense in which affect programs are unmodifiable. Because modifiability belongs to the category of display rules, strictly speaking it would be a “category mistake” to claim that affect programs are modifiable—hence are “unmodifiable” as a contrast class—given what they are narrowly about, and given the substantial amount of evidence for their existence.

1 The contrast is between projectible affect programs—and thus are “properly scientific”—and display rules, which, to the extent that they are “scientific,” borrow concepts from social psychology, anthropology, etc.; more generally, display rules borrow concepts from the “human sciences” (see fn.3, Preface). Note that projectibility is not merely a mark of “science” qua physical science, for projectibility also concerns functional natural kinds within the special sciences (see Chapter 1, section 1.2; see also Chapter 8).
and how that evidence bears upon the characteristics of being complex, coordinated, and automatic. 2

This brings us to the last point, that “innateness” is often confused with various senses of “universal.” There are two senses of “universal” that need discussion to properly situate Ekman’s softened notion of “universal” (Griffiths 1997, p.62). The first sense is that a trait may be called “universal” when it occurs without exception in all cultures. Ekman argues not for this claim, but rather for the weaker claim that affect programs are (to use the term that Griffiths cites from Ekman) “pan-cultural”. What is being contested is the “uni” in “universal.” Biologically speaking, an affect program is not one “ready-made” thing; more subtly, each affect program is found with constrained variations across cultures. What the term “pan-cultural” intends to capture is the statistically significant agreement of how, say, facial expressions are interpreted across cultures in spite of all the cultural variation. 3 (Note that this is stronger than just the cross-cultural “similarity” of affect programs.)

The second sense of “universal” is that a particular trait is universal to a species, which is misleading since the biologically proper distinction is between monomorphic traits and polymorphic traits. Both concepts occur relative to populational and statistical considerations. Monomorphic traits are exhibited by (nearly) all individuals within a “normal” environment, whereas polymorphic traits maintain significant (“discrete”) variation of the relevant character

---

2 The “unmodifiability” of affect programs pertains to 1) a statistical, population-level perspective, where affect programs are stably manifested across cultures (they are “pan-cultural”); and 2) the notion that relative to certain “norms of reaction” in environments, affect programs will develop into roughly “the same” things. While particular individuals, say statistical “outliers,” may have (significantly) modified programs, the focus is not on tracking particular developmental changes through time, but rather on “unmodifiability” as a statistical phenomenon, and from the standpoint of examining developmental outcomes (which are on average “the same”).

3 From a broader evolutionary viewpoint, the general idea conveyed by affect programs also makes sense for other species—ethological studies exhibit the homologous nature of affect programs (e.g., LeDoux’s work on fear uses rats as model organisms). Thus taking into account these additional ethological factors, Ekman’s work is better interpreted as establishing “panversal” affect programs and not merely pan-cultural programs.
within a population (for example, eye color—green, brown, etc.). While it isn’t wholly clear that Ekman intends a polymorphic interpretation of affect programs regarding his appeal to evolution, affect programs probably should not be understood as monomorphic because this interpretation’s appeal to “traits” overextends what is warranted by Ekman’s statistical research (cf. fn.13, Chapter 4). The denial of the monomorphic reading is closely related to the above characterization of affect programs as pan-cultural, although there is a slight difference having to do with scope, since the pan-cultural reading is about cultures, while the non-monomorphic interpretation is more broadly about evolutionary considerations (see fn.3 and “panversal”).

The upshot of disentangling the various senses of “innate” and “universal” is that most critiques of Ekman (e.g., see Russell’s 1994 critique, and Ekman’s 1994a response) stem from caricatures of affect programs, where the intuitions underlying the critiques appeal to some sense of “innate” or “universal,” but lack clarity about which sense is being used, and what associated implications are appropriate. Keeping in mind the above distinctions, to further situate affect programs our attention next turns to linking the general characterization of modules with the features of affect programs (as complex, coordinated, and automatic), and the way in which the experimental evidence indicates that the link is plausible. I start first with a brief overview of Ekman’s earlier experimental work and then proceed to show that the more recent experimental techniques and evidence build upon, and complicate, this research program.

An Overview of Ekman’s Experimental Work

Ekman’s earlier work can be divided into four experimental classes (Griffiths 1997, p.50). The first type involves judgment tests that compare literate cultures, where the judgments made by subjects range from 1) self-reports and how they correlate with the relevant facial expressions, 2) “forced choice” tests (similar to multiple choice) requiring subjects to match pictures of facial expressions with the best available words, or 3) the reverse situation where
subjects assign freely chosen labels to the given pictures. (The expectation in this first experimental class is that if affect programs are pan-cultural, there will be statistically significant clustering between word choice and relevant facial expressions.) The second type involves judgment tests that compare literate and preliterate cultures to see if the facial expression pictures correlate with "the same" words. The third type employs component analyses of spontaneous facial expressions; this is nicely illustrated in a particular experiment exposing subjects to a stress-inducing film, whereupon (unknown to the subjects) the spontaneous facial reactions are recorded. The components of a facial expression (mouth position, brow position, etc.) are gauged for the stereotypical expression of fear, for example, and then the expressions across subjects are measured to find the degree to which they correlate with the stereotype. The fourth type of experiment with blind infants is along the same lines, correlating their expressions with stereotypical facial expressions. (Note that the third and fourth types do not invoke linguistic factors.) The import of all of these experiments is that they deliver statistically significant results. What Ekman did was to conduct a long-running series of experiments involving multiple set-ups which triangulate upon a stable phenomenon, namely the existence of basic affect programs. (Recall the characterization of affect programs as complex, and some of the elements involved—(a) through (e). These elements configure in different ways to give rise to coordinated and distinct programs.)

Affect Programs as Fodorian Modules

There are at least seven marks of affect programs that these four types of experiments control for, where according to Ekman the marks are characteristic of the (roughly) six basic emotions. They are: "automatic appraisal, commonalities in antecedent events, presence in other primates, quick onset, brief duration, unbidden occurrence, and distinctive physiology" (Ekman 1994b, p.18). The feature of automatic appraisal requires a lower-order processing of appropriate
stimuli, which then generates the relevant emotional behavior. The characterization of modules as opaque generally fits this condition. As for the second feature, the commonalities in antecedent events, this was expressed in association with mandatoriness, namely the idea that only certain kinds of inputs get processed (the inputs are "domain specific"). The ethological evidence for affect programs in other species (including primates) was alluded to previously (fn.3); it supports the evolution of (homologous) affect programs. The features of quick onset and brief duration fit with the notion that modules are fast. And fastness, recall, closely relates to the claim that modules are mandatory; the feature of unbidden occurrence conforms to the latter.

Lastly, affect programs have distinctive physiological signatures (usually demarcated by correlations between facial expressions and the autonomic nervous system); this final feature squares with the important condition that modules are informationally encapsulated—if the modular view of affect programs is right, the information processed by the autonomic nervous system ought to be distinguishable with respect to each program.

Recent Evidence and Complications

More recent evidence complicates the evolving view of affect programs, although the core idea that they are modular in form remains intact. The last two things I want to cover are the relation between such evidence and causal homeostasis, and how this research program can be oriented within our four-dimensional framework. I will contextualize both of these topics within a contemporary debate between "dimensional" approaches to emotions and "discrete" approaches; the former are toward the groupers' side of emotions research, while the later are toward the splitters' side.

One of the complications recent research raises has to do with the taxonomy of affect programs. Most experimental studies have focused on the "big six" emotions, but current evidence indicates that the list might need to be expanded to include so-called "nonverbal
displays" (e.g., love, sympathy, amusement, and embarrassment are nonverbal insofar as the correlations between terms and facial expressions are unstable), which apparently exhibit "enough" of the modular characteristics of affect programs. There has also been some doubt cast upon the strength of the correlations between facial expressions and verbal reports for the emotions of fear and sadness, perhaps due in the case of fear to the failure to control for the different pathways of the fear response system (the high road and the low road). In general, it turns out that the types of complications current research raises are about extending the scope of affect programs, and also about finding where the bounds of this approach might be (see Davidson et al. 2003).

The earlier research relies mostly on correlational studies at the concrete, phenomenological level, where the degrees of "resiliency" occur within and between studies—that is, the degree of statistical significance within each experimental study, and the degree of consilience between studies. Some of the results lean toward the causal end of the spectrum, especially those that track autonomic nervous system signatures. However these earlier results deal mainly with cruder measures of heart rate, skin conductivity, etc., where the actual causal pathways are not fully known. More recent studies still do not know the full causal pathways (which, incidentally, is generally the case in nearly all of science), but with the advent of new technologies—especially for brain-imaging and lesion research—the emphasis is increasingly placed on mapping less crude causal proximate mechanisms that mesh with the earlier body of phenomenological-statistical research. What the new studies (e.g., see Dolan 2000) strongly suggest is that affect programs, whose features had been delineated at the level of (phenomenological) correlations between inputs, black-boxed processing, and outputs, have underlying causal accounts that are separable into "kinds" (e.g., signature neurophysiological circuits activated for disgust). The modular account of affect programs already makes a strong
bid for projectibility in light of the earlier research; the new emphasis on explicating the neural, physiological, and developmental pathways ("opening up the black-box") currently underway adds further support for the causal projectibility of affect programs, from which it is plausible to claim that they exhibit causal homeostasis.

**Discrete Approaches Versus Dimensional Approaches**

Although current research is extending the scope of affect programs via causal homeostasis, the issue still remains what the bounds are of this discrete approach. The discreteness of affect programs stems primarily from the capacity of modules to be (reliably) distinguished from one another. By contrast, the question dimensional approaches raise is whether we ought to restrict emotions just to those that qualify as affect programs (recall Ekman's claim that only affect programs are genuinely scientific emotions; all else is display rules juxtaposed upon the "canvas" of the basic emotions; see fn.1). Briefly, dimensional approaches hold that there are, for example, two axes that account for emotional phenomena: one valence axis ranging from positive to negative affection (like pleasure and pain), and another intensity axis ranging from low to high intensity. Continuous variation of these two parameters classifies (and perhaps gives rise to) emotions, depending on what developmental and cultural factors guide such variation. Dimensional approaches emphasize the temporal aspects of emotional expression and development, whereas discrete approaches emphasize the synchronous aspects of "emotions" (or more properly, the modular characteristics of "affect programs").

Dimensional analyses also emphasize the role of culture in development. The temporal-developmental-cultural orientation of dimensional theories makes them holistic, where they are phenomenological to a greater extent compared with discrete theories. Additionally, dimensional models are less concrete (or more abstract) than discrete theories, whether they apply to individual accounts of "2-D" emotional processes (the individual's "web of emotions") or to a
communal web of emotion-practices. Because of the wider scope of these models, the dimensional approach generates heterogeneous results and differing accounts of the best means to organize disparate data. While this might be interpreted as support for the discrete approach and concomitant criticism of the dimensional approach, Ekman draws a reconciliatory lesson: discrete theories account for synchronic phenomena with far greater traction than what dimensional theories afford (e.g., "boundary effects" for facial expressions; see Kelner and Ekman 2000), but this occurs at the price of discounting temporal factors—factors that the machinery of the dimensional approach is designed to accommodate.

Unfortunately dimensional theories fall prey to some of the same criticisms that Fodor and Griffiths mount against holistic accounts; the cost of temporality is less projectibility. And regarding discrete theories, although the causal homeostasis of affect programs is the primary reason why Ekman's program is the most viable candidate for being a putative paradigm of emotions research, the tradeoff is the restriction to a rather narrow, technically defined notion of "emotion" that is removed from the manifest messiness of lived emotional experience. Thus it makes sense to try and extend the psychoevolutionary approach—and the virtues of modularity—to higher-order emotions, while also recognizing the messiness of lived experience. Accordingly, in Chapter 4 we turn our attention to evolutionary psychology—an extended modular research program, classifiable as a discrete theory, which still addresses some of the virtues of the dimensional approach. Before turning to evolutionary psychology, I give a higher-level perspective on affect programs in the following section.

---

4 An example of the latter would be anthropological studies that aggregate individual reports taken from a range of cultures, and then map where in the N-dimensional space the various emotion terms cluster. The results may indicate overlapping regions, but more often attention is drawn to the gaps between cultures, "incommensurable" or otherwise.

5 In both senses as an exemplar, and as a "form of life"—the triangulation via multiple techniques (exemplar) to psychological kinds is indicative of consilience, and such consilience coheres with the evolutionary worldview (form of life) by which the paradigm hangs together.
3.2: Reverse Engineering and Understanding Affect Programs as Systems

Since the larger concern is with understanding emotions from "scientific" viewpoints, the emphasis in section 3.1 is on Fodorian modularity; for part of what apparently qualifies Ekman's research program as "properly scientific" has to do with considerations about modularity and kinds.

This section steps back from the coverage of affect programs, and attempts to organize the conceptual features of modularity from a higher-order perspective. Specifically, Fodorian modules are functional modules. I describe a particular sense of "functionalism" involving a top-down method to understanding (synchronic) complex processes, where the process is broken down into parts in order to grasp how those parts "function" in relation to the (functional) whole; this is often called "reverse engineering." In this sense, functionalism is primarily a research strategy used to explain certain phenomena. I then outline what is "functional" about reverse engineering. Lastly, I explain how affect programs are particular kinds of (functional) systems.

Reverse Engineering

A representative example of a top-down research strategy for understanding information processing systems is David Marr's computational, algorithmic, and implementational levels of analysis (1982). The highest level is the computational level, whose features are "(1) that it contains separate arguments about what is computed and why and (2) that the resulting operation is defined uniquely by the constraints it has to satisfy" (p.23). While Marr calls this a "computational" level, I claim we should view it as the level that discloses the overall "architectural role" of the system (cf. Sterelny 1990). Modular programming in software engineering provides a good illustration of the computational level. Usually a complex programming task needs to be broken down into subcomponents/modules, where each
subcomponent carries out a "subcontracted task" quickly and accurately. Each subcomponent thereby performs a clearly defined function—a particular role clearly defined in relation to the overall architectural role—and the aim is to specify as efficiently as possible all the functions required to perform the overall architectural task. The subcomponents are the separate arguments/modules that explicate (through the function each performs) what is computed, and why (i.e., their contribution to the overall task). The resulting operation—the successful solving of the task—is further characterized by the constraints placed upon the task, including available memory, public/private coding, etc. Lastly, the design should be optimal; i.e., the modular programming’s specification of the subcomponents provides the most efficient run-time configuration. As with Fodorian modules, time is of the essence.

The second level is the algorithmic level. Continuing with the modular programming example, the algorithmic level involves actually coding each module in an appropriate language. It is crucial to Marr’s picture that the language be “appropriate,” meaning the programs operate on parameters that are “representational.” In accordance with computation theory, it is presupposed that a program is a set of rules operating on a string of well-defined symbols. (Although “representation” is a term of art, it is worth keeping in mind the idea of rules-operating-on-symbols as a guide through the various uses of “representation” in this section.) Hence as representational, the parameters aren’t merely symbols or strings of symbols; more strongly what counts as a symbol is the capacity in which it serves as an input to a program that can use that information. In sum, at the algorithmic level, the coded programs “carry out” the functions specified by the computational level—these lower-level, “causal role” functions carry out the architectural function(s) specified at the computational level.

The last level is implementational, which is simply the physical level at which the programs are executed. Here the algorithmic level’s programs are translated into the
“causal/physical” functions (or operations) implemented on electronic circuits. All together, the three levels yield, according to Marr, a proper understanding of a system as an information processing system.⁶

Reverse Engineering Applied to Understanding Affect Programs

Affect programs, which exhibit the features of Fodorian modules, are methodologically framed by reverse engineering. While all three levels (the computational, algorithmic, and implementational levels) situate an understanding of affect programs, emphasis is placed on the algorithmic level.⁷ The computational level, which brings into relief the architectural roles of affect programs (e.g., fear as a particular type of survival-based response), is placed in the background; such roles matter, of course, but Ekman’s research program focuses primarily on the modular features of affect programs, and evidence for their being pan-cultural. In other words, the algorithmic level concentrates on the causal-role functions of affect programs—specifically, the three major (pan-cultural) features of affect programs (they are complex, coordinated, and automatic; see Glossary) outline the causal roles of the elements of affect programs. Informally, if affect programs are conceptualized as “boxes” which process specific inputs and produce specific outputs, the computational level inquires into the general roles played by a box; the algorithmic level looks at that box, and attempts to figure out how its parts (causally) operate to process specific inputs and then produce specific outputs; and the implementational level looks at the neuro-physiological mechanisms underpinning how these “parts” work. Investigations at the

⁶ It should be noted that Marr’s view has been criticized as being too bottom-up, since the implementational level constrains the algorithmic level (and what can count as a representation), and in turn the computational level ends up as merely a conceptual convenience to understand the lower levels. My focus on Marr’s research strategy makes no commitment to these ontological issues.

⁷ Recall that Figure 1, given at the beginning of Chapter 2, lists three levels: the ecological (see Glossary), computational, and implementational levels. Somewhat confusingly, the last two levels map to Marr’s algorithmic and implementational levels, respectively; but the ecological level only partially maps to Marr’s computational level. I will discuss these levels, and their differing terminology, in Chapter 4, section 4.2.
implementational level, as noted in section 3.1 (see "Recent Evidence and Complications"), are currently underway.

**Function and Reverse Engineering**

There remains an "unanalyzed" term latent in the computational, algorithmic, and implementational levels of analysis, namely "function." Although I suggest that the question "what is a function?" is inappropriately phrased, since it isn't properly a thing *qua* thing, it is worthwhile to explore the better-phrased question, "what are the conceptual features required for framing 'function'?" As a clean, illustrative (though artificial and technical) case, let us consider the notion of a recursive function.  

An iterative recursive function provides a simple model for delineating some of the general *conceptual* features of functionality. First, we need a set of arguments that serve as inputs to a function. Then a base condition defines the value of a function given a range of arguments; for example, \( f^{(0)}(x) = x \). The iterative condition claims that the value of \( f^{(n)} \) is determined by the previously computed value: \( f^{(n+1)}(x) = f(f^{(n)}(x)) \). An easy case is when \( f \) is the successor function \( s \) and the set of arguments is the natural numbers, whence \( s^{(0)}(1) = 1, s^{(1)}(1) = s(s^{(0)}(1)) = s(1) = 2 \), etc.  

The first conceptual point to notice is that the arguments are inputs that can be individuated from one another. For mathematics precision is the greatest, as objects (and languages in general) are required to be "well-defined." But in less precise circumstances, the lesson remains that inputs ought to be individuable. For example, DNA is often conceptualized as a code that takes information from the chemical environment and digitally "instructs" the process of synthesizing proteins (usually structural proteins). In order for the code to preserve the

---

8 The reason for considering a recursive function is its direct relation to Turing machines; it also has family resemblance relations to other computationalist orientations (some of which only metaphorically utilize Turing machine ideas).
informational integrity of the “messages” contained in the genetic “blueprint,” the code must digitize/individuate the pieces of the message. It is a well-established fact that analog signals cannot transmit information reliably over extended ranges of space and time; if a complex analog signal is an intricate piece of instrumentation P that requires transportation from A to B, P risks damage due to its intricate design and the contingencies of transportation (“noise”). By contrast, digitizing the “signal” breaks P into parts, sends the parts from A to B, and reconstructs P upon arriving at B. This offers a solution to the problem of transmitting complex information while maintaining integrity (i.e., controlling errors and noise).

The second conceptual point to notice is that the inputs only make sense relative to how the inputs get used. A function generates a reliable pattern—it is a transformer of inputs into patterns. (For example, a simple iterative logistic equation takes data and yields patterns for modeling population growth.) With these two conceptual points, the picture thus far is that inputs are things that can be individuated, and inputs are needed to generate patterns. This picture suggests the following three conceptual moments of understanding “function”: individuation is crucial because of the issue of informational integrity; the generation of patterns requires the use of information; and the pattern itself is an informational pattern.

Concerning Fodorian modules, “individuation” pertains to the domain-specific inputs; the “generation of patterns” concerns the manner in which modules quickly process inputs; and the outputted pattern is itself an “informational pattern.” In particular, affect programs—as complex, coordinated, and automatic—are functionally distinguished by the different ways in which they 1) process specific inputs, 2) coordinate their complex constituents, and 3) automatically generate a (stereotypical) affective response. The functionality of Fodorian modules (and affect programs)

9 Note that a far more appropriate distinction is often made between replicator and interactor, rather than between blueprint and message. See Dawkins (1982, 1995), and Brandon (1990, Ch.3).
is brought into relief primarily by the algorithmic level (the computational and implementational levels, as indicated previously, play important but less prominent roles).

Affect Programs as (Functional) Systems

Lastly, affect programs are conceptualized as functional information processing "systems." Systems, broadly speaking, are defined as individuals whose constituents are situated in relation to each other and to the system as a whole. Affect programs are "systems" in that their constituents—which include (a) expressive facial changes, (b) musculoskeletal responses such as flinching and orienting, (c) expressive vocal changes, (d) endocrine system changes and consequent changes in the level of hormones, and (e) autonomic nervous system changes—are situated in relation to each other—the elements are coordinated in regular, patterned ways—and to the system as a whole—the overall architectures of affect programs are distinguishable from one another by virtue of the different ways in which whole coordinated packages operate.

3.3: Summary

* A more analytic (though highly general) characterization of function is given by Robert Cummins (1975), which is compatible with the above remarks. A nice summary of "Cummins functions" is provided by Amundson (1993):

In functional analysis, a scientist intends to explain a capacity of a system by appealing to the capacities of the system's component parts. A novel feature of Cummins's analysis is that capacities are not presented as (necessarily) goals or purposes of the system (i.e., the algorithmic level is emphasized, not the computational level (which concerns itself with the roles/purposes of the system)). Scientists choose capacities which they feel are worthy of functional analysis, and then try to devise accounts of how those capacities arise from interactions among (capacities of) component parts. The functions assigned to each component are thus relativized both to the overall capacity chosen for analysis [the overall behavioral output] and the functional explanation offered by the scientist. Given some functional system s:

\[
X \text{ functions as an } F \text{ in } s \text{ (or: the function of } X \text{ in } s \text{ is to } F) \text{ relative to an analytical account } A \text{ of } s \text{'s capacity to } G \text{ just in case } X \text{ is capable of } F-\text{ing in } s \text{ and } A \text{ appropriately and adequately accounts for } s \text{'s capacity to } G \text{ by, in part, appealing to the capacity of } X \text{ to } F \text{ in } s. \quad (\text{Cummins 1975: 762; variables renamed for consistency})
\]

Cummins's assessments of function do not depend on prior [emphasis mine] discoveries of the purposes or goals served by the analysed capacities [i.e., the analysis carried out at the algorithmic level is largely "synchronic"] (p.232).
In conclusion, reverse engineering brings into relief a “4-D subspace” within our four-dimensional space that captures the conceptual features of affect programs. This subspace emphasizes the causal-phenomenological axis by virtue of its concrete orientation. The fidelity dimension depends primarily on what the reverse-engineered models emphasize: affect programs exhibit a high degree of fidelity since the models triangulate across cultures, across studies, etc. Lastly, affect programs seem to refer to natural kinds; thus Ekman’s program is appropriately categorized as occupying the paradigm-and-research-program region more prominently in comparison to the programs I discuss in the following chapters; for the degree of integration between theories (modeling affect programs), methods (reverse-engineering the features of affect programs), and values (commitment to evolutionary biology) is much stronger than the research programs to come.

Conceptually speaking, this 4-D subspace can be thought of as a 4-D “shape” whose general contour is an arc (“half-oval”) that is best outlined by the concrete-abstract and causal-phenomenological axes:

![Diagram](image)

**Figure 5. Reverse Engineering and the Four-Dimensional Space**

In chapters four through six, I discuss three research programs in relation to this diagram.

Evolutionary psychology is situated (via the “functional stance”) in a (higher arced) half-oval whose orientation is less concrete than Ekman’s program. The Cognitivist Program, in turn, is situated (via the “strategic stance”) in a 4-D subspace above and overlapping with this higher arced half-oval, whose shape is a “crescent.” And the Social Constructivist Program is situated (via the strategic and “semiotic” stances) in a 4-D subspace that heavily overlaps with this 4-D
crescent but also is distinguishable from it. The semiotic stance delineates where the (fuzzy) distinction occurs. The following picture relates this chapter to the next three chapters:

Figure 6. The Three Informational Stances and the Four-Dimensional Space
Chapter 4: Evolutionary Psychology and Extended Modularity

Recapitulation

Affect programs exhibit a modular approach to understanding a restricted, technical sense of “emotions”; higher-order senses appeal to “display rules” (which use concepts from the human sciences). The emphasis was placed on Fodorian modularity, for part of what apparently qualifies Ekman’s research program as “properly scientific” has to do with modularity and its relation to natural kinds (see Chapter 1; Chapter 3, section 3.1; and Chapter 8). Additionally, reverse engineering provided a conceptual understanding of affect programs as functional systems.

Prospectus

I discuss an “extended modular” approach in this chapter, which falls under the broader “Jamesian Program” (the reason for this name is simple: the program can be traced to William James). The particular approach I examine is an “information processing” view of emotions given by evolutionary psychology. In relation to Chapter 3, both Ekman’s program and evolutionary psychology have an evolutionary orientation and employ reverse engineering, although they focus on different types of emotions and different interactive relations classifying such emotions. Specifically, Ekman’s program focuses on “peripheral” affect programs, while evolutionary psychology focuses on more “centralized” emotive-cognitive processes.

There are two further differences between these programs. The first difference is that while Ekman’s modular approach has a general evolutionary orientation, evolutionary considerations operate in the background; by contrast, evolutionary psychology makes explicit use of evolutionary concepts in understanding emotions. The second difference is that (a)
Ekman's program separates affect programs from human-science concepts by virtue of the
distinction between peripheral affect programs—which are really the only foci of Ekman's
program—and display rules—which simply reference the "other category" concerning various
socially informed processes; by contrast, (b) evolutionary psychology's extended modular
approach separates functional "modules" (that are not just peripheral systems, but importantly
are centralized systems) from non-functional (non-adaptive) processes.

Section 4.1 discusses evolutionary psychology's approach to "higher-order" (centralized)
emotions. The approach—compared to Ekman's concrete orientation—is less concrete; and
while evolutionary psychology is committed to a causal "ontology" of emotions as physico-
chemical processes, the emphasis is on a middle-level phenomenological orientation that seeks
functional-emotive kinds. As with Ekman's program, issues concerning fidelity stem from a
reverse-engineering method that delineates the features of emotive processes. Although there are
a few well-worked-out illustrations of this approach, it is best classified as a research program
whose plausibility lies in the fruits it promises to generate, as it claims to be an extension of an
established science, namely (evolutionary) biology.

Section 4.2 organizes the conceptual features of evolutionary psychology through
something called the "functional stance" toward information processing (which, briefly, is an
enriched kind of reverse engineering). Applying our four-dimensional framework, this stance
brings into relief a "4-D subspace" that is situated along the causal-phenomenological axis by
virtue of its degrees of concreteness; the fidelity dimension is more disparate, depending on what
is stressed (for example, certain reverse-engineered phenomenological models have high-fidelity
at the phenomenological level, but don't have high fidelity with respect to causal concerns).
Lastly, the functional stance extrapolates from the best that modular (or modular-like)
perspectives currently have to offer, and as such is appropriately categorized as occupying the
paradigm-and-research-program region more prominently in comparison to the programs I discuss in chapters five and six.

4.1: Evolutionary Psychology and Functionally Individuated Information Processing Systems (FIIPS)

The Jamesian Program, like the Darwinian Program, has a number of research agendas that fall under the “Jamesian banner.” Two rather prominent examples are Antonio Damasio’s work on emotions, which takes its cue from James’s characterization (very roughly speaking) of emotions as feelings issuing from perceptions of (certain) bodily disturbances, and evolutionary psychology’s appropriation of James’s idea that we have diverse modes of behavior because we have more, not fewer, instincts. These ideas are found in James’s The Principles of Psychology (Damasio’s work is a modern extension of the James-Lange theory; evolutionary psychology picks up on different, though related, themes in James’s tome). There are significant overlaps between the two agendas—most prominently, both reject a division between “emotion” and “cognition”—but given that (functional) modularity is the focus of Chapter 3 and Chapter 4, I will discuss evolutionary psychology as the main representative for the Jamesian Program, and how evolutionary psychology extends the modular approach by utilizing the more general concept of information processing systems.

Overview of the Argument

I start with a broad characterization of emotions offered by perhaps the most recognizable school of evolutionary psychology, the “Santa Barbara School,” and then unfold the concepts configuring this characterization of emotions. I discuss a number of potential confusions in order to show what the program is not claiming. I then distinguish several senses of “module” to situate what the research program does claim; closely related to the confusions over modules are
similar confusions regarding encapsulation—which is distinct from modularity. It is relative to these various clarifications that two philosophical critiques of evolutionary psychology receive coverage: first I examine Griffiths’ charge that this school subscribes to a monomorphic view of the mind; then I look at the core of David Bueller’s critique in Adapting Minds. The focus on emotions suffices in illustrating many of the critiques’ broader concerns with central processes. I shall argue that the critiques do not adequately address the considerations brought to the fore by our four-dimensional framework, and as a result the questions that should be asked require alignment along each of the four axes—in other words, the two philosophical critiques, as well as evolutionary psychology’s research program, require alignment.

Characterizing Emotions

John Tooby and Leda Cosmides, two of the main founders of the Santa Barbara School, claim that

an emotion is a mode of operation of the entire cognitive system, caused by programs that structure interactions among different mechanisms so that they function particularly harmoniously when confronting cross-generationally recurrent situations—especially ones in which adaptive errors are so costly that you have to respond appropriately the first time you encounter them.¹

This passage, in essence, captures the evolutionary psychology research program. More generally, evolutionary psychology seeks to map specific types of mind-brain relations from an evolutionary perspective:

The brain is an organ of computation that was built by the evolutionary process. To say that the brain is an organ of computation means that (1) its physical structure embodies a set of programs that process information, and (2) that physical structure is there because it embodies these programs. To say that the brain was built by the evolutionary process means that its functional components—its programs—are there because they solved a particular problem type in the past. In systems designed by natural selection, function determines structure. …There is no single algorithm or computational procedure that can solve every adaptive problem. The human mind is composed of many different programs for the same reason that a carpenter’s toolbox contains many different tools: Different problems require different solutions (Tooby and Cosmides 2000, pp.1167-8).

¹ See http://www.psych.ucsb.edu/research/cep/primer.html.
In what follows I focus on expounding the above characterization of emotions, as it illustrates evolutionary psychology's more general research program.

"An emotion is a mode of operation of the entire cognitive system...."

First, the cognitive system does not separate emotion from reason; it is a system of mental operations ("cognition" in the broadest sense) and the consequences of those operations. Accordingly, the entire cognitive system expresses an indefinite range of species-typical cognitions, where emotion is a particular mode of operation within this system. As a mode of operation, there is no hard division between emotive processes (so-called "hot" processes) and more reasoned, "cold" processes. The notion of a mode emphasizes (as I explain below) that the total cognitive system contains modular-like subsystems (or "information processing systems"), which behave in ways regular enough to distinguish these subsystems; what receives the label "emotion" concerns the relation between the subsystems and associated folk classifications of emotional phenomena.

A Distinction Between Fodorian Modules and Information Processing Systems

However, note that these subsystems are not equivalent to Fodor's notion of modules (what will be called "Fodorian modules"² henceforth). The extended modular approach holds that the mind—both its peripheral and central processes—consists of evolved information processing systems that are "instantiated in the human nervous system" (Tooby and Cosmides 1992, p.24). There are a number of subtle differences between Fodorian modules and information processing systems; the primary difference is that not all of the marks of modules apply to information processing systems—the features of Fodorian modules (see section 3.1) apply selectively to these systems, where weight is differentially placed on certain features

² Although Fodor does not claim that the nine characteristics of modules define what modules are, he has been misread in this way (cf. Fodor 1983, p.37).
depending on the particular information processing system being explored (e.g., a cheater-detection system, a kin-detection system, systems dealing with jealousy, aggression, parental love, friendship, disgust, etc.).

"...caused by programs that structure interactions among different mechanisms..."

To continue with expounding the above passage, these information processing systems are caused by programs, meaning they are produced by canalized, developmental processes, and are adaptive (i.e., products of natural selection). The programs structure interactions among different mechanisms because an adaptive information processing system is a coordinated, functional entity drawing from an array of subelements/submechanisms. However, as Cosmides and Tooby note, while emotions are subsystems (modes) within the entire cognitive architecture, emotions are not equivalently subprograms, since subsystems are the products of developmental programs but are not identical to the programs themselves. The distinction is one of degree, not kind, since emotions are claimed to be superordinate programs; the function of these programs is to

- direct the activities and interactions of the subprograms governing perception; attention; inference; learning; memory; goal choice; motivational priorities; categorization and conceptual frameworks; physiological reactions (e.g., heart rate, endocrine function, immune function, gamete release); reflexes; behavioral decision rules; motor systems; communication processes; energy level and effort allocation; affective coloration of events and stimuli; recalibration of probability estimates, situation assessments, values, and regulatory variables (e.g., self-esteem, estimations of relative formidability, relative value of alternative goal states, efficacy discount rate); and so on. An emotion is not reducible to any one category of effects, such as effects on physiology, behavioral inclinations, cognitive appraisals, or feeling states, because it involves evolved instructions for all of them together, as well as other mechanisms distributed throughout the human mental and physical architecture [emphases mine] (Cosmides and Tooby 2000, p.93).

The list is considerable. But the point of this non-naive account of superordinate "programs" is that programs make evolutionary sense (as I discuss below), since the investigation into levels of proximate mechanisms and their interrelations to ultimate explanations is a complicated
balancing act. The large list of elements that superordinate programs (modally) coordinate is one way in which information processing systems differ from modules, for recall that affect programs are complex, coordinated, and automatic; but in the above case there are far more elements, the elements are coordinated in different ways, and the processing isn't always automatic. Still, evolutionary psychology is trying to extend a “modular-like” account to central processes—processes that tap into peripheral systems, but unlike peripheral systems, have far greater flexibility regarding temporal considerations, the information accessed in the web of belief, and so on. Emotions are the perfect particular topic for illustrating in general why modular peripheral systems make scientific sense, and that a similar approach might work for increasingly centralized processes. Accordingly, I will not focus on accounts of individual emotions, in part because none are “fully” cashed-out by evolutionary psychology, but more importantly because what matters most are the general features exhibited by “emotion” systems—features that should be understood in terms of information processing systems since there is presumed to be no genuine emotion-cognition divide.3

3 Cosmides and Tooby (2000) do use fear to illustrate the some of the (modal) features of these systems, although fear is not “fully” cashed-out; rather they give a sketch of fear as a superordinate program (one can conceptualize fear as a “family” of potential coordinated responses which help the organism negotiate a range of specific complex problematic situations):

Consider the following example. The ancestrally recurrent situation is being alone at night and a situation detector circuit perceives cues that indicate the possible presence of a human or animal predator. The emotion mode is a fear of being stalked. (In this conceptualization of emotion, there might be several distinct emotion modes that are lumped together under the folk category “fear”, but that are computationally and empirically distinguishable by the different constellation of programs each entrains.) When the situation detector signals that one has entered the situation "possible stalking and ambush," the following kinds of mental programs are entrained or modified:

1. There are shifts in perception and attention. You may suddenly hear with far greater clarity sounds that bear on the hypothesis that you are being stalked, but that ordinarily you would not perceive or attend to, such as creaks or rustling. Are the creaks footsteps? Is the rustling caused by something moving stealthily through the bushes? Signal detection thresholds shift: Less evidence is required before you respond as if there were a threat, and more true positives will be perceived at the cost of a higher rate of false alarms.

2. Goals and motivational weightings change. Safety becomes a far higher priority. Other goals and the computational systems that subserve them are deactivated: You are no longer hungry; you cease to think about how to charm a potential mate; practicing a new skill no longer seems
"...so that they function particularly harmoniously when confronting cross-generationally recurrent situations..."

"Programs" blurs the distinction between developmental programs and their products (information processing systems), since the products are themselves superordinate programs. It appears as if the mind is nothing but types of programs. Two things need discussion: what is a proper understanding of a "program," and what is the scope of these programs (i.e., is the mind "nothing but" programs)? Answering the first question requires clarifications about modules and

rewarding. Your planning focus narrows to the present: Worries about yesterday and tomorrow temporarily vanish. Hunger, thirst, and pain are suppressed.
3. Information-gathering programs are redirected: Where is my baby? Where are others who can protect me? Is there somewhere I can go where I can see and hear what is going on better?
4. Conceptual frames shift, with the automatic imposition of categories such as "dangerous" or "safe." Walking a familiar and usually comfortable route may now be mentally tagged as "dangerous." Odd places that you normally would not occupy—a hallway closet, the branches of a tree—suddenly may become salient as instances of the category "safe" or "hiding place."
5. Memory processes are directed to new retrieval tasks: Where was that tree I climbed before? Did my adversary and his friend look at me furtively the last time I saw them?
6. Communication processes change. Depending on the circumstances, decision rules might cause you to emit an alarm cry, or be paralyzed and unable to speak. Your face may automatically assume a species-typical fear expression.
7. Specialized inference systems are activated. Information about a lion's trajectory or eye direction might be fed into systems for inferring whether the lion saw you. If the inference is "yes," then a program automatically infers that the lion knows where you are; if "no," then the lion does not know where you are (the "seeing-is-knowing" circuit identified by Baron-Cohen, 1995, and inactive in individuals with autism). This variable may automatically govern whether you freeze in terror or bolt. Are there cues in the lion's behavior that indicate whether it has eaten recently, and so is unlikely to be predatory in the near future? (Savanna ungulates, such as zebras and wildebeests, commonly make this kind of judgment; Marks, 1987.)
8. Specialized learning systems are activated, as the large literature on fear conditioning indicates (e.g., LeDoux, 1995; Minsky & Crook, 1993; Pitman & Orr, 1995). If the threat is real, and the ambush occurs, the victim may experience an amygdala-mediated recalibration (as in post-traumatic stress disorder) that can last for the remainder of your life (Pitman & Orr, 1995).
9. Physiology changes: Gastric mucosa turns white as blood leaves the digestive tract (another concomitant of motivational priorities changing from feeding to safety); adrenaline spikes; heart rate may go up or down (depending on whether the situation calls for flight or immobility); blood rushes to the periphery, and so on (Cannon, 1929; Tomaka, Blascovich, Kihler, & Ernst, 1997); instructions to the musculature (face, and elsewhere) are sent (Ekman, 1982). Indeed, the nature of the physiological response can depend in detailed ways on the nature of the threat and the best response option (Marks, 1987).
10. Behavioral decision rules are activated. Depending on the nature of the potential threat, different courses of action will be potentiated: hiding, flight, self-defense, or even tonic immobility (the latter is a common response to actual attacks, both in other animals and in humans). Some of these responses may be experienced as automatic or involuntary (pp. 93-4).

(For the above referenced works, see either the article, or go to: http://www.psych.ucab.edu/research/cep/emotion.html.)
encapsulation. These clarifications in turn relate to how information processing systems function particularly harmoniously when confronting cross-generationally recurrent situations—the last piece of the initial characterization of emotions. The second question about scope pertains to interpreting “cross-generationally recurrent situations.” I will discuss a competing evolutionary psychology research program, where I argue that the issue of scope is crucial in uncovering what philosophical assumptions guide the differing research agendas.

Programs and Evolutionary Biology

The notion of a program is never fully cashed-out, as far as I can tell, within the evolutionary psychology literature (cf. Tooby and Cosmides 1990), although its employment is apparently justified by evolutionary psychology’s grounding in evolutionary biology. Accordingly, the following is a brief overview of programs and their relation to the concept of function in biology. Ernst Mayr contextualizes the idea of a program by noting its role in “teleonomic” processes—the mechanical processes that distinguish law-like processes in the life sciences from law-like processes in physics. “A teleonomic process or behavior is one which owes it goal-directedness to the operation of a program” (1988, p.45). Programs for Mayr are materially based and exist prior to the unfolding of the teleonomic process. He makes a distinction between closed and open programs: for example, assuming “normal” developmental conditions, a (hatchling) alligator’s closed program enables it to break its shell and head towards water; and for some bird species their open program allows them to copy the sounds of conspecifics. The task of studying programs and their functions is anything but easy; it is “the

---

4 “Tentatively, program might be defined as coded or prearranged information that controls a process (or behavior) leading it toward a given end” (p.49). This “prearranged information” does not necessarily imply that a program is a “blueprint” (see Chapter 3, fn.9).

5 Note that the distinction between closed and open programs must be soft because of the integral mutualism of program and environment: e.g., a “gene for” the construction of X is really about the complex of genetic/developmental/organismic/ecological pathways—(structural) genes work in a biochemical
most difficult area of biology” (p.50). By practical necessity, programs are epistemically vague. In spite of this, a computational/information-processing view offers a working definition of programs: programs read and process information to serve stereotypical functions; they generally transform energy into order, where the most prominent form of biological order is biological function (adaptations or otherwise—e.g., anatomical or morphological “functions”).

But why are programs important for understanding the fact that evolution occurred, and the implications of this acknowledgement? According to Mayr, “as soon as one accepts the simple conclusion that the totality of the genotype is the result of past selection [broadly construed; see Mayr’s (1988) essay “How to Carry Out the Adaptationist Program?”], and that the phenotype is a product of the genotype (except for the portions of the program that are filled in during the lifetime of the individual), it becomes one’s task to ask about any and every component of the phenotype what its particular functions and selective advantages are” [emphases mine] (p.55). The notion of “phenotype” includes anatomical characters (“traits” or otherwise), patterns of behavior, “emotions,” instincts, impulses, and so on.

Thus evolutionary psychology, acknowledging that we are “evolved organisms,” attempts to apply a biologically informed sense of “programs” to understanding the human mind. Indeed, as Tooby and Cosmides write, “in species like humans, genetic processes ensure that complex adaptations virtually always are species-typical (unlike nonfunctional aspects of the system). This means that functional aspects of the architecture will tend to be universal at the genetic level, even though their expression may often be sex or age limited, or environmentally contingent” (Tooby and Cosmides 2000, p.1169). (It is important to observe that no explicit reference to

environment (in conjunction with regulatory genes) and need additional machinery to synthesize proteins; gene interactions have diverse classifications (linkage, pleiotropy, epistasis, etc.); other phenotypic levels require consideration; and so on—that result in X for an organism-in-an-environment. This background contextualizes closed and open programs; programs are placeholders for explicating the relevant context.
"traits" is made; discussion of this point occurs in a later subsection.) Furthermore, consistent with Mayr’s claims above, Tooby and Cosmides proceed to note that

the genes underlying complex adaptations cannot vary substantially between individuals because if they did, the obligatory genetic shuffling that takes place during sexual reproduction would break apart the complex adaptations that had existed in the parents when these are recombined in the offspring generation [see Mayr’s (1988) essay, “The Unity of the Genotype”]. All the genetic subcomponents necessary to build the complex adaptation rarely would reappear together in the same individual if they were not being supplied reliably by both parents in all matings (p. 1176).

In sum, talk of programs in evolutionary psychology is shorthand for talking about functions.6 It is here that the broad similarity between evolutionary psychology and Fodor’s view of modularity is most clearly seen, as Fodorian modules are compiled transducers that process information in very specific ways. However, unlike Fodor’s view, evolutionary psychology emphasizes information processing systems and biological function; to see just how significantly different these two views are, we must first examine two general senses of “modularity” and their relation to encapsulation.

Representational and Computational Modules: Distinguishing Two Conceptual Moments of Information Processing Systems

There is a general distinction between “representational modules” and “computational modules.” “To a first approximation, representational modules are domain-specific bodies of data (organized and integrated in the right kind of way); computational modules are domain-specific processing devices” (Simpson et al. 2005, p. 13). The difference lies in the emphasis placed on data on the one hand, and processing on the other. Conceptually speaking, it is useful to distinguish these two senses because we may want to emphasize computational modules (processes) whose outputs are then used to generate a representational module (data); or we may want to emphasize domain-specific representational modules whose data are manipulated by domain-general processes (processes which are not computational modules) (p. 14). To reiterate,

---

6 I discuss the epistemic vagueness of “program” and its relation to function in section 4.2.
we have the conceptual distinction between domain-specific representational *data*, and domain-specific *processing* which highlights the transformation of data into other kinds of data—data that may be representational (specific or otherwise) or not if the data aren't "sufficiently" representational.

**Fodorian Modules Versus "Functionally Individuated Information Processing Systems" ("FIIPS")**

Fodorian modules, I think, are best characterized as *narrow* "computational modules" since the criteria for these modules are rather specific. The extent to which Fodorian modules help generate "representational modules" depends on the degree to which the outputs are interpreted as representations. This difficult issue is not central to my concerns, so it is placed on the side. What matters is that Fodorian modules are perhaps the most narrowly defined type of *computational* module. The distinction between computational and representational modules is important for evolutionary psychology because while evolutionary psychology studies both types of modules, its overarching concern is with framing "modularity" from an evolutionary viewpoint. To explain, let me begin with a contrast. Fodorian modules are framed by a research program that is a part of cognitive science, and whose assumptions stem from traditional cognitive psychology and artificial intelligence—assumptions about peripheral systems, data structures, syntax, memory, etc. The implicit orientation regards architectures that solve the "combinatorial explosion problem" (or the closely related "frame problem"), which, informally speaking, is the problem: what are the requirements for a computational platform to effectively solve difficulties posed by the environment without being sidetracked by an indefinite number of other solutions (optimal, satisficing, suboptimal, or otherwise)?

---

7 Although it should be noted that language is a prominent example of a "*representational* module," which Fodor acknowledges as the one anomalous *central* process that is also "modular" in *roughly* his computational sense as well (see, for example, Fodor 2000).
Fodorian modules are geared towards solving the combinatorial explosion problem; the most important feature is that they are informationally encapsulated, a notion qualified by the remaining (eight) features of modularity (recall the crucially nuanced claim that X is informationally encapsulated to the degree that X does not have access to other systems as central processes do; this "degree" is qualified by the other features). But evolutionary psychology is not primarily interested in Fodor's narrow sense of a computational module; and while it is not unconcerned with the combinatorial explosion problem, it is not explicitly concerned with this problem either. Rather evolutionary psychology focuses on representational or computational modules in light of the wider search for "functionally individuated cognitive mechanisms" (Hagen 2005)—what I will call "functionally individuated information processing systems" ("FIIPS"). The combinatorial explosion problem is too general in form, and in some ways too narrow: the general problem does not suggest particular constraints on actual human minds, as the combinatorial problem applies to a broad array of computational platforms; and it is too narrow because Fodorian modules do not focus on biological solutions to problems posed by ancestral environments, the details of which need to be pursued ("narrowness" has to do with the focus on synchronic, not phylogenetic, solutions to the problem). Evolutionary psychology significantly shifts its focus by attempting to "determine how the brain changes the environment [via functionally individuated information processing systems] to facilitate and enable the reproduction of the organism. For [this research program], a radically different set of ecological concepts is critical: finding food and mates, besting competitors, avoiding predators and toxins, and helping kin [for example]" (p.163). These ecological concepts concern the Pleistocene "crucible" in which Homo sapiens (sapiens) appears, and thus are "radically different" compared to the traditional concerns of cognitive psychology and artificial intelligence—concerns that Fodor shares. Functionally individuated information processing systems are not sharply distinct
from representational and computational modules, and they certainly aren't confined to the
characterization of Fodorian modules. Evolutionary psychology's research program simply starts
in an importantly different, though related, place.

**Reverse-Engineering Functionally Individuated Information Processing Systems: Encapsulation, Inaccessibility, and Domain-Specific Inputs as Three Major Parameters**

So given that evolutionary psychology searches for "functionally individuated cognitive
mechanisms," and given that Fodorian modules are informationally encapsulated, what exactly is
the relation between encapsulation and functional individuation? It might appear that
encapsulation implies individuation; and if X is also functional, it seems that encapsulation is the
same as functional individuation. The converse also seems to hold, namely that if X is
functionally individuated then it is encapsulated. However, it is worthwhile to distinguish
encapsulation from functional individuation—and both of these from modularity—to avoid a
bevy of confusions. For example, encapsulation may be confused with Fodorian modularity
(informational encapsulation is just one feature of Fodorian modularity); encapsulation and
modularity are ambiguously referenced (would encapsulation refer to the domain-specific data, or
process?); domain-specific inputs could be conflated with encapsulation; and so on. To clarify
the concept of a functionally individuated cognitive mechanism and its importance to
evolutionary psychology, a range of conceptual possibilities will be given.

Generally speaking, encapsulation is distinct from "inaccessibility" (opaqueness), and
both are distinct from domain-specific inputs (Carruthers 2005a). Picture a processing system as
a "box"; the three concepts are distinguishable by virtue of the distinct informational relation each
has to the box. Encapsulation holds that the *internal operations* in the box cannot draw
machinery/information from outside; inaccessibility holds that *other systems* cannot have access
to the box’s internal operations/information; and domain-specific inputs are about the information
"coming into" the box. To use an analogy, if the box is an engine in a car, domain-specific inputs
would be things like gas; inaccessibility would be the (rather odd) idea that the machinery in “the engine” cannot be lent to other systems like the radiator (assuming the engine is just the stuff with pistons, etc.); and conversely, the engine cannot borrow from the radiator’s “machinery,” as the engine is encapsulated. All of these “functional parts” are separable, if only because these parts play a larger functional role for working cars. Schematically, the following diagram highlights the different relations.

Figure 7. Encapsulation, Inaccessibility, and Domain-Specific Inputs

Fodorian modules are all three of these things (in addition to their other features). Evolutionary psychology asks: what is the relation between these three things and functional systems? First, separating domain specificity from encapsulation results in the broader distinction between “access specificity/generality” and “process specificity/generality,” respectively (Barrett and Kurzban 2005). This is important because one thing can be widened and the other narrowed, among other combinations. For example, “some mechanisms might have [wide] access to large amounts of information in the mind but only [narrowly] process information that meet its input criteria [see their paper for examples]” (ibid). There can also be systems that exhibit narrow access and narrow processing. Both of these possibilities would be functionally individuated information processing systems; in the first case we would have systems that function according to processing abilities narrow enough to be individuated, and in the latter case a good example would be Fodorian modules, which are clearly classifiable as functionally individuated information processing systems.
A Conceptual Map of Extended Functionally Individuated Information Processing Systems: Tweaking the Three Parameters

More generally, once it is recognized that we can “tweak” all three distinctions, a number of other information processing systems are possible. Since the previous distinction between access specificity/generality and process specificity/generality “tweaks” the relation between domain-specific inputs and encapsulation, let us tweak the relation between encapsulation and inaccessibility. Encapsulation and inaccessibility are about process specificity/generality; permuting specificity and generality yields four possibilities:

Figure 8. Tweaking Encapsulation and Inaccessibility

To be clear, the distinction between access and process specificity/generality either holds fixed the input and then varies the specificity/generality of the encapsulated process, or the converse. The most problematic case is when the inputs are wide (access generality) and the processing is also wide (process generality), for it does not appear that these systems are functionally individuated information processing systems individuated in any reliable and non-arbitrary way. Likewise, the rightmost case in the above diagram is most problematic when the inputs are wide. As this diagram expresses degrees of process specificity/generality, the most specific case is the first one, where the process is both encapsulated and inaccessible; the second and third cases are less specific (or more general); and the last has the least degree of specificity. Given that Fodorian modules already fall under the broader notion of functionally individuated information

---

8 The third relation between domain-specific inputs and inaccessibility does not need explication, since the story is the same as the domain-specific inputs/encapsulation story; the only difference is the direction of the “arrow” indicated in the previous diagram.
processing systems (henceforth "FIIPS" for short), the question is, which of these cases are compatible with evolutionary psychology?

Case 1: Fixing Encapsulation and Inaccessibility, and Tweaking the Inputs

If we fix the inputs as (domain) specific, Fodorian modules fall under the first case. If we then remove the other characteristics of Fodorian modules and have only the characterizations of encapsulation and inaccessibility ("no borrowing and no lending"), this would clearly count as a FIIPS, conceptually speaking. (Note that the current and ensuing discussion is at the abstract level of mapping conceptual possibilities for understanding the broad features of FIIPS.) And if the inputs are not specific (access generality), encapsulation and inaccessibility (process specificity) would still be enough to have a FIIPS—indeed, encapsulation and inaccessibility would individuate the system; figuring out its function additionally requires examining the system’s behavior and, since it has access generality, the centralized processes or representational data in play.

Clarifying Some Scope Ambiguities

For the remaining cases, it must be noted that there is a scope ambiguity for the initial characterization of encapsulation and inaccessibility—to repeat, encapsulation holds that the *internal operations* in the box cannot draw machinery/information from outside, and inaccessibility holds that *other systems* cannot have access to the box’s internal operations/information. The ambiguity concerns the scopes of the "internal operations" and "other systems." The stronger scope for the former claims that *all internal operations* cannot draw information from the outside, and for the latter, *all other systems* cannot have access to the box’s internal information. The weaker reading claims that *many* internal operations cannot draw information from the outside, and that *many* other systems cannot have access to the box’s
internal information. (These scopes apply in the obvious ways to the first case.) On the strong reading an unencapsulated system means either that (SU1) some internal operations can draw information from the outside, or (SU2) it is possible that all internal operations draw information from the outside (the first reading is a denial of the universal quantifier; the second reading is a denial of the modal operator). On the weak reading it means either that (WU1) a few internal operations can draw information from the outside, or (WU2) it is possible that many internal operations draw information from the outside. Likewise, on the strong reading accessibility means either that (SA1) some other systems can have access to the box's internal information, or (SA2) it is possible that all other systems have access to the box's internal information; WA1 and WA2 are defined in similar fashion. With these distinctions we are ready to discuss the last three cases.

Case 2: Inaccessible but Unencapsulated Systems

The reason for these somewhat pedantic distinctions is to make clear the relation between the last three cases and classifying types of FIIPS; with these distinctions we don't need to consider all combinations, just the worst-case scenarios. For if FIIPS make conceptual sense in these scenarios, then they will make sense for the other combinations. The worst scenario for the second case (an inaccessible but unencapsulated system) is with SU2 and weak inaccessibility, since the system would be the least specifiable. SU2 poses a problem, since if all internal operations actually draw information from the outside, individuating the system would not appear feasible. As for the second condition (many other systems cannot have access to our system's internal information), this constraint might be enough to individuate the system (e.g., a higher-order system draws from any number of lower-levels in accordance with SU2, but most lower-level systems cannot access this higher-level). All that matters is that the system's outputs are

---

9 A similar distinction is made in Carruthers (2005a).
specifiable enough to determine the functional role it plays. While the issue of narrow or wide inputs matters, isn’t of decisive importance in this case; what bears most upon finding a FIIPS are the constraints expressed by weak inaccessibility, and the outputs that determine the system’s functional role.

Case 3: Encapsulated but Accessible Systems

For the third case, suppose the inputs are specific. Intuitively, if a process is encapsulated but accessible ("no borrowing, lending OK"), just as long as lending information to other systems doesn’t interfere with the internal machinery of the encapsulated process—so that it outputs whatever it is supposed to output—then it doesn’t much matter that it is accessible. The issue of “lag” in the time it takes to process information when the system is accessed by other systems might be a legitimate concern, but if an encapsulated, accessible system operates on specific inputs and is reliably “fast enough,” it will qualify as a FIIPS. The same lesson may apply when the inputs aren’t specific; just as long as the behavior is reliable and temporal considerations do not hamper the individuation of function, the system will qualify as a FIIPS. The worst-case scenario occurs with the weak reading of encapsulation and SA2. The reasoning is similar to the second case—when all other systems actually have access to the box’s internal information, the prospects for individuation are placed on weak encapsulation.¹⁰ For the present case, the slightly different emphasis is that when temporal considerations are not crucial—that is, temporal considerations don’t make these systems intractable—FIIPS may in principle be found.

Case 4: Unencapsulated and Accessible Systems

The fourth case, as mentioned, is the most problematic. The worst scenario is with SU2 and SA2. Suppose it is actually the case that all internal operations draw information from the outside, and all other systems have access to the system’s internal information. It does not look

¹⁰ E.g., a crucial low-level system is “monitored” by all other systems, yet many of the system’s operations are internal to the system—most machinery cannot be borrowed from other systems.
as if there are narrow inputs or narrow outputs. If anything is a global system, this case is it, and it appears that finding FIIPS will not occur in any tractable manner. The other combinations may be examined, although I will not do an exhaustive exploration; rather, I will look at the best-case scenario. The best case is WU1 and WA1: a few internal operations can draw information from the outside, and a few other systems cannot have access to our system’s internal information. This case seems no worse than the worst scenarios in cases two and three; I will not argue for the claim that FIIPS can be found here, but merely suggest that it may be viable.

**How Far Can the Extended “Modular” Approach Extend? Applying the Above Map**

The question arises at this juncture, how far can the search for FIIPS extend—i.e., what is the extent of the “extended modular” approach? According to the above conceptual map, each of the cases allow for the possibility of FIIPS, although to differing degrees. The question of extent brings us to the issue of scope—is the mind nothing but FIIPS (or programs)? Our conceptual map reveals that there are a number of avenues for pursuing the scopes of FIIPS; whether the issue is about the mind consisting of all FIIPS, or many FIIPS, it must first be clarified what sense of FIIPS is employed.

**Reverse-Engineering FIIPS and the Environment of Evolutionary Adaptedness (“REA”)**

Evolutionary psychology includes Fodorian modules, so clearly the issue of scope exceeds that particular sense of “module.” The type of FIIPS that evolutionary psychologists examine may draw, in principle, from any of the four cases. Accordingly, the empirical research program searches for specialized abilities in the cognitive system, examines the stereotypical format of inputs for that specialization, and speculates on the specialization’s function relative to the distinction between the “proper domain”—the environment in which the purported function evolved—and the “actual domain,” which concerns the function’s behavior in the current environment (Barrett and Kurzban 2005). The first two tasks result in a flowchart of boxes; these
tasks search for the range of inputs to specialized boxes that are connected in particular ways to generate a specialized range of outputs.\textsuperscript{11} Each box may be a FIIPS, or may just be a processing device that is not really "functional" on its own; but the total flowchart is itself a FIIPS (e.g., a fear FIIPS, a cheater-detection FIIPS, a folk-biology FIIPS, a folk-physics FIIPS, etc.). The first two tasks, which are "established" through experimental methods (similar to those mentioned in Ekman's work), are guided by the third task, which is to find a plausible working hypothesis about the adaptive, evolutionary value of the FIIPS. The general (philosophical) framework assumes that most psychological adaptations evolved in the Pleistocene when our ancestors were hunter-gatherers (the "EEA"—the Environment of Evolutionary Adaptedness). Selective pressures postulated within the proper domain of the EEA—checked against reconstructions offered by anthropologists—grant adaptive value to the FIIPS. Thus the claim that programs "function particularly harmoniously when confronting cross-generationally recurrent situations" is a (educated) conjecture about the evolution of FIIPS in the EEA.

However, we need to be clear about what conception of FIIPS guides research. A conservative approach keeps FIIPS as tightly characterized as possible; and should future research demand more generous conceptions (like Case4), this demand will only be seen by adopting first a narrow conception that foregrounds any subsequent inadequacies. The conservative approach is a piece of procedural methodology, distinct from the general concepts that give the research program its direction. It is with respect to the latter concern that Tooby and Cosmides claim that emotive-cognitive systems are superordinate programs. The concern is not with any particular account of FIIPS—accounts which may differ from system to system—but rather with the coordinated role that programs play, whether this coordination falls under our first case or under the more generous conceptions from the remaining three cases. The "specialization

\textsuperscript{11} This is a particular type of reverse engineering, which I link to the "functional stance" in the following section.
of an adaptation for a function does not lie in the specialization of all the parts to its function. The specialization lies in the way the particular interrelationships [—that vary from system to system—] of the parts is [sic] coordinated to solve the specialized adaptive problem with particular efficiency” [emphasis mine] (Tooby, Cosmides, and Barrett 2005, p.336).

Not any coordinated information processing system will do. Even if the first two empirical tasks of the research program issue in a flowchart of boxes (an individuated information processing system), the question remains whether it is a functional system that evolved in the EEA; for if it is an artifact of ontogeny and the exposure of experimental subjects to similar (current) environmental factors, then evolutionary psychology would not count it as a FlIPS. The third task is really about the idea that FlIPS appear to express species-typical adaptations. But what sorts of adaptive “traits” are these? The question seems to be whether these “traits” are monomorphic or polymorphic (see section 3.1). Griffiths, in his critique of evolutionary psychology’s account of higher-order emotions, charges that this research program is inaccurate because it is committed to a monomorphic view of the mind, a biologically inaccurate view because many of our traits are polymorphic; the monomorphic view portrays too narrow a view of how the mind works.

“Traits” and Evolutionary Psychology?

It is not clear exactly what evolutionary psychology claims on the issue of traits, although evolutionary psychology searches for species-typical characteristics that are “universal”—a concept rooted in the general genetic architecture of the species. That is, the genetic potential is universal, while the expression of that potential differs in relation to environmental affordances.

---

12 This view holds that the major (species-typical) cognitive “traits” for people are monomorphic “traits” (see fn.13). This thesis argues that apparent cognitive “adaptations” acquired/learned by people across cultures are not a result of genetic differences, but rather a result of differences in environments. These apparent “adaptations” are really expressions of underlying, species-typical adaptations. Monomorphic psychic “traits” concern the same set of genetically “determined” developmental potentials—the same “developmental program.”
Traits like sex, which are standardly viewed as polymorphic, are more generally conceptualized as expressing this universal potential; the potential for sex determination is species-typical, and differentiation into discrete types is a matter of certain switches being turned on and others being turned off in the course of development (e.g., Tooby and Cosmides 2005). Notions like “genetic architecture,” “potential,” etc. are placeholders whose mechanisms other disciplines explore. By employing these placeholders, evolutionary psychology emphasizes that we are a species, meaning in its barest form that there is interbreeding viability; “universal genetic potential” is foregrounded by this “viability.” On a sympathetic reading, it is inappropriate to attribute either a monomorphic or polymorphic view to evolutionary psychology, as the more precise biological senses of these terms are not wholly addressed by the research program. Again, the technical senses are black-boxed so that researchers can get on with looking for functional regularities (FIIPS) at a higher level of investigation. Griffiths’ objections to the monomorphic view of mind and its reliance on genetic “programs” (see pp.122-32) are slightly misleading; as I’ve argued previously, the notion of “programs” is shorthand for “functions,” in particular FIIPS.13

FIIPS as Reverse-Engineered “Boxologies”

The legitimate issue concerns the phenomenological level evolutionary psychology focuses on when it produces “boxologies” (flowcharts mapping the inputs, processing systems, and outputs that jointly delineate FIIPS). Evolutionary psychology black-boxes the causal mechanisms of actual genetic-neural-physiological-developmental pathways and emphasizes the semi-causal and phenomenological nature of such boxologies (which are “semi-causal” in that the models still appeal to black-boxed causal pathways; and “phenomenological” in that the

---

13 Actually I am quite sympathetic to Griffiths’ critique, as Tooby and Cosmides (1990, 2005) appear to strongly suggest a monomorphic reading. While “monomorphic traits” is not mentioned, their 1990 article employs the language of “monomorphisms” that are “functionally integrated.” The issue of monomorphic versus polymorphic traits at the psychological level would need to get straight on other multiple levels involved, how “trait” is being employed, how function relates to discrete and continuous characters, etc. Here I “wisely” (fearfully) avoid this bog with my “sympathetic reading.”
flowcharts track correlations which “save the phenomena”). An aspect of evolutionary psychology’s philosophical framework wagers that starting first with investigating FIIPS and subsequently exploring various “non-functional” processes will not only prove a fruitful method for understanding the mind, but also as other interdisciplinary areas concerned with genetics, neuroscience, and so forth are brought into the fold, it will be revealed that the causal factors mesh with evolutionary psychology’s higher-order results (e.g., see Gazzaniga 2000, section X). Accordingly, Griffiths’ critique (slightly) misses the mark because his focus on monomorphic traits misconstrues what evolutionary psychology’s models are primarily about—the fidelity of FIIPS are evaluated at the semi-concrete (semi-causal and phenomenological) level; it is at this level that boxologies are produced, as guided by adaptive hypotheses for why a FIIPS evolved in the EEA.

**Competing Research Programs within “Evolutionary Psychology” (Broadly Construed)**

However, as alluded to (fn.13), Griffiths’ critique has real substance; an element of his critique goes to the heart of competing research programs within “evolutionary psychology,” broadly construed. Specifically, there is a separation between the Santa Barbara School and “behavioral ecology.” Additionally, Griffiths describes another competing program different from evolutionary psychology, namely “Developmental Systems Theory” (I cover this program in greater detail in the following chapter). One major source of conflict between the three research programs is the EEA and its relation to FIIPS. On the behavioral ecologists’ side, William Irons argues for an alternative to the EEA—the Environment of Evolutionary Adaptedness—namely the “ARE”—the Adaptively Relevant Environment (Irons 1998). The commitment remains to the idea that an organism has special-purpose adaptations keyed to aspects of the environment, but Irons differs from the Santa Barbara School in his claim that when a part of the environment changes, the changes will probably affect some adaptations and not others. For Irons, the general
picture of evolution is that evolution occurs in "mosaic" (p.194) fashion, which contrasts with the EEA since the EEA suggests a period of stasis—a crucible in which our species-typical adaptations were formed. Ever since the Pleistocene our cognitive architecture has been (more or less) the same according to the Santa Barbara School; but this claim has problems, as other evolutionary mechanisms could plausibly have been in operation; and if so, a different picture would result of how our minds evolved. Avoiding the details, the upshot is that mosaic evolution presents the alternative view that things haven’t been fixed since the Pleistocene. Rather the adaptive pieces of our minds "stabilized” over time, which makes it appear as if our cognitive architecture has been “the same” since the Pleistocene, but more probably mosaic changes occurred within the Pleistocene and after that as well.

Why this shift to the ARE matters is that behavioral ecologists like Irons are not primarily concerned with finding FIIPS. Instead the focus is on environmental changes that may have led to changes in evolution. It isn’t that Irons is opposed to the idea that the mind consists mainly of FIIPS. Indeed, he believes that this element of the Santa Barbara School is worth exploring. What he is drawing attention to in his critique of the EEA is that the Santa Barbara School’s experimental studies are framed by philosophical assumptions about the EEA. His nuanced critique is that experimental evidence for domain-specific decision making ("emotional," "rational," or otherwise) is based upon a contrast class between the actual domain of decision making and the proper domain. For example, in the experimental setting for the Wason selection task (which involves verifying the material conditional), the actual domain—the results of the experiment—suggests that people are poor reasoners; but when the experiment employs the material conditional within a social context for detecting cheaters, people appear to be rather good reasoners. The proper domain of the material conditional is domain-specific to social
exchanges dealing with the detection of possible cheaters. And this proper domain, according to the Santa Barbara School, is actually the EEA, where the cheater-detection FIIPS evolved.\textsuperscript{14}

The experimental studies comparing the cheater-detection FIIPS with the actual domain (of poor reasoning) are framed by the search for what evolved in the EEA. Irons' objection is that if mosaic evolution is a more accurate picture of our evolution, it is difficult to know the referent of the "proper domain," and this in turn has consequences for understanding developmental programs and their relation to FIIPS. Even if robust results are found supporting the existence of FIIPS, we would still need to figure out when the mosaic "pieces" of the mind evolved, what environmental pressures were present, and the exact meaning of "pieces" (especially if a piece is a response to the pressure of fluctuating environments). In essence, the manner in which I've presented the above critique is the core of David Bueller's argument against the Santa Barbara School. Bueller's general critique, stated in slightly different terms, holds that because a mosaic view of evolution requires attention to different evolutionary mechanisms—that operate in different ways, with different degrees of strength, for different environmental pressures, and at different times—our minds haven't been adapted since the Pleistocene; our minds evolved "piecemeal" to arrive at a species-typical architecture, but this is the result not of the EEA, but the result of variegated environmental and temporal details (neglected by the Santa Barbara school) within our hominid line. In addition, Bueller crucially shifts the argument by drawing the philosophical implication that, if our minds evolved mosaic-style, our minds are potentially still adapting. The Santa Barbara school offers a view of the adapted mind, but mosaic evolution suggests that part of our "species-typical architecture" is the ability to

\textsuperscript{14} See Cosmides and Tooby (2005) for an overview of roughly twenty-five years of research on this topic. (This article also contains some implicit replies to aspects of Bueller's (2005) critique of cheater-detection FIIPS.)
reorganize aspects of that architecture; an adaptation of our minds is that they are adapting minds (cf. Bueller 2005, p.102).15

**The Santa Barbara School, Behavioral Ecology, and Developmental Systems Theory (“DST”)**

The philosophical frameworks of the Santa Barbara school and the behavioral ecologists can be summed-up by different attitudes to the relation between FIIPS and human cognitive adaptations: the Santa Barbara school strongly suggests a monomorphic mind will be found, while behavioral ecologists search for mosaic environmental pressures that indirectly bear upon possible FIIPS. Bueller lies along this “continuum,” although he is father to the “left”; he is closer to left-wing Developmental Systems Theory. This approach focuses on polymorphic factors involved in the lifecycle of an organism. If it searches for a view of the mind, it would be the “polymorphic mind.” Developmental Systems Theory is both more philosophical in its framework compared to evolutionary psychology, and also (strangely enough) more causal-concrete in its orientation. Fidelity is a key issue since it tacitly bridges these two things; it is also a primary reason why the programs talk past each another to a significant extent (as I discuss in chapters five and six).

Another way to situate Bueller’s emphasis on adapting minds is in light of Fodor’s claim that the most important central processes are holistic; domain-general processes like learning and creativity are thought to be inadequately accounted for by FIIPS. While even Bueller grants that various modes of processing can be captured by the FIIPS approach—an approach which he

---

15 Evolution “has not designed a brain that consists of numerous prefabricated adaptations, but has designed a brain that is capable of adapting to its local environment” (p.136). In contrast to the Santa Barbara School’s use of “developmental programs” (see Griffiths 1997, pp.127-32 for an overview of the types of developmental programs), Bueller offers the alternative language of “phenotypic plasticity,” which “refers to cases in which a genotype builds a mechanism or process that is capable of producing phenotypic change or reorganization in response to changing conditions in the organism’s environment” (p.137). Note that such plasticity is a domain-general process—the very sort of thing that evolutionary psychologists argue would be highly unlikely to evolve as an adaptation (i.e., a “hopeful monster”). For responses to Bueller (2005) go to http://www.psych.ucsb.edu/research/cep/bueller.htm.
believes can make respectable progress—it probably cannot capture the more “cognitive” processes: evolutionary psychology “may be right about some of our more basic emotional adaptations, but nonetheless wrong in its claims that we possess a lot of cognitive adaptations devoted to very specific forms of problem solving” (p.143). And so the debate continues as to how far the extended modular approach can extend—how far it can account for emotive-cognitive centralized processes.

**A General Overview of the Santa Barbara School’s Research Program: Spaces for the Human Sciences?**

To briefly recap, Ekman’s program focuses on “emotions” technically defined as affect programs; higher-order notions appeal to (unspecified) relations between affect programs and cultural “display rules.” Display rules utilize a (indefinite) range of concepts stemming from the human sciences. Ekman’s program, though, spends most of its time understanding affect programs; display rules are simply acknowledged as demarcating the “other category” concerning concepts from the human sciences. Evolutionary psychology also creates a space for more humanistic concepts, however the distinction is not between FIIPS and display rules. Rather, to the extent that evolutionary psychology would utilize specific humanistic concepts, these concepts must be constrained by information processed by FIIPS. For example, given the distinction between the “actual domain”—the current environments—and the “proper domain”—the EEA—evolutionary psychology would utilize “humanistic” concepts (analyzing, say, anthropological aspects of religious practices) only insofar as comparisons could be made between information-processing relative to a current environment, and information-processing relative to the EEA (e.g., Boyer 2000).

More generally, the Santa Barbara School’s research program lays out the following model of how its understanding of mind relates to cultural concerns:
Their model connects [not "reduces"] the social sciences to the rest of science by recognizing that:

- the human mind consists of evolved information-processing mechanisms instantiated in the human nervous system;
- these mechanisms, and the developmental programs that produce them, are adaptations, produced by natural selection over evolutionary time in ancestral environments;
- many of these mechanisms are functionally specialized to produce behavior that solves particular adaptive problems, such as mate selection, language acquisition, family relations, and cooperation;
- to be functionally specialized, many of these mechanisms must be richly structured in a content-specific way;
- content-specific information-processing mechanisms (regarding those mechanisms that are functional (see "Programs and Evolutionary Biology" in section 4.1)) generate some of the particular content of human culture, including certain behaviors, artifacts, and linguistically transmitted representations;
- the cultural content generated by these and other mechanisms is then present to be adopted or modified by psychological mechanisms situated in other members of the population;
- this sets up epidemiological and historical population-level processes; and
- these processes are located in particular ecological, economic, demographic, and intergroup social contexts or environments (emphasizes mine) (Tooby and Cosmides 1992, p.24).

Rather than highlighting a distinction akin to that between affect programs and display rules, more prominently, connections should be made between species-typical FIIPS and (the constraints they impose on) "epidemiological" population-level processes (points e through h). These connections would be "borderland" connections between fields that mutually inform one another. Thus insofar as there is a distinction to be made between evolutionary psychology and the human sciences, it would be (according to the Santa Barbara School) a distinction between respective semi-autonomous fields of investigation (that still mutually inform each other). This first distinction would create a space for certain human sciences—a space concerning mutually informed relations between evolutionary psychology and the relevantly constrained human-science concepts.

However, from a broader perspective, there is a further (fuzzy) distinction between evolutionary psychology and the human sciences, since, as point e acknowledges, FIIPS only
generate some of the content of human culture. Hence if a fuzzy distinction is to be drawn between FIIPS and the human sciences—"fuzzy" since the human sciences include an indefinite array of social sciences, like those mentioned in point h—it would be between FIIPS as (functional) adaptations and non-"functional" processes (those not generated by adaptations) which are artifacts of ontogeny, environment, certain social processes, etc. This second distinction would create a space for other concepts from the human sciences—a space concerning processes that are not explicitly generated by cognitive adaptations.

4.2: The Functional Stance to Information

It might appear that my focus on FIIPS has lost the thread of emotions. But the emphasis in this chapter is on "extended modular" approaches to understanding emotions systems, as the larger concern is with different approaches to emotions from "scientific" viewpoints; for part of what apparently qualifies certain approaches as "scientific" has to do with considerations about modularity and kinds. My exposition of the general features of FIIPS, by way of emotions, illustrates possible extensions to the modular approach. Accordingly, I have not lost the thread of "emotions," at least insofar as accounts of particular emotions are not the relevant topics to pursue concerning what makes these accounts "scientific." Thus what matters most are the general features exhibited by "emotion" systems, precisely because these features carry over to understanding FIIPS. (Most importantly, remember that evolutionary psychology rejects a hard distinction between "emotive" and "cognitive" processes; rather evolutionary psychology places emphasis on modes of evolved affective operation—on evolved centralized systems.) The question I am pursuing is whether there can be a science of emotion. The promise of the extended modular approach is that if the mind is nothing but (or more weakly, very many) FIIPS, then it would appear that it is possible to establish a science of emotion. But as I’ve argued
above, this claim is subject to controversy since it is unclear just how far this approach can be extended. While the program has promise, as it is a part of evolutionary biology and offers (via its philosophical framework) a “vision” of its general scope, from an empirical standpoint the bounds of the approach are still unclear.

Organizing Affect Programs and FIIPS

This section organizes the conceptual features of “modularity” discussed in section 4.1 and Chapter 3. First, observe that the language of FIIPS encompasses affect programs, as affect programs are Fodorian modules (see section 4.1, “Reverse-Engineering Functionally Individuated Information Processing Systems” ff.). More generally, FIIPS are systems: their constituents (the elements of superordinate programs) are situated in relation to each other (the elements are coordinated so as to work as modes of operation) and to the system as a whole (these modes are configurations of the entire system). Individuation occurs relative to the information that is inputted, processed, and outputted; and these last three notions, in turn, are aspects of functional systems.

A Conceptual Approach to “Function”

It was noted in section 4.1 that an emotion is a “superordinate program,” which is shorthand for “function.” It was also noted that a program transforms energy into order, another way of stating that programs engage in information processing. So function appears to be inextricably intertwined with information. What, then, are function and information? I don’t think these abstractions can be precisely defined without losing their broad and useful import; while there are more precise definitions of “information” and “function” in mathematics, the less precise use of these abstractions elsewhere (e.g., in numerous biological disciplines, in technology) suggests that they are 1) theoretical terms not subject to tight technical definition, and 2) conceptual placeholders for organizing other concepts. My concern here is with
"function" and "information" as conceptual placeholders for organizing FIIPS; what I call the "functional stance" to information provides this organization.16

Some Brief Historical Background

Before describing the functional stance, I first give some historical background to situate from where the stance originates. The predominant orientation in cognitive science is a functionalist/computationalist orientation. As Howard Gardner notes in The Mind's New Science, the "cognitive revolution" began with John von Neumann's work The Computer and the Brain. Reflections on the electronic computer, computation theory, and an emerging picture of the brain's neural operations gave way to conceiving of mental processes as computational; the metaphor "the mind is a computer" (however that is understood—and there are many interpretations) stands at the core of computationalism's rise to prominence. Additional work by Claude Shannon and Norbert Wiener on information, and the multifaceted applications of information in early artificial intelligence led to computational approaches in biology,17 psychology, linguistics, and computer science.18 Cognitive science emerged from the rich exchanges between these key disciplines. More recently, a fair amount of what is at the forefront of contemporary science has a computational orientation.19 Given that "information" is intertwined with computationalism/functionalism in multiple ways, the functional stance towards information captures the conceptual features of these multiple interrelations.

The Functional Stance

16 Here I focus on "function" (see fn.20 below); I defer discussion of "information" and the notion of a "stance" to section 9.4. In Chapter 7 I discuss more precise senses of "function" in relation to biology.
17 As noted, Mayr makes a qualified appeal to the informational notion of a "program" as what distinguishes teleonomic processes from telomatic processes (1988).
The functional stance encompasses two senses of "functionalism." The first sense was discussed in Chapter 3, section 3.2; to recap, the first sense involves a top-down approach to understanding a complex process, where the process is broken down into parts in order to grasp how those parts "function" in relation to the (functional) whole. The approach is often called "reverse engineering." In this sense, functionalism is primarily a research strategy used to explain certain phenomena. The second sense employs Daniel Dennett's notion of the "intentional stance," which, in brief, tracks and predicts the behavior of a complex system by adopting an as-if attitude towards the "beliefs," "desires," and "goals" of the system in question. What this stance adds to reverse engineering is a way to reduce the number of "live options" regarding putatively functional systems. Together, reverse engineering and the intentional stance comprise what I am calling the "functional stance."

Terminological Clarifications

Recall from section 3.2 that reverse engineering consists of the computational, algorithmic, and implementational levels of analysis. Recall also from the beginning of Chapter 2 that Figure 1 lists three levels of analysis for cognitive psychology: the ecological, computational, and implementational levels (see also Sterelny 1990). While the terminology differs slightly, basically the former "algorithmic level" is the same as the latter "computational level" (both "implementational levels" are the same). However, while the former "computational level" and the latter "ecological level" are roughly the same—as both levels are concerned with the architectural roles of a system under study—the ecological level also includes a dimension

---

20 In Chapter 3, section 3.2 fn.10, an analytic characterization of "Cummins functions" was given for analyzing a system in terms of certain functional capacities. Cummins functions have been criticized as applying to potentially any and all systems which a researcher chooses to analyze as "functional." There are two things to note: 1) reverse engineering is a more explicit research strategy for analyzing certain functional capacities—a strategy which already has widespread contextualized application in the cognitive sciences; and 2) the intentional stance provides further constraints on reverse-engineering putative functional systems.
not explicitly included within the computational level. The additional dimension is Dennett’s intentional stance, which further constrains the types of systems analyzed by the computational level (see fn.20).

The Intentional Stance

Reverse engineering starts at the computational level, whose generality allows for application to many systems. But if the computational level yields too many “parsings” (i.e., it is not clear how best to break down the overall task into subcomponents), or the search for constraints appears infeasible, the intentional stance is required to distinguish behaviors. Dennett illustrates the need for the intentional stance by first discussing where it ought not to be applied: “Why should we disqualify the lectern? For one thing, the strategy does not recommend itself in this case, for we get no predictive power from it that we did not antecedently have” (1987, p.23); its behavior is sufficiently boring to be explained by appealing to its physical properties, its “functional” design, etc. But in “the case of people or animals or computers, the situation is different. In these cases often the only strategy that is at all practical is the intentional strategy; it gives us predictive power we can get by no other method” (ibid). The way the intentional strategy/stance works is that “first you decide to treat the object whose behavior is to be predicted as a rational agent”—that is, the agent has more-or-less consistent beliefs relative to dealing with aspects of the world, and is rational enough to draw certain implications from such beliefs. Then “you figure out what beliefs that agent ought to have, given its place in the world and its purpose”—that is, figure out what “true” (i.e., warrantedly assertible) beliefs the agent ought to have to cope in its environment. Figuring out its beliefs is intertwined with figuring out the desires it ought to have, desires satisfying the most basic needs like survival, food, etc. And
finally "you predict that this rational agent will act to further its goals in light of its beliefs" (p.17).²¹

A host of objections have been raised to Dennett, usually based on the charges that the stance is too behavioristic, and too externalist since it is a third-person account.²² Responding to the first charge, Dennett privileges not behavior but more subtly the role that internal representations play in regulating behavior for intentional systems (p.32). As for the second charge, the stance only makes sense on the assumption that we "are" intentional systems who understand other intentional systems by implicitly adopting the intentional stance; informally, the other person is like myself and has folk states that make certain behaviors predictable, and even surprising.²³ Our first-person experiences (for "competent" experiencers) are already bound up with third-person considerations. Another objection is that just as people may predict wrongly, the intentional stance may fail to predict fruitfully, thus casting doubt on its efficacy to help understand various systems (which, let us suppose, are also not captured by reverse engineering). But as Dennett notes, even if the intentional stance produces a host of predictions but does not distinguish the most likely outcome (or a reasonably-sized set of likely outcomes), it can still "dramatically reduce the number of live options" (p.24). Dennett in fact turns this objection in his favor—the strength of the intentional stance is its avoidance of fine-grained predictions that would lose the qualitative behavior(s) of interest. The intentional stance generally orients the

---

²¹ Dennett gives two nice examples that are also clearly dispositional: "The strategy even works for plants. In a locale with late spring storms, you should plant apple varieties that are particularly cautious about concluding that it is spring—which is when they want to blossom, of course. It even works for such inanimate and apparently undersigned phenomena as lightning. An electrician once explained to me how he worked out how to protect my underground water pump from lightning damage: lightning, he said, always wants to find the best way to ground, but sometimes it gets tricked into taking second-best paths. You can protect the pump by making another, better path more obvious to the lightning" (p.22).
²² For a sympathetic reconstruction of Dennett, see Elton (2003).
²³ Humor, for example, depends on a fund of expectations in order to juxtapose the "unexpected" punch-line.
investigation by suggesting live possibilities (or reducing the number of live options), upon which reverse engineering can further be deployed, should it be appropriate.

The **functional stance** simply consists of the two moments of reverse engineering and the intentional stance. The upshot of all this is that the functional stance delineates the conceptual, "boxologized" features of FIIPS, Fodorian or otherwise. Clearly reverse engineering captures a fair amount of what occurs in Ekman's program and evolutionary psychology; but it may appear that the intentional stance does no work in accounting for FIIPS. This would be mistaken since if sense is to be made of the link between FIIPS and folk categories (like those expressed in folk psychology and folk biology), the intentional stance must be presupposed in order for the project to get underway. In other words, the intentional stance orients the empirical search for FIIPS since the investigation of evolved cognitions requires a prior fund of background beliefs *about* the "beliefs," "desires," etc. that are being boxologized.

Specifically, the intentional stance, as utilized in evolutionary psychology, frames FIIPS from an *evolutionary* viewpoint, placing specific types of constraints on FIIPS that are not brought into relief by reverse engineering alone. As Tooby and Cosmides write,

> to *reverse-engineer the brain*, one needs to discover functional units that are native to its organization. To do this, it is useful to know, as specifically as possible, what the brain is for — *which specific families of computations* it was built to accomplish and what counted as a biologically successful outcome for *each problem-type*. The answers to this question must be phrased in computational [reverse-engineered] terms because that is the only language that can capture or express the functions that neural properties were *naturally selected* to embody. They must also refer to the ancestral activities, problems, selection pressures, and environments of the species in question because jointly these define the computational problems each component was configured to solve [emphases mine] (2000, p.1168).

"Reverse-engineering the brain" concerns *primarily* the algorithmic level, which outlines functional units — what I’ve called "boxologies" — as "families of computations" — the component parts of each box, and how they interrelate to each other and to the box as a whole. These families, in turn, are *framed* by the computational level, which outlines the architectural roles of
the functional units—the wider functional roles of the boxes concerning “each problem-type.” 
Lastly, since computational problems are defined in relation to “natural selection” (“the ancestral activities, problems, selection pressures, and environments of the species in question”), the intentional stance is required to further constrain computational problems—the intentional stance reduces the number of live options concerning links between FIIPS and folk categories (e.g., folk physics, folk biology, etc.).

4.3: Summary

In conclusion, the functional stance brings into relief a “4-D subspace” within our four-dimensional space that captures the conceptual features of FIIPS. This subspace emphasizes the causal-phenomenological axis by virtue of its degrees of concreteness; the fidelity dimension depends primarily on the what the reverse-engineered models emphasize: for evolutionary psychology the FIIPS are mainly phenomenological, and have high-fidelity relative to that aim; affect programs have a greater degree of fidelity since the models at the computational, algorithmic, and implementational levels triangulate. Lastly, the functional stance extrapolates from the best that various “modular” perspectives—like Ekman’s program and evolutionary psychology—currently have to offer, and as such is appropriately categorized as occupying the paradigm-and-research-program region more prominently in comparison to the programs I discuss in the following two chapters; for the degree of integration between theories (modeling FIIPS), methods (reverse-engineering the features of FIIPS), and values (commitment to evolutionary biology) is much stronger than the research programs to come.

Conceptually speaking, this 4-D subspace can be thought of as a 4-D “shape” whose general contour is an arc (“half-oval”) that is best outlined by the concrete-abstract and causal-phenomenological axes:
In chapters five and six, the next two programs are situated in relation to this diagram. The Cognitivist Program is situated (by the “strategic stance”) in a 4-D subspace above and overlapping with this half-oval, whose shape is a “crescent”; and the Social Constructivist Program is situated (by the strategic and “semiotic” stances) in a 4-D subspace that heavily overlaps with this 4-D crescent but also is distinguishable from it. The semiotic stance delineates where the (fuzzy) distinction occurs.
Chapter 5: Cognitivist Approaches to Emotions

Recapitulation

Chapters three and four concern "functional" approaches to emotions. In particular, affect programs are (peripheral) functional systems whose functional features are primarily understood via reverse engineering. Evolutionary psychology's more general notion of (centralized) functionally individuated information processing systems (FIIPS) situates these systems as "functional" both in the reverse-engineering sense and, most importantly, in terms of evolved cognitive functions/adaptations (see also Chapter 7).

Prospectus

As we move towards the groupers' end of the spectrum, the emotions research programs downplay the role of "function" (sensu chapters three and four), and the broader concerns with temporality and social relations come to the fore—"mid-level" concerns that are situated squarely within the human sciences. The two research programs illustrating these temporal and social considerations are the Cognitivist Program and the Social Constructivist Program (which I discuss in Chapter 6). To give a brief overview of each program, individual "appraisals" are the former program's key idea, where emotions involve appraisals of events of significance for the organism (emotions are "cognitive" in this sense), and where the theoretical framework for understanding appraisals is expressed by a dimensional approach. Social "syndromes" are the latter program's central notion, which are social scripts (similar to Ekman's display rules) bearing upon individuals' judgments about interpreting emotions; these social judgments are cashed-out in terms of active appraisals of events of (social) significance. Situating chapters five and six in relation to the functional approaches discussed in chapters three and four, the two programs are
dimensional approaches to emotions contrasting with discrete approaches. The Cognitivist and Social Constructivist Programs emphasize temporality—considerations regarding how emotions experientially have a continuous and ever-changing nature, how emotions are centrally bound up with coping-in-the-world, and how coping is enmeshed with group-related considerations (social scripts, communication, institutions, etc.).

In spite of the differing foci of dimensional and discrete approaches, there is one important commonality between evolutionary psychology and the Cognitivist and Social Constructivist Programs: all three of these programs hold that there is no strong divide between “emotion” and “cognition.” Evolutionary psychology, though, examines emotive-cognitive systems from the standpoint of (evolved) function, and is less holistic in this respect; the other two programs examine emotive-cognitive systems from the standpoint of developmental and social processes, and thus are relatively more holistic. Generally speaking, these three programs investigate increasingly centralized emotive-cognitive systems utilizing differing “conceptual frames.”

The focus on more holistic factors in emotions pays the general price of less projectibility, and as a result the Cognitivist and Social Constructivist research programs are to a greater degree philosophical frameworks compared with the functional approaches. However, it would be misleading to claim that because they rely more heavily on philosophical frameworks they are not scientific. For remember that our four-dimensional framework locates different types of sciences from the standpoint of heterogeneity; accordingly, a sympathetic interpretation of these holistic approaches holds that they seek to organize a diverse array of higher-order factors respecting both our lived experience of emotions and the social structures that inform such experience (even though we might not be fully aware of how these structural considerations operate).
In section 5.1 I discuss the Cognitivist Program. I expound the general import of the dimensional approach and the way in which the Cognitivist Program adds to that approach. I also relate the program to functional approaches showing that the program differs in degree, not kind, from functional accounts; I argue that the program has a higher-order orientation with a strong phenomenological flavor and both a semi-concrete and a very abstract framework. Issues dealing with fidelity stem from this framework. I then discuss an extension to the Cognitivist Program; the extension concentrates on the latent communicative dimension of appraisals, as illustrated by Robert Frank’s commitment model of emotions.

Section 5.2 steps back from the coverage of the Cognitivist Program and situates things within the larger framework of Developmental Systems Theory (DST). The focus is on the key conceptual features of DST that encompass the conceptual features of the Cognitivist Program; as a “corollary,” the Cognitivist Program can benefit from some of the added “machinery” of DST. I consequently outline the second of my informational stances, the “strategic stance,” and show that it accommodates the Cognitivist Program, as well as aspects of DST. The primary reason for why the strategic stance is compatible with these two frameworks is that its methodological and epistemic features cohere with the manner in which the Cognitivist Program and (aspects of) DST start with similar “ecological” perspectives. In particular, in terms of our four-dimensional framework, these two frameworks carve out a subspace that forms a “crescent” above (and overlapping with) the subspace carved out by the functional approaches; this 4-D crescent is both more causally-based (in some ways) than the functional approaches, and more expressive of a philosophical framework whose models exhibit a strongly abstract flavor.

5.1: The Cognitivist Program

Distinguishing the Modern Cognitivist Program from Two Other Traditions
An "appraisal" seems to connote a deliberative, reasoned judgment about some subject matter. While early cognitivist theorists (e.g., Magda Arnold) didn't have quite this association in mind, they did claim that emotions are "cognitive" appraisals which (although not necessarily deliberative in their actual unfolding) are still in some sense "reasonable" judgments—the perception of an object isn't enough to elicit an emotion; an intervening appraisal of the object's significance is also required to bring about the emotion. Perhaps the cognitivist orientation towards emotions can be traced to Aristotle and his notion that the expression of emotions reveals the kind of character we have (the (in)appropriateness of an emotional response in a particular context reveals our powers of judgment—i.e., our dispositions (hexis) to judge); however, the Aristotelian tradition needs to be distinguished from early cognitivist theories (such as Arnold's; see below), and both of these traditions need to be distinguished from the contemporary version of the Cognitivist Program.

The modern rehabilitation of the Aristotelian/Stoical view of emotions-as-judgments begins with Anthony Kenny (1963). Many related "propositional attitude" views of emotions have followed Kenny's lead in attempting to give conceptual analyses of what the emotions are. However, the propositional attitude approach has numerous problems, as Griffiths argues (1997, Chapter 2), to which the interested reader is referred. Griffiths' general conclusion is that this particular brand of a priori conceptual analysis is impoverished since it does not take seriously current scientific attempts to illuminate just what the emotions are—a lesson that ought to be taken seriously by philosophers still overly preoccupied with a priori approaches. While I will not discuss the Aristotelian/Stoical view of emotions, nor its propositional attitude inheritors, it is worthwhile to note that this tradition is distinct from early cognitivist theories (rooted in Magda Arnold's work), since the former holds that emotions involve higher-order judgments which are far more "cognitive" compared to the latter's softened notion of "cognitive." For Arnold,
appraisals are not really intellectual judgments, but are primarily sense-judgments that accord with the Darwinian and Jamesian view of emotions as survival-related responses. What Arnold adds to the Darwinian and Jamesian accounts is the idea that Jamesian bodily perceptions do not suffice, as the salience of certain objects over others in eliciting an emotion requires the concept of appraisal (to make sense of "salience"). Her working definition of an emotion, accordingly, is that an emotion is a felt tendency toward whatever is appraised as "good" and away from whatever is appraised as "bad," and in this light emotions are "reasonable" strategies for coping/surviving in the world.

An Overview of the Argument

If emotions-as-appraisals are strategies for coping, are these strategies learned, modifiable, "innate," and so on? The meaning of "strategies for coping" goes to the heart of the Cognitivist Program and the way in which it moves beyond both the Aristotelian approach and early cognitive theories by inquiring into the degree that appraisals are weighted as more intellectual or more sensory judgments. The modern Cognitivist Program begins with this issue in a famous debate between R.B. Zajonc and R.S. Lazarus about the "cognitive" status of emotions. However, before covering the essence of the debate and its importance to some of the central features of the Cognitivist Program, first I will outline the general architecture of the program via a discussion of the dimensional approach, and the manner in which the Cognitivist Program appropriates the dimensional framework while moving beyond it. By first expounding the general framework, the debate between Lazarus and Zajonc and its contribution to understanding appraisals can better be seen.

What is "Dimensional" About Dimensional Approaches?

Recall from Chapter 3 that dimensional approaches use axes representing the major parameters for classifying emotive phenomena (e.g., a two-dimensional framework with a
positive-negative valence axis, and an intensity axis). Comparatively speaking, dimensional approaches may vary in the number of axes, and may use different types of axes with differing foci (focusing on individual or social categorizations of heterogeneous emotion types). These approaches are dimensional in part because of the shared commitment to the idea that emotions are heterogeneous; associated with heterogeneity is the additional commitment to distinguishing the experience of emotions from the representational schemes tracking experience. Whatever “experience” is taken to mean, it operates in the background guiding the formation of various dimensional accounts. By contrast, the representation’s axes and the manner in which the representation accommodates a heterogeneous range of emotional phenomena are made explicit.

Broadly speaking, the background intuitions about emotive “experience” hold that emotional phenomena can be general (e.g., mood, temperament) or specific (rage, guilt); can involve an object (the object of one’s love) or not (a diffuse sense of anxiety); and can naturally be expressed as “bipolar pairs (happy-sad)” or not “(tension, outrage, terror, and agitation all seem without an exact opposite, or perhaps to share the same opposite of calmness)” (Russell and Lemay 2000, p.495). These intuitions, which undergird the generation of dimensional representations, are theoretically framed by eight characteristics (Russell and Lemay 2000)—characteristics that stem from reflection upon the field of experience and the relation of vernacular terms to this field.

The first characteristic is that emotion-representations have “fuzzy boundaries.” Fuzziness may issue from cultural language games whose terms, from a cross-cultural perspective, seem comparable and yet are still difficult to categorize in a single representational scheme; fuzziness may also issue from terms that are linked together within one “form of life” by family resemblance relations (or other categorization mechanisms), which would preclude precise boundaries for a family of emotion-terms. Closely related is the second characteristic of “typicality”: prototypical examples anchor the graded extensions of a category. For example,
Ekman's prototypical features of facial expressions anchor gradations of membership (although as alluded to in Chapter 3, there are still “boundary effects”; see “Discrete Approaches Versus Dimensional Approaches”). The third characteristic claims that emotion categories are nested in a “fuzzy hierarchy.” The fuzzy categories form family resemblance links between other categories, where the highest categories are general (like mood, temperament, etc.), the middle categories are particular prototypical emotions (like love, anger, etc.), and the lowest categories are “subspecies” of middle-level categories (like puppy love, parental love, outrage, etc.).

The fourth and fifth characteristics are “dimensionality” and “inter-category structure.” The fourth characteristic holds that for an empirically collected range of terms, the task is to orient the terms with appropriate bipolar parameters (e.g., parameters like positive-negative valence would be general enough to include particular bipolarities like happy-sad). And even for terms not easily given a bipolar representation (such as tension, outrage, terror, and agitation as mentioned above), the task is still to examine the extent to which these terms have an opposite term (calmness); the point of other parameters like intensity is to capture this vaguer type of “bipolarity” (tension has high intensity, whereas calmness has low intensity). The fifth characteristic of inter-category structure claims there ought to be systematic relationships between emotion categories and the dimensional representation; in other words, the systematic relationships ought to be empirically established, with proper statistical analyses done on those results. This means the dimensional representation fits well with a practice’s associated emotional “experiences.” A further implication of inter-category structure, according with the heterogeneity of emotions and the diverse range of systematic relationships, is that any particular

---

1 Note that this is not a classical hierarchy of strict inclusion. Fuzzy hierarchies differ in the following two interconnected ways: first, in breath, as some emotional categories like unhappy “are quite broad (and hence appear in the upper regions of the hierarchy), whereas others, such as outrage, are quite narrow (and hence appear in the lower regions)”; and second, in degree of overlap, as, for example, “most but perhaps not all instances of outrage are also instances of anger; and most but perhaps not all instances of anger in turn are instances of unhappiness” (p.497).
instance of a category does not fall exclusively under one category, but may fall under a number of different categories (e.g., parental love not only falls under the higher categories, but as a complex emotion may intersect with a number of other categories like anger, as an expression of parental love).

The sixth, seventh, and eighth features are “scripts,” “theory-embeddedness,” and the “relativity of categories.” Scripts are the mechanism by which prototypical examples of emotion categories within a cultural milieu operate to produce the emotional experience; a script is a “prototypical sequence of causally connected and temporally ordered events” (p.496). The deployment of scripts in a folk context, in turn, expresses the theory-embeddedness of emotion categories: scripts in a cultural milieu require an appeal to the folk psychological categories of beliefs, desires, and so on. And lastly, the relativity of categories means that categories differ relative to historical periods, socio-cultural factors, etc., so it is important not to reify the particular dimensional representation being employed, as one must be sensitive to the semantic field’s configuration of terms in a particular social matrix.

Dimensional approaches, as the above eight characteristics disclose, make much stronger use of human-science concepts compared to the functional approaches in chapters three and four. In other words, dimensional analyses of “fields of experience” are more sensitive to phenomenological considerations regarding cultural practices, and how best to represent some of the subtleties within a practice, as well as differences between practices.

Linking the Cognitivist Program to Dimensional Approaches

These eight characteristics exhibit a higher-order orientation with a strong phenomenological flavor; in particular, they are about the meta-language by which the dimensional approach represents emotional experiences. These meta-scientific conceptual tools assume a particular view of categorization (which correlate with experience); they also assume
that the most appropriate means for representing emotion-categories is through dimensional representations. Cognitivists have a similar orientation, where they primarily examine the situations eliciting emotional appraisals, and differentiate emotions by way of the (graded) activation of the parameters in the dimensional space. There is one important and subtle difference, though, between dimensional theories and appraisal theories. Dimensionalists, to repeat, are careful to separate “the experience” of an emotion (inferred via testimony, questionnaires, etc.) from the representation of “that experience” (the explicit dimensional representations employing data from testimony, etc.). Conservative dimensionalists claim that their representational schemes only describe certain emotional phenomena (relative to any of the eight characteristics built into their representations); less conservative dimensionalists go further in claiming that their approach is not merely descriptive, but also prescriptive: experimental evidence seems to indicate that emotions really do have blurry boundaries; cross-cultural studies seem to indicate that the English use of “emotion” is itself a folk concept that has particular culture-bound associations, as apparently there are other cultures where the linguistic practices do not square with the semantic constraints of “emotion”\(^2\); and as a result of “descriptive” considerations like these, when examining culturally-bound uses of “emotion” concepts, investigators ought to be sensitive to how representations are set up and that the concepts may differ significantly from English-based representations (see Russell and Lemay 2000).

Cognitivists are bolder than dimensional theorists in that they commit what John Dewey calls “the philosopher’s fallacy,” which roughly is the reification of the representational scheme one employs to understand a process (i.e., the representation is taken for the process itself). To illustrate the “fallacy,” I start with a contrast: firstly, appraisal/cognitive theories are different

\(^2\) For example, in Chinese philosophy people are thoroughly affective. When English speakers say “he’s emotional” an implicit contrast is made (e.g., “he’s rational” or “he’s cool as a cucumber”); whenever work is being done by “emotion,” it would be prudent to distinguish this folk concept from the relevant semantic field in Chinese philosophy. I discuss this general point in Chapter 6.
from discrete approaches since discrete approaches argue for a finite number of separate basic emotions (e.g., affect programs), where each emotion has rather complex machinery; cognitivists instead claim that there are a potentially infinite number of emotions, as continuous variation in the parameters of the dimensional space can give rise to any number of emotions.³ They also claim that emotions-as-appraisals are structurally “simpler” than basic emotions because appraisals are just “parameterizations” in the dimensional representation. It could be objected that since there is room for different types of axes, different numbers of axes, and so forth, representations aren’t “real”—i.e., they don’t delineate “genuine” phenomena, as the phenomena may just be artifacts of the representation, where changing the parameters may “erase” the phenomena. Cognitivists respond by claiming that while researchers have produced axes that differ in number and type, upon closer inspection the representations and their associated empirical results actually show striking overall similarities to one another (Ellsworth and Scherer 2003); apparently these representations have some claim to projectibility.⁴

Secondly, cognitivists also contrast themselves with dimensional theorists, as they think that dimensional approaches focus too much on subjective experience and the relation of such experience to folk language; cognitivists emphasize that folk language is too crude and narrow to capture the richness of emotions. And while dimensional approaches would be able to express a potentially infinite number of emotions, the framework is too descriptive for cognitivists, whether the theorists are conservative or “less conservative” dimensionalists. Beyond description, explanation is desired, and according to the Cognitivist Program that requires a causal account.⁵

³ In Humean terms, in principle there is no “missing shade of blue.”
⁴ Despite differing dimensions, some of the major agreed upon “appraisal-dimensions” are novelty, goal-conduciveness, coping potential, and norm-compatibility. These are actually parameters situated within broader dimensions such as valence and intensity.
⁵ Scripts mention causality, but are not causal enough since the neural-physiological connection is not explicitly made. The causal links in scripts appear to involve psychological events or folk states that are merely “linked” to one another in a temporal ordering.
In order to give such an account, first one must be careful not to interpret an appraisal as merely an antecedent condition to an emotion, where an appraisal of a percept is a necessary precondition for an "ensuing" emotion. The problem with this view (a view that early cognitive theorists held) is that it is too simplistic an account of how appraisals relate to emotions because it leaves out the explicit causal relation between appraisals and emotions. The modern view of appraisals holds that, in similar fashion to evolutionary psychology, appraisals are "modes" of operation within the "cognitive" architecture (i.e., the arena of mind-neural-physiological happenings), and are also modes/components of "emotive" processes. More precisely, cognitivists claim (as do evolutionary psychologists) that there is no strong distinction between "emotion" and "cognition"; these modes are really emotive-cognitive processes. The locution "emotions are appraisals" is not an identity statement; rather it conceptually expresses the crucial role played by appraisals in emotive-cognitive processes, where the proper way to understand appraisals as "components" in these systems is by viewing them as modes of operation.

**Appraisals as Modes (Systems) of Activation**

Modes are expressed in the dimensional representation by the particular manner of configuration of the parameters—it is a structural pattern of (causal) activation. As structural patterns of activation, appraisals (see fn.4) are not merely necessary preconditions for ensuing emotions. Thus the reification ("fallacy") of the dimensional schemes used by cognitivists is actually part and parcel of the way in which their project moves beyond the dimensional approaches, since the explanatory element of the Cognitivist Program is linked to the claim that the particular appraisal-modes are causally responsible for how emotions arise; parameters like valence and intensity refer to real somatic-neural processes, and the dimensional representation in turn is interpreted as an "activation space" tracking those processes, some of which give rise to our subjective experiences of emotions. Hence the dimensional models used by the Cognitivist
Program are semi-concrete in the way they relate to our phenomenological experiences and in the way they refer to causal somatic-neural processes; and the (causal) models also express highly abstract capacities by virtue of the quasi-mathematical notion of a dimensional activation space.6

Lastly, observe that appraisals are systems: appraisals are dynamical structures whose elements—the particular "values" of parameters like valence and intensity—are situated in relation to each other—the parameters are activated in a particular configuration—and to the system as a whole—the overall "emotion" enacted by the modal configuration.

The Debate Between Zajonc and Lazarus

The notion that this activation space is causally linked to our subjective experience of emotions raises the following question: if our emotional experiences are usually felt, conscious experiences, does this also mean that appraisal processes are felt, conscious processes? For the former, appraisal processes are not necessarily felt, as appraisals are "components" of emotive-cognitive systems and not the felt aspects of emotions. For the latter, conscious awareness of the appraisal requires examination of two further issues: what is meant by "cognitive" appraisals? Are they required to be conscious and at least semi-deliberative? And secondly, what is really at stake in claiming that appraisals are cognitive? The first issue brings us back to the roots of the modern Cognitivist Program and the debate between Zajonc and Lazarus; the second issue motivates the need for a developmental account.

The two central cognitivist ideas—there aren't a basic and finite set of emotions, and the experience of emotions is a continuous process—can be attributed to Lazarus. Lazarus generally argues that emotions are inferences based on "core relational themes" (such as "a demeaning

6 I discuss fidelity-related concerns in section 5.2 after expositing the broader philosophical framework of DST. In brief, fidelity has to do with the reification of the dimensional representation and the claim that its explanatory value is linked to its causal nature.
offense against oneself' in the case of anger). Zajonc, by contrast, argues that emotions don’t require those inferential processes; the emotion system and the cognitive system are (semi) independent since emotional judgments can operate without the standard marks of the cognitive system’s operations. Zajonc’s conception of an emotion system is akin to the operation of affect programs, although for Zajonc emotions are less narrowly defined, as they are judgments that implicate the self. The debate has often been portrayed as one with each person talking past the other since they each have different conceptions of “cognitive,” and so the dispute is really over the semantics of this term. The more socialized ideas that Lazarus appeals to in his core relational themes are grounded in notions like “motivational relevance,” “other-accountability,” etc.; whereas the types of emotions Zajonc focuses on are affective reactions, where the appraisal component of these reactions are about more “primitive” notions such as “liking,” “disliking,” “preference,” etc. So on the portrayal that the dispute is about semantics, both sides are right, depending on what is meant by “cognitive”: Zajonc is right that certain affective responses can operate disjointedly from the “cognitive” (i.e., conscious and semi-deliberative) system while still involving judgments, but Lazarus is also right that many higher-order notions about emotions do not allow for a meaningful division between the emotion system and the cognitive system.

Appropriating Zajonc and Lazarus into the Cognitivist Program

Extracting from the debate, the modern Cognitivist Program claims that appraisals may be conscious as well as unconscious, and automatic and well as deliberative; all that matters in the end is that the parameters in the dimensional representation be able to accommodate these types of appraisal processes. The above interpretation that the debate is over different senses of “cognitive” is partly correct, but what is really at stake are the “sorts of information processing involved in the production of emotional responses” (Griffiths 1997, p.25). The shift from “cognitive” to “information processing” is significant since the former term, used widely and
loosely in psychology, is traditionally associated with problem-solving tasks; and in philosophy, the former term's association is with propositional attitude theories. Information processing sidesteps these narrow associations by focusing on modes of operation; similar to evolutionary psychology's concern with more centralized systems, the modern Cognitivist Program is concerned with understanding a range of emotive-cognitive systems. As a result of this shift to emotive-cognitive modes, the Zajonc-Lazarus debate discloses (from the perspective of the modern Cognitivist Program) that the dispute is really over what representational language to employ: from an information processing perspective, the issue for Zajonc is really about states, while the issue for Lazarus is really about experiences.

The view that emotions are states means that emotions are specific patterns of embodied activity, where these states normally occur prior to conscious awareness—they are more peripheral and less centralized. The states are basically affect programs associated with specific neural-physiological changes, but a cognitivist additionally characterizes states as "response tendencies" associated with specific judgments implicating the organism's well-being, making cognitivist states wider than affect programs. Zajonc's view differs from Lazarus' view because emotional states need not involve the types of inferences that usually bear upon our emotional experiences. That is, emotional states may occur without direct awareness, we may for whatever reason deny these states, or we may just ignore them; the multiple modes of emotional systems allow for these possibilities. While Lazarus would also grant some of these possibilities, his point is that "disjointed" modes of operation for emotional states focus on the wrong aspect of emotion systems. Lazarus' emphasis on appraisals-as-experiential holds, by contrast, that states make no sense without looking at how these putative entities relate to coping in the socialized world—these concerns indicate that appraisals are more centralized systems. The shift in emphasis from (emotion) modes-as-states to (emotion) modes-for-coping is really about the
difference between a more synchronic orientation for the former, and a more
diachronic/developmental orientation for the latter, where development is about some of the
interactive relations between organism and environment.

The “Continuum” Between Discrete and Dimensional Approaches

Ekman’s work on affect programs acknowledges developmental considerations, but his view holds that core, discrete emotions constrain and organize what other affects develop—i.e., they form the “scaffolding” upon which display rules operate. Moving up a level, Zajonc occupies a middle position between Ekman and Lazarus—a position with more human-scientific concerns—where most importantly, emotional modes-as-states involve judgments that implicate the self (the well-being of the organism). The modern Cognitivist Program adds to the modes-as-states view the idea that states are represented by a (continuous) dimensional framework, whence states have a potentially infinite number of gradations. In turn, as states change through time, Lazarus’ view comes to the fore—a view explicitly referencing human science concepts—where emotive-cognitive modes-for-coping track the development of appraisals and the formation of dispositions, moods, etc. Developmental concerns raise two questions: if the Cognitivist Program emphasizes individual appraisals that serve socialized functions, what role does communication play in the formation of appraisals? And secondly, is there a framework for thinking about development that fits with the Cognitivist Program? I explore the first question by discussing Robert Frank’s “commitment model” of emotions; I address the second question in section 5.2.

The Tacit Communicative Dimension of Appraisals

The tacit dimension of individual appraisals is that they often occur in relation to communicative signals. Cognitivists do not account for this parameter explicitly in their models.

---

7 This isn’t to say that the Cognitivist Program denies the importance of “basic emotions,” for as Ellsworth and Scherer note, “there still may be magnetic regions in this [dimensional] space, perhaps named regions, that attract ambiguous emotions and are salient in folk psychology” [emphases mine] (2003, p.589).
because the study of communicative signals is a study all its own; that added complexity would detract from the primary aim of providing representations for (the multi-dimensional aspects of) appraisals in emotions. Still, given that communication is presupposed in many emotional judgments (e.g., the core relational theme of anger is a demeaning offense against oneself), it would be worthwhile to explore an extension of the Cognitivist Program that focuses on how emotions are embedded in these larger concerns with communicative signals. Robert Frank, in *Passions Within Reason*, gives one of the most promising theories of the communicative dimension of emotions (and thus what qualifies them as "higher-order" emotions). (Most importantly, Frank's account has an evolutionary orientation, which fits with the tenor of my overall project.) As I think Griffiths accurately notes, on Frank's account an emotion can be characterized as an "irruptive pattern of motivation" (Griffiths 1997, p.120). There is an ambiguity, though, in the interpretation of an "irruptive pattern": is this pattern more akin to a state (*a la* Zajonc) or more akin to an experience (*a la* Lazarus)?

**Emotions and the Commitment Problem**

On the first interpretation, which Griffiths attributes to Frank, these irruptive patterns of motivation are the effects of affect programs (or their related extensions). Frank "suggests that the conscious affect (feeling) associated with emotion [affect program] acts as an internal source of reinforcement for behavior. Behaviors which would be reinforced by external rewards are punished by these internal reinforcers, and vice versa" (p.121). In this passage Griffiths is referencing Frank's appropriation of Darwin's work on the (facial) expression of emotion. For Frank, the modern significance of Darwin's work is that emotions serve both "external" and "internal" functions—the internal functions deal with the different sentiments one feels towards certain courses of action, and the external functions deal with the reception of these possible actions. To illustrate, consider a person who is tempted to cheat a business partner by skimming
off some of the profits in secret. Suppose that the “internal sentiment” of greed\(^8\) is felt, where engaging in cheating would be reinforced by the external reward of cash and its associated sentiments (e.g., glee, satiating a hoarding urge, and so on). Let us also suppose that the rewards from cheating are considerable enough, so that the short-term gain from cheating gives rise to a serious degree of temptation. The question posed by Frank is, why would an agent want to be honest when the incentives favor cheating? This problem, in brief, is the “commitment problem”—i.e., what makes it possible for an agent to commit to being honest in the face of short-term gain?\(^9\)

**Emotions as Game-Theoretic Equilibria**

To be more accurate, the commitment problem stems from the acknowledgement that in fact people enter into business with one another without engaging in cheating; yet if opportunities for cheating present themselves, as they invariably do, what makes it possible for people to commit to cooperative ventures *in the first place*? There are two approaches to the commitment problem discussed by Frank; the first (alluded to above) is the “sincere-manner account,” where honesty is the mechanism postulated in accounting for commitments to cooperative ventures.

This isn’t to say that temptation or cheating won’t occur; it is just that the *internal* “sentiment” of honesty functions to enable the very possibility of cooperation in the first place.\(^10\) Thus if one is

---

8. Note that the language of “sentiments” stems from Frank’s appropriation of the moral sentiments tradition (of Hutcheson, Hume, and Adam Smith). Accordingly, greed should not be interpreted as a “felt sentiment” (colloquially speaking); rather emotive-cognitive sentiments (like greed) serve specific *game-theoretic* roles, as I expound in the following sections.

9. Note the problem’s homology with the problem in biology concerning the evolution of altruism. In essence both are Prisoner’s Dilemma problems. Commitment problems concern the possible *origins* of moral sentiments, as examined from an evolutionary perspective. Technically, commitment problems concern evolutionary game-theoretic equilibria—these convergence points help to explain why certain behaviors (like altruism) are evolutionarily possible.

10. Honesty, as Frank recognizes, is a character trait/disposition. The moral sentiment of sympathy assists in configuring this disposition, whence honesty is a “higher-order” moral sentiment. This obviously stretches the meaning of “sentiment” (even when read in light of the moral sentiments tradition). Character traits (like honesty and prudence), though, are still emotive-cognitive modes (remember that “emotion” is a term of art); see fn.11, and section 9.5.1.
honest, the feeling of guilt felt by cheating (or the prospect of cheating) serves as an internal source of reinforcement for honest behavior. Cheating behaviors that would be reinforced by external rewards such as cash, for example, are punished by internal reinforcers like guilt; additionally, the "externality" of getting caught should one cheat carries the associated burden of shame.\(^\text{11}\)

By contrast, the other approach to the commitment problem is expedient in flavor. It postulates that the reason why cooperation can get off the ground is because of the internal sentiment of prudence (not honesty)—one is worried about one's reputation, especially if one is caught when cheating (the fear of getting caught). So if a person could cheat without the prospect of getting caught, it would be prudent to engage in cheating; but since cooperating with others is needed—and in order to do that one's reputation must be cooperation-worthy—one cannot cheat when the chance of getting caught (relative to the "amount" of short-term gain) exceeds the value of sacrificing reputation. For this "reputation account," the less-than-moral sentiment of prudence is postulated to solve the commitment problem (unlike the mechanism of honesty, which carries a nobler air for why we cooperate).

The point of these two approaches is that they are not mutually exclusive, as they both have plausibility depending on circumstances; but more importantly, whether we postulate a sincere-manner account or a reputation account, commitment problems are solved by the same kinds of emotions/sentiments. Thus both accounts can appeal to emotions like guilt and fear (as

---

\(^{11}\) This brief account may appear circular, since being honest makes certain commitments, and so the commitment problem would be "solved" because these commitments are made. Actually the postulation of honesty as a mechanism for solving commitment problems concerns the possibility that honesty may be an evolutionary game-theoretic equilibrium point, just as (biological) altruism is postulated to solve long-term cooperative problems regarding the origins of such behavior. In the latter case the circularity charge is inappropriate, since we are concerned with evolutionary (iterated) games, not one-shot games. Frank adopts a similar game-theoretical account for the evolutionary roles of "sentiments" (like honesty and prudence) in solving commitment problems. From this perspective, honesty—as a "higher-order" sentiment—taps into emotions like guilt; the resulting (modal) emotive-cognitive configuration regards the possible evolutionary origins of such a complex moral sentiment. (These observations also hold for the following discussion of prudence and the reputation account.)
they bear upon the moral sentiments of honesty and prudence, although each account would preferentially assign more weight to certain emotions. Getting back to Griffiths' first interpretation of Frank (see “The Tacit Communicative Dimension of Appraisals” above), the best-fitting research program is Ekman's approach (and its extensions). Accordingly, the way to read the claim that emotions are “irruptive patterns of motivation” is in relation to solving commitment problems, since their irruptiveness (i.e., affect programs are mandatory, fast, etc.) makes them difficult to fake—emotions solve commitment problems since they are genuine signals. This additionally means that potential cheaters are more easily detected (e.g., love, as it relates to the commitment problem of marriage, is difficult to wholly fake without any of love’s typical patterns of feeling, motivation, and behavior—especially facial expression).

The Communicative Dimension of Game-Theoretic Equilibria (“Strategies”)

I think Griffiths is largely right in claiming that the first interpretation squares with much of Frank’s presentation. However, Frank situates his view of emotions in a larger framework where emotions are (game-theoretic) strategies for solving commitment problems; they are interpretations of social signals to engage in cooperative behavior. Only if emotions are viewed as narrow states is the affect program interpretation justified. But it seems to me that this interpretation is not definitive since the emphasis in Frank's work is on the communicative dimension of emotions, where emotive-cognitive dispositions only bear upon the commitment problem if others consistently read these behaviors as signals (cf. Frank 1988, p.64). Because Frank is not clear whether emotions must be states, a better interpretation would be that if emotions are states, they are more akin to the cognitivist’s view of emotions as modes—in particular, the modes-as-states aspect of emotion systems.

---

12 Even though guilt, honesty, prudence, and reputation are not really affect programs, some of the key features of affect programs signaled by these sentiments are what Frank emphasizes.
There is a second interpretation to Frank’s work that is consistent with the view of emotions as (experiential) modes-for-coping (see “The Tacit Communicative Dimension of Appraisals” above). If emotions are still irruptive patterns of motivation, they may be “additional pieces of mechanism. These need not be related to the known affect programs” (Griffiths 1997, p.121), since these additional pieces of mechanism may be quite separate from affect programs, or may be only indirectly related to these programs (e.g., the mechanism developmentally relies on these programs, but acquires additional pieces such that the mechanism when “mature” is significantly different in form). This second interpretation creates a space for emotional signals as modes for social coping, especially modes for engaging in cooperation. As Frank notes, although some emotions show evidence of being affect programs, even if they (and other emotions) “were transmitted only by cultural indoctrination, they would serve equally well [in solving commitment problems]” (p.11). The metaphor that captures his view of the strategic role of emotions is to “imagine that nature’s role is to have endowed us with a capacity that is much like a gyroscope at rest; and that culture’s role is to spin it and establish its orientation. Each of these roles is indispensable” (p.65), since culture must establish some of the parameters for interpreting signals as social signals, and since nature sets up irruptive patterns of motivation that are difficult to fake (the ability to express genuine signals).

Both the first and second interpretations are actually about the communicative dimension of emotions; and whether we view emotions as states or as modes for coping, the lesson is that the latent dimension of appraisal operates in assessing signals in their social capacity. But if the roles of “culture and nature” are both indispensable for signaling processes, the question remains how the parameters for interpreting signals develop through time. The temporal emphasis of the Cognitivist Program assumes that development occurs, but “development” can be understood in any number of ways. For example, dimensionalists start with individuals whose emotions are
already developmentally (and ontogenetically) mature, and a dimensional representation then
describes the usage of emotion-terms by individuals within a culture\textsuperscript{13}; cognitivists start with the
assumption that ontogenetic development happens in the dimensional activation space, and while
there can be an indefinite number of developmental-emotional processes (as causal), it is still
acknowledged that emotion terms tend to cluster and to form fuzzy categories. Adding a further
twist to the issue of development, developmental cognitivists like L. Alan Sroufe (1995) hold that
it really makes no sense to speak of a “morphology” of emotions. It is misleading to speak of
developmental “outcomes,” and even if there are fuzzy emotional categories, this occurs at the
price of failing to emphasize that mature emotions do not “appear”; they develop and continue to
develop “even after the emergence of mature expression” (p.65). What we need is a framework
for these various developmental considerations forming the backbone to the Cognitivist
Program’s temporal orientation.

5.2: Developmental Systems Theory and the Strategic Stance

There is a latent problem in the above reconstruction of Frank’s account and its fit with
the Cognitivist Program. The problem is that if emotions are hard to fake due to characteristics
associated with modes-as-states (e.g., mandatoriness), and where because of such “hardness-to-
fake” they have conferred on them the status of being genuine signals, this appears not to square
with the second interpretation of emotions as modes-for-coping. The experiential aspect of
modes-for-coping means that there are legitimate modes of emotional experience which are
conscious and “deliberative” (i.e., the appraisal component of an experience is deliberative); but
if appraisals are deliberative, are the relevant emotions still genuine signals? Granted there are

\textsuperscript{13}This is not wholly true, as dimensionalists acknowledge the need to integrate a more explicit
developmental-ontogenetic account with dimensional representations. However, their emphasis is on the
latter, since they want to track cultural language games and the games’ relations to types of experiences.
ways in which "deliberative" modes-for-coping could display marks of genuineness (e.g. if one is a good actor; if one is a "master rationalizer"—consciously or not—and makes a piece of deliberation "real"; if one is pathological; etc.), at least with respect to "normal" behavior the problem remains. The answer to this problem involves amending Frank's account, which consequently sets up the need for a broader perspective provided by DST.

Distinguishing "Static" Equilibria and Evolutionary Equilibria

The amendment stems from a confusion in the commitment problem. Frank's commitment problem assumes a distinction between self-interested rationality and commitment rationality (in Prisoner's Dilemma terms, it is the distinction between short-term gains from defecting and long-term gains from cooperating). Theoretically, this means a mathematical "agent" has these two forms of rationality, and is a different type of rational agent depending on which rationality is adopted (Ross and Dumouchel 2004, p.282); in other words, given that the theoretical foundation of Frank's work is game theory, the agent is a different type of rational agent relative to playing two different strategies within a game. However, mathematically speaking this is confused, since there aren't conflicting preferences (with resulting payoffs) within an agent; there aren't two rationalities "vying for control" within the individual. Rather what is really happening is that for such conflicting preferences, the conflict would actually be between two agents: one sees the situation as a one-shot Prisoner's Dilemma (PD) game, whereas the other sees the game as an iterated PD game, in which case there are simply different preference structures associated with separate games (p.255).

The shift is subtle but important. In game-theoretic terms, it is muddled to claim that an agent simultaneously plays two games, a classical PD game (one-shot) with an optimal payoff to defect (which would make commitment rationality an irrational choice for that agent), and an evolutionary PD game (iterated) with an optimal payoff to cooperate. The problem with the
picture of two rationalities vying for control within an agent is that if one is classically rational then one is evolutionarily irrational, and vice versa. But (classical) agents don't really play evolutionary games; *strategies* are the units of evolutionary game theory, and as such, strictly speaking, strategies play these games.\(^{14}\) Strategies are “external” to (classical) agents in the sense that they are “meta-games” played over the games that agents engage in—the games that (classical) agents play form the “base-level” upon which various coordination games and their higher-order equilibrium points build. Metaphorically, if classical agents have local basins of attraction, evolutionary strategies explore possible landscapes (over attractive and non-attractive basins) in the search for macro-basins (see fn.15 below).

**Summing Up**

To take stock, *if* emotions are viewed from the standpoint of modes-as-states, the general picture that Frank portrays still makes good sense, as emotions are still commitment devices (since they are hard to fake). However, it is theoretically misleading to claim that there are two rationalities within an agent who encounters a commitment problem.\(^{15}\) The revised theoretical backdrop views strategies as “external” to (classical) agents, where emotions are commitment devices that shift the “strategy horizon” by drawing attention to a *restricted* range of meta-basins (i.e., emotions function to narrow the range of possible Nash equilibria (p.282)—and thus

---

\(^{14}\) As Ross and Dumouchel note, “there’s simply no mathematical sense to be made of the idea of a solution concept for games according to which agents can behave ‘irrationally’ in equilibrium” (p.257). Thus classical agents are distinguished from strategic agents. Saying that “strategies play these games” is similar to claiming that functors (“second-order agents”), ranging over functions (“first-order, classical agents”), delineate second-order mappings.

\(^{15}\) Note that this is a game-theoretic point. It is left open whether the *psychological* mechanism is one of conflicting “voices” (e.g., a short-term “instinct” versus a long-term instinct) or otherwise. Indeed, the previous two interpretations provided above—see “The Tacit Dimension of Appraisals” ff.—indicate that the general game-theoretic account, as it relates to “actual” humans, is compatible with either a mostly unreflective process (Griffiths’ interpretation of Frank that emotions are *irruptive* patterns of motivation—what I have called the “modes-as-states” interpretation), or a more culturally informed and deliberative process (the second interpretation that emotions are “modes-for-coping”—recall that even if emotions “were transmitted only by cultural indoctrination, they would serve equally well in solving commitment problems”).
coordination games—that can be played). Additionally, this revised view fits with the claim that emotions may be modes-for-coping, because even if the modes are conscious and deliberative processes (e.g., sympathetic engagement), these emotions can still function as commitment devices which enable cooperation “for the sake of welfare enhancement in social interactions” (p.282). With the emphasis on strategies, emotions become “externalized judgments”: they are signals according with the conventions that govern them, where the conventions’ associated punishments make the signals genuine (thus conventions play a role similar to the hard-to-fake condition).

Overview of the Remaining Argument

While I’m getting slightly ahead of myself with the above discussion of emotions as modes-for-coping and the way in which the external (socialized) aspects of emotions are highlighted, the reason for this discussion is that it provides a bridge between the Cognitivist Program and the Social Constructivist Program (Chapter 6)—both of which utilize appraisals. Getting back to the Cognitivist Program, since it acknowledges communication and the “external” role of social structures but emphasizes individual appraisals (whether they be about states or coping), there are two remaining things I discuss in this section. First I give a brief overview of DST to show that it accommodates and broadens the central temporal/developmental orientation assumed in the Cognitivist Program; and secondly I outline the strategic stance, which captures the features of the Cognitivist Program and its extension from a meta-level view. Lastly I situate the strategic stance within our four-dimensional framework. (The reader should keep in mind that DST’s general framework is more “dialectical” compared to the strategic stance. See fn.18 below; I save fuller discussion of these topics for Chapter 6.)

Game theory underpins both Frank’s account and the above amendment to his account. Game theory is indirectly related to the Cognitivist Program via the latter’s quasi-mathematical
dimensional representations. In order to make the link between the two clear, the logic will go as follows: DST is schematically represented by a set of coupled differential equations interrelating organism and environment (and the mutual changes on both); the Cognitivist Program's dimensional representations track changing modes in an activation space, and while explicit "mathematical functions" governing these changes are not provided, the general idea of a set of coupled differential equations that express the shifting "topology" of the activation space coheres with the characterization of the program in section 5.1; if DST accommodates the representations used by the Cognitivist Program, what captures the relevant features of DST from a meta-scientific level? The strategic stance does the job, where the connection between game theory and (aspects of) DST occurs through the general recognition that evolutionary game theory (which encompasses classical game theory) is really just a branch of dynamical systems; conceptually speaking, the dimensional activation space with its "differential equations" can be organized by the strategic stance.

**Developmental Systems Theory**

DST is not really a "theory," rather it is primarily a (philosophical) framework for approaching empirical problems and their associated causal models. Thus a "central theme of the [DST] research tradition has been that distinctions between classes of developmental resource(s) should be fluid and justified by particular research interests, rather than built into the basic framework of biological thought" (Oyama et al. 2001, p.206). These particular research interests generate particular causal models for how some developmental process occurs, while the general framework is highly abstract in that the "unit of both development and evolution is the developmental system, the entire matrix of interactions involved in a life cycle" (*ibid*).¹⁶ Thus

---

¹⁶For example, neural network models of *C. elegans* express the general principles of genetic pleiotropy and the multifunctionality of phenotypes; these models build upon molecular studies linking specific genes to various behaviors, and thus are causal in this respect (see Schaffner 1998).
DST is both causal and abstract in attempting to "dialectically" unfold the way in which particular causal models of specific developmental processes are situated within this general framework; these models illuminate certain details of what the framework is about, and in turn the framework helps to further orient future investigations of interactive relations in the “matrix of a life cycle.” Since DST is intended as a biological research program, I will discuss it in greater detail in Chapter 7; for now all I wish to draw attention to are the broader conceptual features framing the matrix of interactions.

Generally, the matrix of interactions is understood by way of a set of coupled differential equations that express the interrelations between organism and environment (Lewontin 2000). More accurately, the differential equations are really organizational placeholders for the idea that there is no legitimate separation between organism and environment. This isn’t to say that these nodes are fully “interpenetrating”; rather the point is that we ought to focus on the shifting relations between these nodes (and that these nodes can change in certain significant respects—e.g. organisms “create” niches). Accordingly, Lewontin gives two coupled equations to (conceptually) express the essence of DST:

\[
\frac{dO}{dt} = f(O, E) \\
\frac{dE}{dt} = g(O, E)
\]

for organism O and environment E. The rate at which the organism changes is a function (f) of how the environment bears upon the organism; and the rate at which the environment changes is a function (g) of how the organism bears upon the environment. Since these are two potentially different functions, it is illegitimate to claim that the organism-environment relation is fully “interpenetrating” as there is an indefinite number of particular relations between the two nodes (expressed by different fs and gs). Most importantly, the equations are not separate; they are coupled equations, which means that the empirical task for researchers is to model “real”
processes in "real" ecosystems that cash out specific fs and gs and their actual interrelations—the particular "solutions" to the coupled equations which reliably track developmental processes.¹⁷

So quite clearly, the orientation of DST is both causal and highly abstract. In turn, issues relating to the fidelity of models trade on both of these characteristics: particular developmental models have high fidelity relative to being causal models; and these models, as they express the philosophical framework of DST, are interpreted to have a high degree of fidelity with respect to illustrating the importance of developmental interactions. (In terms of the reified dimensional representations of the Cognitivist Program, both the particular causal models and the general dimensional orientation are taken to have high fidelity as they "dialectically" explicate one another; see fn.6.)

Some Differences Between DST and the Cognitivist Program

These coupled equations occur relative to a space of possible developmental processes that is not dissimilar to the dimensional activation space employed by cognitivists. Additionally, dimensional representations have causal import, as do the models in DST (expressed by particular fs and gs). There are two conspicuous differences, though, between DST and the Cognitivist Program. The first difference is that the Cognitivist Program brackets communicative (or more generally "environmental") relations in its models, and emphasizes the manner in which appraisals process external information for emotive-cognitive systems. The program also brackets developmental concerns, for although development is the crucial backbone of the causal-dimensional accounts, the focus isn’t on organism-environment relations but rather on individual appraisals. However, the extension to the program as found in Frank (and the subsequent

¹⁷ Note the use of scare quotes. Developmentalists are realists of sorts, but the emphasis on constructive interactions between O and E shifts the focus from O and E as nodes to the relations between the nodes. That emphasis on change (and the appropriate language for expressing change) is why developmentalists are so sensitive to the terms used (e.g., "real" is not meant to connote a fixed environment, or a fixed genetic-developmental potential for an organism, etc.).
amendment to his account) does take a more "ecological" view where agents are situated in an environment of communicative signals (that is, a strategy space). So while the Cognitivist Program emphasizes the individual's implicit relation to her environment ("dO/dt"), the extension focuses on the environment's dynamics in relation to the individual ("dE/dt"). From a higher-order conceptual perspective, DST's broader dialectical framework can accommodate both of these (less dialectical) approaches. (Note that DST is only viewed as "incommensurable" with these approaches if one privileges (reifies) the dialectical interrelations between the conceptual equation types. However, as noted above, for DST "distinctions between classes of developmental resources should be fluid and justified by particular research interests, rather than built into the basic framework of biological thought." Thus the Cognitivist Program's focus on the individual's implicit relation to her environment—focusing more on "dO/dt" and less on "dE/dt"—and the extension's focus on the environment's dynamics in relation to the individual—focusing more on "dE/dt" and less on "dO/dt"—are compatible with DST's pluralistic and fluid approach to understanding various dialectical processes.)

The second difference is that the Cognitivist Program doesn't use structural equations to map changing modes. But since cognitivists acknowledge there can be areas of "magnetic" attraction that develop within the activation space, an additional piece of machinery consistent with the tenor of the research program would be the conceptual employment of a set of (continuous) "differential equations." What this adds is twofold: it adds the ability to supplement metaphoric images like "magnetic attraction" (and similar talk of fuzzy boundaries, scripts, etc.) with more explicit structural relations, relations that mesh with the causal import of dimensional representations so highly valued by cognitivists; it also fits better with the extension to the program, as the communicative elements of the extension are grounded in evolutionary game
theory—a branch of dynamical systems (whose primary tool is the employment of differential equations).

The Strategic Stance

Hence what we have is the manner in which DST encompasses the Cognitivist Program and its extension, and also one way the program can benefit from DST's additional machinery. The last thing to do is to provide a meta-level conceptual framework for the Cognitivist Program and its extension. The strategic stance does the trick, I claim. For what frames aspects of DST (see fn.18) is the employment of certain dynamical system considerations (as organizational placeholders); and because, generally speaking, these considerations are approachable from the standpoint of evolutionary game theory, a meta-level stance that captures the conceptual features of evolutionary game theory will suffice. 18

The strategic stance, which builds upon the functional stance covered in Chapter 4, looks at relations between "boxes"/agents to delineate patterns of interaction that the functional stance might not so readily discern. The way in which the strategic stance reveals patterns is through two uses of "strategy," a static use and a dynamic use, both of which find technical application in game theory. While the formalism of game theory would be too precise for our purposes, it still assists in characterizing the conceptual features of the strategic stance.

Several ideas briefly need to be put in place before describing the static sense of "strategy." First, "utility" is defined as whatever behavior an agent consistently desires (i.e., would act upon); the tautology "agents act so as to maximize utility" stems from the definition

18 More strictly speaking, the central equation types for population ecology flow from the Lotka-Volterra equations, which are mathematically equivalent to replicator dynamics, the central notion of evolutionary game theory (see Hofbauer and Sigmund 1998). The fs and gs that make biological sense for DST are variations on the Lotka-Volterra equations; it is with respect to these dynamical system considerations that a game-theoretic approach is viable. However, DST is conceptually wider than the restriction to population ecology; in Chapter 6 I give an exposition of how DST's "more dialectical" philosophical framework goes beyond the strategic stance.
The tautology ought to be understood in a dispositional sense, since utility is really "a measure of relative behavioral dispositions given certain consistency assumptions about relations between preferences and choices" (ibid). Secondly, a "game" is defined as all "situations in which at least one agent can only act to maximize his utility though anticipating the responses to his actions by one or more other agents" (ibid). Agents are assumed to be rational, a notion profitably understood in Dennett's qualified sense (cf. section 4.2); agents are "loaned" intentional-system status, whence preferences and choices are consistent relative to the intentional stance. Lastly, a "strategy" is defined as a "predetermined 'programme of play' telling an agent what actions to take in response to every possible strategy other players might use" (ibid); strategies are cashed-out by "expected values," in lieu of specifying a "decision problem."19 Putting the above together, expected values use utility assignments, a game sets up the "static" space of possible payoffs, and strategies give the means to maximize utility. In static game theory, the search for equilibrium points occurs in non-repeated games where the space of possible strategies depends upon the information available to the agents ("perfect" or "imperfect" information); the upshot is that in static games, strategies look at possible dispositions (expected value payoffs) without considering repeated encounters with agents. In short, the picture is a "time-slice," where the problem is finding and classifying solutions relative to the information available.

The move that extends Dennett's intentional stance to strategies in general (strategic or dynamic) involves dynamical games. It has been noted that the future of game theory lies in evolutionary game theory (another way of speaking of dynamical games), since evolutionary

---

19 For technical details, see Myerson (1991).
games capture and extend the basic static game structures. The intentional stance accommodates this dynamical shift through a change of perspective, where strategies are viewed as intentional systems themselves. That is, the units of action in static game theory are ("first-order") agents, and now the units are strategies (systems, in turn, are sets of strategies); give the strategies intentional system status as "second-order agents," then look at the ways in which strategies interact with other strategies.

The strategic stance is just this shift in perspective: it gives intentional system status to strategies (in addition to "first-order" agents) where as an epistemic stance it assists in disclosing the behavior of higher-order systems. Accordingly, whether we are emphasizing the environmental aspects of emotion systems (as with the extension to the Cognitivist Program) by granting intentional system status to those systems in order to comprehend strategic behavior, or whether we are emphasizing the individual aspects of appraisals by granting intentional system status to agents in order to explain why they have the appraisals they do—why it is "rational" for them to have their "beliefs," "desires," etc.—the conclusion is that the strategic stance accommodates the conceptual features of each of these approaches. Additionally, should the emphasis be on development, the stance can also accommodate (intentional) aspects of DST by treating an organism-and-environment system as an intentional system embedded in matrices of interactive relations (strategies within strategies, and so on); this possibility is grounded in the deeper relation between evolutionary game theory and dynamical systems.

5.3: Summary: Relating the Strategic Stance to the Four-Dimensional Framework

20 The founders of modern game theory (von Neumann and Nash) also had dynamical concerns in mind when initially working on static games. See Gintis (2000) on evolutionary game theory and its subsumption of classical, "static" game theory.
In the language of our four-dimensional framework, the strategic stance, with its ability to capture certain conceptual features of DST (those features that encompass the Cognitivist Program and its extension), carves out a subspace that forms a “crescent” above (and overlapping with) the subspace formed by the functional stance. This 4-D crescent is more causally based (in some ways) than the functional approaches, as exhibited by the particular causal models of DST (cf. fn.16) and by the causal reification of dimensional representations—representations that presumably refer to actual neural-physiological processes. And this crescent is also more expressive of a philosophical framework whose models exhibit a strongly abstract flavor, which is clearly seen in DST’s conceptual use of coupled equations, and in the Cognitivist Program’s use of a continuous dimensional activation space that allows for a potentially infinite range of emotions. (Because of this causal-abstract orientation—more causal and more abstract than the functional approaches—the shape is crescent-like; compared with the functional stance’s “half-oval,” the “pinching” of the causal and phenomenological ends and the “bowing” of the more abstract arc yield a crescent. Note that the phenomenological orientation of the Cognitivist Program stems from its appropriation of dimensional representations, which purport to partially map types of emotive “experience”—see “What is ‘Dimensional’ About Dimensional Approaches?” ff.)

Lastly, fidelity-related issues flow from the “causal-abstract” orientation of DST, the Cognitivist Program, and its extension; for as previously mentioned, fidelity stems from the dialectical interrelation between causal models and the (developmental) philosophical framework. As a result of this dialectical causal-abstract orientation, to a significant extent the cognitivist approaches talk past the functional approaches. More precisely, in terms of our four-dimensional framework, they have overlapping but significantly different 4-D subspace “shapes.” Without these framing shapes, the mistake too easily made is to engage in misleading attributions—e.g.,
functional models neglect development and emotional experience, and so cannot be adequate representations of the phenomena; or cognitivist models are merely descriptive phenomenological models that aren't scientific because they don't yield causal and falsifiable predictions.²¹ (Both are wrong: functional models, like FIPS, acknowledge developmental processes and folk experiences, but these are bracketed in order to produce boxologies. And cognitivist models, like dimensional appraisal models, not only can be more causally-based than functional models, they can also produce local predictions; but the emphasis is on the dialectical interrelations between these models and the developmental framework. For generally speaking, development is notoriously difficult to predict in detail; yet it is still a scientific notion, and so the criterion of providing falsifiable predictions is the wrong tool for the (holistic) job that the developmental framework performs.) With the strategic and functional stances in hand, we are now ready to cover the last program, the Social Constructivist Program, and its relation to the final of our three stances, the semiotic stance.

²¹ Recall from Chapter 2 that the four-dimensional framework helps to disclose when debates are at crosspurposes; see “Implications of the Three-Dimensional Space.”
Chapter 6: Social Constructivist Approaches to Emotions

Recapitulation

The Cognitivist Program claims that appraisals are emotive-cognitive systems, or dynamical modes of activation, represented via dimensional activation spaces. Part of what makes these representations "scientific" concerns their capacity to track "real" causal (i.e., somatic-neural) processes. These representations also implicitly start with a mid-level phenomenological orientation respecting various types of emotive "experience"—an orientation squarely falling within the human sciences. The strategic stance was used to frame the Cognitivist Program within the four-dimensional framework, and to situate its causal-abstract-phenomenological orientation ("crescent") in relation to the functional approaches of chapters three and four.

Prospectus

Both the Cognitivist Program and the Social Constructivist Program emphasize developmental (temporal) processes; the former program concentrates on individual-level processes, while the latter focuses on social-level processes. The two programs are interrelated: recall that individual "appraisals" are the former program's key idea, where emotions involve appraisals of events of significance for the organism; social "syndromes" are the latter program's central notion, which are social scripts (similar to Ekman's display rules) bearing upon individuals' judgments about interpreting emotions. These social judgments are cashed-out in terms of active appraisals of events of (social) significance. The Cognitivist and Social Constructivist Programs emphasize concerns flowing from temporality—concerns regarding how emotions experientially have a continuous and ever-changing nature, how emotions are centrally
bound up with coping-in-the-world, and how coping is enmeshed with group-related considerations (social scripts, communication, institutions, etc.). Such temporal concerns utilize an array of concepts from the human sciences.

Building upon the Cognitivist Program, in section 6.1 I discuss the Social Constructivist Program. There is a shift in the interpretation of "appraisals," where individual appraisals are placed in the background and social structures are highlighted. Social structures are seen as emotion-structures/systems; these structures express abstract capacities or tendencies that exhibit variegated relations to the causal and phenomenological elements upon which they rely; also the perspectives afforded by philosophical frameworks bring into relief emotion-structures. The issue of perspectives leads to coverage in section 6.2 of the last informational stance, the "semiotic stance." This stance in certain ways carves out a 4-D space "above" what the strategic stance discloses, but the added twists are that the semiotic stance is presupposed in both the functional and strategic stances, and it can be used to inquire into the very representations which it affords. I will discuss how this occurs and what it means for our four-dimensional space in section 6.2.

6.1: The Social Constructivist Program

Overview of the Argument

There are at least two forms of social constructivism: a strong form, which holds that emotions are appraisals largely constructed by culture (where other non-cultural factors are negligible); and a weak form, which claims that the examination of emotion systems should emphasize cultural factors, but where these factors are intertwined with other (non-negligible) "lower-level" systems (e.g., biological and psychological systems). Both forms of social constructivism have deep relations to the Cognitivist Program; the main link between the
programs is the reliance on appraisals. Accordingly, I first relate the strong form to the discussion of dimensional approaches covered in section 5.1 (see “Linking the Cognitivist Program to Dimensional Approaches”). On the conservative dimensionalist reading, the focus shifts from individual appraisals to descriptions of cultural-level emotion-scripts; on the less conservative reading, the construction of emotion-scripts suggests that scripts are, in a sense, genuine macro-level entities.

I then discuss the weak form of social constructivism; I will argue that this approach falls within the framework of Developmental Systems Theory (DST). However, the emphasis is now placed on cultural factors (not merely on organism-environment relations), and the manner in which the cultural system bears upon the “individual” system. While DST accommodates weak social constructivism in a most general sense, a more explicit framework tracking the interrelations between systems needs to be given; the framework assisting in this regard is “Dual Inheritance Theory” (DIT), which encompasses some of the conceptual features of weak constructivism. An ensuing discussion of where DIT cannot capture other aspects of weak constructivism discloses the need for the final informational stance, the semiotic stance, which I discuss in the last section.

Strong Constructivism and the Conservative Dimensional Approach

Starting with the descriptive use of dimensional representations (the conservative dimensionalist reading), if one is employing these representations as a strong constructivist, it is argued that given the wide range of data collected on various cultures, the attention to differences in language use reflects differences in affective concepts, and thereby different experiences. And while there are cross-cultural “homologies” between certain affective1 terms, upon closer

---

1 With respect to cross-cultural considerations, rather than using “emotion,” the less problematic (and also more general) term is “affect.” However, I will use both interchangeably, keeping in mind that “emotion” is a term of art.
insensitive the homologies mask the subtle differences in how affective terms are configured in semantic fields; these terminological differences reveal conceptual differences, which make enough of a difference for experiences. This might not be so apparent when terms have strong "homologies," but the contrast is made clear with "exotic" affect terms like amok, liget, amae, and fago (see Griffiths 1997, Ch.6; Harre 1986; Cornelius 1996, Ch.5; and Averill 1982). The practices associated with "exotic" terms suggest significantly different concepts, and thereby different experiences.

The focus on linguistic practices (i.e., the dimensional linguistic representations) leads to the conclusion that "emotion" is a (Western) term whose semantic field is judged to be "different enough" from affective terms used elsewhere. These differences are disclosed via a "theory-free" categorization scheme which "properly" represent practices. That is, when confronting a practice, first gather a range of affective terms, mark each term as a "character," and then do a best-fitting statistical analysis to produce a similarity diagram/dimensional representation showing the degree to which the terms (within a practice) are related to one another (e.g., White 2004).² Further analyses comparing diagrams across linguistic practices generate the argument that because diagrams express different organizations, and so different concepts, affective experiences can be starkly different across cultures. The high-level focus on the linguistic elements of affective phenomena is largely responsible for the conclusion.

Strong Constructivism and the Less Conservative Dimensional Approach

However, note that even if these dimensional representations are just descriptive (which is subject to dispute), it still remains that the inferences from differences-in-representations to

---

² In the biological field of systematics, a similar type of approach is offered by "numerical phenetics," where the outward appearances of organisms are used as a basis for assigning "characters" to organisms (characters identify certain anatomical features). Applying statistical methods to these characters, the best-fitting measure of "overall similarity" produces a "phenogram." Because all characters are initially given the same weight, the number-crunching procedure is deemed to be "theory-free" (or at least as theory-free as possible). See Mayr (1988, Ch.16).
differences-in-concepts, and then to differences-in-experiences are not "theory-free" descriptions. Further, even if this conservative approach is considered "scientific," as it is a research program using accepted scientific methods, standards, etc., the problem arises, how is this social construction? The descriptive project needs to be latched into a larger research program that asks how and why these representations make sense—that is, if the representations give what is constructed, then how and why does this occur? Such questions lead to the less conservative dimensional approach. For a strong social constructivist, on the less conservative approach social factors are emphasized; yet these factors still implicitly rely on individual appraisals. Appraisals help in part to explain how construction occurs: it is through "embodied mechanisms" like individual scripts, prototypes, and related concerns (e.g., metaphor, radial categories, frames, etc.) that affective phenomena are constructed.

These individual mechanisms, in transaction with environmental affordances, bring forth meaningful structures. For example, a communicative signal in the form of a question elicits an improvised "online construction" in the receiver through the above embodied mechanisms, whereupon an answer (a signal) is given. And in the process of dialogue, an answer raises further questions; the whole meaningful process is a constructive process. The less conservative dimensional approach is not merely descriptive, since it attempts to provide mechanisms for the construction of affective practices described by dimensional representations. More subtly, descriptive representations are also partly generated by individual appraisals, and both representations and appraisals mutually change in relation to communicative transactions.

---

3 In biology these days, there are few strict numerical pheneticists, since 1) the number-crunching game discounts phylogeny, and does not attempt to integrate its methods and results into the general evolutionary framework; and 2) the inherent "arbitrariness" of what counts as a character (as well as what statistic-measure to use) problematizes the "descriptive" claim of numerical phentetics (see Ridley 1993). Likewise, the point for conservative dimensionalists is that the project needs to move beyond the merely "descriptive" aspect if it to have promise as a future science.

4 For more on these mechanisms, see Lakoff 1987 and Kövecses 2000. (Note that I make no claim on whether they are classifiable as strong constructivists.)
However, for constructivists, individual appraisals are not emphasized in the transactional process: social forces are emphasized, where the emphasis on communicative relations shifts the focus from individual scripts to cultural scripts. For less conservative social constructivists, then, the dimensional representations track cultural scripts and their influence on individual affective appraisals. Cultural scripts are social constructions that “transactionally construct” individual appraisals, and in this sense cultural scripts are “real” entities (thus answering why construction occurs—it is “for the sake of” cultural scripts, as I elaborate below).

Dynamic Nominalism and Higher-Order Entities

But in what sense do these scripts become higher-order entities? First, as noted above, they are parasitic on individual appraisals, so they aren’t self-standing entities. Secondly, since less conservative dimensionalists who are still strong social constructivists want to privilege social systems over other systems (that is, the other systems are negligible for affective phenomena that matter regarding diverse experiences), there is a sense in which these higher-order entities are “real”—they aren’t just a collection of individual appraisals. Griffiths, appropriating Ian Hacking’s work on multiple personality syndrome (1995), offers a plausible way in which to interpret the claims that affective social entities aren’t self-standing, and yet are more than just a collection of individual appraisals. He notes that things like “citizens, members of parliament, and licensed dog owners are social constructions while electrons, magnesium, and clades are not. The categories referred to in the first list are social constructions, whereas those referred to in the second list are not” (1997, p.145). What makes each former category more than merely a collection of individuals is the idea of “dynamic nominalism,” which “differs from simple nominalism in that the members of a category do share something over and above the fact

---

5 Griffiths distinguishes this stronger sense of “social construction” from a weaker sense, where the concepts mentioned in both lists above “cannot exist independently of a community of speakers and thinkers, and each was created by a sociolinguistic process” (p.145).
that they are members of that category. However, the fact that the members have these shared properties reflects the existence of the category and the social practices in which it is embedded (p.146); in particular, these practices are “modes of being” that make a difference to the members who are actually engaged in practice. The dynamic element has to do with the genuine difference practice makes for its members. Applied to affects, affective cultural scripts express categories that are socially constructed; the categories exist, as reflected by the “shared properties” of its members (or more accurately, the categories are parasitic on individual appraisals, where appraisals have “shared properties” by virtue of their shared practices); yet cultural scripts are not merely the enumeration of individuals-with-appraisals (and so are not merely nominalistic); the characteristic that makes these scripts “real” is the dynamic aspect of adopting what the script affords within a practice—the script enacts a mode of being that makes a genuine difference for its practitioners and how they practice.


The problem with the strong constructivist’s interpretation of cultural affects as real, higher-order entities/“appraisals” is that the entities heavily rely upon individual appraisals. Since a strong social constructivist holds that other non-cultural systems are negligible in understanding affective phenomena that matter for practitioners’ experiences, it looks as if the “reification” of the cultural system does not take seriously enough the dynamical interaction between practitioners and systems. For dynamical nominalism relies on both individual scripts and cultural scripts to grasp the manner in which cultural affects are “real.” Since less

---

6 If the strong constructivist claims that cultural entities are “real” and less dependent upon individual appraisals than it appears, it seems to me that the burden is squarely placed upon the constructivist to justify this strong metaphysical reading; that is, it needs to be shown how cultural entities both are self-standing and don’t deeply depend on individual appraisals. A better interpretation of the reality of these entities is that they are abstract tendencies expressing the “ontological ground” (cf. Ch.2) of appraisals. While tendencies are (epistemically) parasitic on individual appraisals, the larger “functional” roles that cultural scripts play are stable enough to deem appraisals “negligible.” Tendencies express these functional roles, and as such are “real.”
conservative dimensionalists who are strong social constructivists desire to move beyond
description by offering explanations for how and why construction occurs, the above
interpretation reveals the need to push beyond reifying cultural systems.

Weak Social Constructivism and Overt Construction

The focus on dynamics makes way for weak social constructivism, which takes seriously
the idea of multiple systems with differential interactions between (changing) levels. Weak social
constructivism's key claim is that cultural emotion scripts are "syndromes." To comprehend this
notion, a distinction first needs to be made between what Griffiths calls "overt" construction and
"covert" construction. Overt construction is the explicit recognition by practitioners that a social
category expresses a conventional practice held together by the relevant beliefs; the general
contrast is between overt social constructions like money and parliament—which are clearly
acknowledged by practitioners as being conventional—and categories like electron, which are not
"merely" conventional. Overt constructions can highlight the existence of cultural affects as
(relatively) self-standing entities (the intent of conservative dimensionalists); or in line with weak
constructivism, overt constructions can emphasize the dynamic aspect of social categories as they
interrelate to individual appraisals. What matters for both possibilities is that cultural affects and
individual appraisals are explicitly acknowledged, or can be made explicit without significantly
changing the nature of the convention. For example, the modern social institution of paper
money depends on the employment of a particular type of (technical) medium that is
conventionally "acknowledged," in its capacity as a sign, as signifying a mode of exchange. If
the conventional nature of how the technical medium (i.e., it needs to be difficult to duplicate) is
used is explicitly spelled-out, that wouldn't significantly change the institution of money; the
response would probably be "obviously!" but surely the explication wouldn't undermine the
institutional practice—there would be no widespread response: "Oh no! All this time I thought I
had money; but all I really had were (technically sophisticated) pieces of processed wood pulp... alas, all is lost...."

Weak Social Constructivism and Covert Construction: The Need for a Higher-Order Perspective on Emotion-Systems

By contrast, covert constructions really do depend on the nature of the social category remaining “covert”; in particular, covert emotion systems are “syndromes” since they are interpreted as “passions” (in the sense of being disclaimed and passively “received”). James Averill represents this position. His weak social constructivism recognizes that systems are interrelated, and that individual appraisals are especially important. However, no set of emotive responses on the part of the individual will give “necessary or sufficient conditions for the attribution of emotion” (1982, p.7). This is because if emotion systems are only properly understood by examining the dynamic relations between other systems (the “cultural system,” the “biological system,” and the “psychological system”), then just examining a set of responses will not be enough to grasp these dynamics. For example, a fully enumerated individual “response system,” if one could be given, is not sufficient to account for what cultural factors elicit various emotions in particular situations (e.g., in social situation X is one experiencing mere fear? anxiety? uncertainty? angst? or perhaps existential angst?). And the system is also not necessary because it is possible to have a socialized attribution of emotion without a “genuine” response (e.g., there might be no overt show of an anger response, although a social occasion is recognized in a practice as an “angry situation”; an emotion could be “cut off” by a master actor—and so there is no proper internal response—yet the actor may adopt the emotive comportment as appropriate for that circumstance; if boredom is attributed to someone as a social emotion, how is the language of inner or outer “response” appropriate?). To move beyond individual response systems, the context of social transactions needs to be considered—a context requiring a higher-order perspective.
Social Syndromes and Dynamic Nominalism: Emotions as Covert Social Constructs (Higher-Order Appraisals)

The dynamic aspect of covert social syndromes holds that members of a practice interpret emotions as passions. A good illustration of this general phenomenon is the (widely cited) syndrome of amok, exhibited in certain Southeast Asian societies (see Averill 1982, Griffiths 1997). Triggered by a socially perceived dishonor (e.g., a victim of an affair, the inability to provide for family), a male “runs amok” by attacking others with varying degrees of violence. The behavior is socially interpreted as a disclaimed action (similar to a “momentary insanity plea”) and others briefly tolerate such violence, but the affair usually ends with the group members killing the male. The sense in which running amok is a passion has to do both with the fact that the man is not pretending to be in a frenzy (the passion is real for that person), and because it is also interpreted as a disclaimed action by others—the dynamic nominalism of amok concerns being affected by circumstances and acting out in response to being so affected.

Regarding the latter aspect— that amok is interpreted as a disclaimed action by others—Averill notes that the cultural script is seen as a passion (as “acted out”) by members of a group; that is, certain emotions, in their capacity as covert social constructs (cultural scripts), are not really passions; they are interpreted as passions, making them “active” insofar as they are parasitic on individual appraisals. The social category is an instance of dynamic nominalism since the man who runs amok “would not be in a [genuine] frenzy unless he had learned that this is an appropriate response to certain unbearable social pressures” (Griffiths 1997, p.141); the cultural script enables this new mode of behavior, and since the practitioners adopt the interpretation afforded by the script, a new mode of being is created by tacitly adopting that interpretation.

Thus amok is, from this higher-order dynamical perspective, both a form of display (for the group), and a felt state (for the individual who runs amok); it is a mode of activation tapping into two systems—a cultural-level system and an individual appraisal system—jointly required for
understanding the dynamical nature of covert constructs. Like evolutionary psychology and the Cognitivist Program, “emotions” are modes. However, unlike these programs, which emphasize individual emotive-cognitive modes of operation, the Social Constructivist Program emphasizes dynamical modes of operation from the perspective of modal interrelations between individual-level and cultural-level systems.

Interpreting Social Syndromes: The Importance of Group-Level and Individual-Level Perspectives

The emphasis on tacit interpretation is crucial, since it makes certain socially constructed emotions (like amok) covert constructions. They are not overt, because if the community of practitioners and the person running amok were to realize that amok is just an interpretation of an emotion as a disclaimed action, the practice would diminish in force (or to illustrate with a different example, multiple personality syndrome is a social pretence that cannot survive the realization that it is merely our invention (Griffiths 1997, p.145–6)). As Griffiths notes, these covert categories “are ontologically on a par with overt social constructions but are treated by the community as if they corresponded to independent distinctions in nature [emphasis mine]. They are treated in the same way as categories like chlorine or motor-neuron disease” (p.147). What makes covert social constructs powerful and strange from a meta-level view is that individuals are appraising an emotional syndrome as a passion, and thus are making “active” (online) judgments using individual scripts, yet for the practice to make a difference for that form of life it must be tacitly understood as a passion.7 This mechanism makes (covert) cultural scripts “real” not only for each individual, but also “real” in terms of how it forms group bonds. The appraisal process and its widespread adoption within a practice are the “glue” by which it makes sense to speak of a

7 Note some of the structural similarity to in-group morality in various forms of life where, in spite of certain principles with more universal import (e.g., universal love, democracy and freedom, etc.), the lessons do not tend to carry over to out-groupers. Still, in-group members generally see no inconsistency.
form of life as a form, and additionally, the tacit interpretation of cultural scripts as passions insulates the practice from "non-believers," further perpetuating this dynamical form.

The (Perspectival) “Interests” of Syndromes: Group-Level, Individual-Level, and “Mixed” Perspectives

At this juncture, it ought to be pointed out that there is an ambiguity in interpreting Averill’s disclaimed action account (for those socialized emotions that count as covert constructs). As Griffiths inquires, is the claim that syndromes are improvised responses on the part of agents who are “manipulating” (or rationalizing) what they know about the operation of cultural norms? Or is a syndrome better interpreted as an emotion inculcated within individuals on the basis of reinforcement, where the emotion develops into an automatic mode of behavior? On the improvisation account, “free-riders” may exploit what the cultural norms afford, but more generally, whether these agents are ill-intended or not, they are consciously tapping into these norms. On the reinforcement account, which Griffiths advocates, affective scripts within the cultural system help to shape and guide what then later become largely automatic responses “appropriate” to the practice (p.142); so if there are “free-riders” they are really just (collateral) expressions of the indoctrination process. The point is that latent within both interpretations is the issue of “interests”—what interests do syndromes serve? From the perspective of the reinforcement account, they primarily serve the “interests” of the social structures (the “preservation of tradition,” for example), while from the perspective of the improvisation account they serve mixed interests, since the “interests” of individuals and group-level norms may be aligned, but may also be in conflict.

Averill’s Three Systems Situating Emotive Social Syndromes (as Modal Systems)

Informally, the upshot of covert social construction is that it would be incorrect to say “an emotion is an emotion is an emotion.” Emotions serve varied and potentially conflicting functions (“interests”) depending on the perspective taken. For Averill, understanding the
dynamical interests of emotions requires adopting the perspective that populations of response-tendencies covary in multiple ways, a perspective foregrounded by the systems and parameters which encompass and configure, respectively, populations and covariance. Since Averill is primarily a (weak) social constructivist, the most relevant system is the cultural system. The general picture portrays an emotion as a system tapping into multiple systems, where highlighting the cultural system then discloses its covariance with other systems; syndromes are actually just one mode/subsystem within the broad conceptual framing of multiple systems and their dynamical interrelations. There are three hierarchical systems in particular that Averill distinguishes: the biological system, the psychological system, and the social system. All of these are distinguished only in theory—not practice—as they are abstractions for better comprehending dynamics (Averill and More 2000). Each of the three systems has four levels of hierarchical organization; from bottom to top the levels are: 1) component elements (e.g., for the biological system the elements are "fixed action patterns"; for the social system they are "typifications" (tokens) of social "roles" (types)); 2) subsystem/syndrome (e.g., for the psychological system the subsystem is one’s folk states; for the social system it is social "roles"); 3) system (e.g., for the biological system this level concerns crucial, species-typical "instincts" (drives, really) like the need to reproduce, attachment, etc.; for the social system this level concerns "institutions"); and 4) suprasystem/inclusive unit (for the biological system the suprasystem is the "species"; for the psychological system it is the "self"; for the social system it is "society"). Observe that within each of the three systems the levels have vertical interrelations, and even more importantly there are horizontal interrelations between systems. Thus dynamical emotion syndromes are systems since their various elements are situated in relation to each other (vertically and horizontally), and to the construct as a whole (the individual-cultural emotive mode enacted by these parameters). Accordingly, Averill’s levels of organization highlight that when investigating social syndromes
(a mode at the subsystem level), both horizontal and vertical relations need to be taken into account. Unlike strong constructivists, Averill recognizes the importance of not reifying any aspect of the abstract scheme for understanding complex processes.

**Developmental Systems Theory (DST) and Social Constructivisms**

The distinction between abstraction and process helps to frame various interactive relations between systems—relations which delineate the functional roles of syndromes. Hence whether we adopt the previously mentioned improvisation interpretation or reinforcement interpretation of syndromes (or a mix), the general systems orientation, with its vertical and horizontal relations, encompasses either type of investigation. It is a natural transition to observe that this systems orientation is capturable by DST; Averill focuses on the structural features of syndromes, downplaying explicit developmental considerations, yet his systems orientation is broadly developmental. That is, although social syndromes interrelate (vertically and horizontally) to psychological appraisals and biological action tendencies, the focus is really on *levels and systems*; developmental dynamics ends up being an implicit category connecting the vertical and horizontal relations. To make clear this developmental “backbone”—and manner in which DST’s general framework accommodates the weak Social Constructivist Program—observe that the biological, psychological, and social systems are conceptually representable as three “differential equations”: one “structural equation” for the biological “organism”/system (“dO_y/dt”), another for the psychological “organism”/system (“dO_y/dt”), and the last for the social system (“dE_i/dt”). The *levels* within each system may be represented by “sub”-functions for the three equation types that capture each system’s vertical relations. And the horizontal relations would be captured by the notion of three *coupled* equations.

So most generally, DST encompasses the weak Social Constructivist Program; it also accommodates the strong program, since if the coupling between equations is made negligible,
the focus then shifts to an equation type expressing the "reified" social system's structure. As I mentioned in the previous chapter (see fn.18), the strategic stance coheres with DST, restricted to certain biologically relevant dynamics. It is clearly not possible to draw sharp a priori boundaries between where the strategic stance ends and where the broader considerations about dynamic relations (pertaining to the social constructivist's systems) begin. Even stronger, there is significant overlap between both, given the strong interconnections between the Cognitivist Program (and its extension) and the Social Constructivist Program. Still, since DST's framework is more general than the scope of the strategic stance, and since DST accommodates the Social Constructivist Program's architecture, a more tractable tool is needed that assists in distinguishing where the Social Constructivist Program differs from the Cognitivist Program. For while DST accommodates both of these "ecological" programs ("ecological" since they deal with communication-and-appraisals)—what Griffiths calls the "heterogeneous construction of emotions" (1997, p.132)—a finer-grained tool is required to bring into relief just where the two programs differ in emphasis.

**Dual Inheritance Theory (DIT): Distinguishing the Cognitivist Program from the Social Constructivist Program**

The tool for the job is Dual Inheritance Theory, which, like DST, comes from biology: it is a particular theoretical approach to understanding systems and their interrelations whose most crucial feature is the theoretical perspective adopted. DIT has several key assumptions (Smith 2000): social information is "heritable" (via imitation, learning, etc., which are distinguished from "non"-social learning mechanisms); the spread of social information occurs by "multiple forces" (e.g., natural selection, appraisals and scripts, parent-child transmission, etc.); "models" of these forces are similar to models in evolutionary population genetics; and behavior is "codetermined" by genetic, social, and nonsocial environmental factors. From these assumptions, DIT generates three general conclusions:
1. Since culture exhibits the three characteristics required for evolution by natural selection (variation, heritability, fitness effects), cultural evolution can be analyzed using neo-Darwinian methods.
2. Since cultural inheritance differs from genetic inheritance in key ways [e.g., many cultural modes of inheritance are Lamarckian] (e.g., non-parental transmission, multiple transmission events over a lifetime), the evolutionary dynamics of culture will also differ in important but analytically understandable ways.
3. Genetically nonadaptive cultural evolution is possible, and it is more likely when the differences just referred to are most marked (e.g., in modern bureaucratic societies and other hierarchical social/culturation structures) (p.32).

A brief expansion of each of these points will be taken up in sequence. For the first point, variation occurs on two “dual” levels: an “individual”-genetic system (i.e., a population genetic system is an “individual”) and a cultural system (if there is a psychological system it straddles the two systems). For both systems, variation is expressed by statistical characters (“traits”) and the measurable differences between characters; in other words, the neo-Darwinian use of statistical tools for thinking about abstract populations, characters, and variation can be applied to both systems. Likewise, the neo-Darwinian methods employ the (statistical) concept of

---

8 A “cultural system” concerns aspects of a “form of life.” As Durham (1991) notes, quoting Clifford Geertz (1973), culture (aspects of a form of life) is an “ordered system of meaning and symbols...[it is] the fabric of meaning in terms of which human beings interpret their experience and guide their action; social structure [actual pattern of behavior] is the form that action takes, the actually existing network of social relations” (p.4). In Durham’s terms, a cultural system “should be thought of not as behavior but as part of the information that specifies its form” (ibid). Moreover, since a cultural system models aspects of a form of life, these aspects are analytically representable. That is, a recent trend in anthropology is the emergence of what Roger M. Keesing (1974) has called “ideational theories” of culture. These new approaches offer a more explicit and more analytic conceptualization of culture than has been possible since Edward B. Taylor introduced the term to anthropology, calling it “that complex whole which includes knowledge, belief, art, morals, custom and any other capabilities and habits acquired by man as a member of society.” (1871, I: I.) For years, anthropologists have debated about the range of phenomena to be included within that “complex whole”.... The trend in recent years has been to move away from the ambiguous and overgeneralized expressions of Taylor and other early writers and to “narrow the concept of ‘culture,’” as Kessing put it, “so that it includes less and reveals more” (p.3).

Accordingly, coevolutionary models in DIT model aspects of a form of life via Lamarckian mechanisms like “learning rules,” direct imitation, etc.; see fn.12 below.
9 Durham writes:

As Geertz argued some years ago, “culture is best seen not as complexes of concrete behavior patterns [what was called “social structure” in fn.8 above]—customs, usages, traditions, habit clusters—as has, by and large, been the case up to now, but as a set of control mechanisms [those “aspects of a form of life” modeled by a “cultural system” mentioned in fn.8 above]—plans,
heritability for tracking the transmission of characters from one generation to another, which is possible because heritability is general enough to abstract away from the differences between genetic inheritance and cultural "inheritance" (heritability, really). Lastly, fitness effects concern population-level characters which differentially "survive" in subsequent generations, whether these entities are represented via population genetic models, or via "group trait models."¹⁰

For the second point, the two systems also differ in some of the details of variation, heritability, and fitness effects. Variation in the cultural system differs from variation in the genetic system primarily because cultural variation depends upon statistics to sustain its "ontology," whereas the genetic system taps into known mechanisms which further constrain the

¹⁰ For example, in the group selection models offered by Sober and Wilson (1998), "a group is defined as a set of individuals that influence each other’s fitness with respect to a certain trait but not the fitness of those outside the group. Mathematically, the groups are represented by a frequency of a certain trait, and fitnesses are a function of this frequency." (p.92). This is just the use of population genetic ideas, but here traits aren’t tied to "genetic points"; rather traits/characters are defined more generally in relation to frequencies.
statistical results. For the genetic system, inheritance involves passing "the genes" to the next generation, whereas for the cultural system there is only inheritance by analogy—the statistical concept of heritability is primarily operative, where statistically significant types are deemed as cultural characters tracked through "generations" (e.g., parent-child lifecycles, larger group lifecycles, script or "memetic" lifecycles, etc.). Lastly, for fitness effects, the cultural system differs from the genetic system since there are two qualitatively different types of mechanism: selection may occur by "biased transmission" of information, where people preferentially adopt certain cultural variants over others, or selection may apply to certain cultural variants ("guided variation"), where the variants "survive" because they "affect the lives of their bearers in ways that make those bearers more likely to be imitated" (Richerson and Boyd 2005, p.79).

---

11 Richerson and Boyd note that the two systems can theoretically "keep pace" with one another, but this only occurs when a genetic difference between, say, two groups exists on average, and "this average genetic difference must be large compared with the average difference in culture and environment between the two groups" (Richerson and Boyd 2005, p.38). To clarify, two possibilities are suggested: first, when the two systems "keep pace," variation in the cultural system does not significantly differ from variation in the genetic system, in relation to other groups and their dual systems (e.g., imagine two groups whose cultural gap is, on average, not as far apart as their average genetic difference; this induces two "islands" of "reproductive isolation," and thus on each island, the dual systems get "locked in"). And secondly, when the systems do not keep pace, the average genetic difference between groups may not be large compared with the (larger) average difference in culture and environment (e.g., commonsensically, we are one species with diverse cultures).

12 The former, "biased transmission," can be direct, indirect, or frequency-dependent; such transmission concerns "what's going on inside the head" of decision makers. Direct biased transmission involves directly imitating/learning a cultural variant/character, where a character initially spreads like "wildfire" and slows down as the frequency increases (think: "juicy gossip" spreads quickly, but as it becomes more widely known, it loses its force). Indirect biased transmission occurs when some preferred character, say "street smarts" associated with Humphrey Bogart, indirectly causes a preference for smoking. Frequency-dependent biased selection occurs when the probability of acquiring a cultural character is higher than its population frequency (think: if a potential fad doesn't reach "critical mass," it won't spread like wildfire; but if it does, it takes off, and then settles down once the fad becomes widely known). The latter cultural selective mechanism, "guided variation," focuses less on "what's inside the head" and looks at the propagation and transformation of "learning rules." A cultural "learning rule" is transmitted via a "parent" to a carrier/"offspring" ("guided"), where the carrier adopts and modifies the rule in accordance with some goal ("variation"). For example, the recipe for chili transmitted from the mainland U.S. to Hawaii has been modified to regional tastes; both the adoption of those "tastes" as what, say, people in Hawaii understand as chili, and the directed actions of chefs in appropriating "the recipe" for (Hawaiian) chili are central to guided variation.
Because the two systems differ, the third point claims that it is possible for the systems to diverge in their "interests." From Chapter 5 (see 5.2, "Summing Up") the view of strategies as external signals provides an illustration of how dual systems could diverge: external emotion systems (as strategies), expressed by conventions or norms, are cultural characters; the multiple stable equilibria in an evolving "culture space" could result in genetically nonadaptive evolution, especially when the strategies are strongly attractive basins that dominate the equilibria of the genetic system (for example, norms can amplify group selective processes leading to "functionally adaptive" groups, some of which end up committing, say, mass-suicide; see Sober and Wilson 1998, pp.149-54). So while the previous two points allow for systems with the "same" interests (or at least interests that aren't divergent), the third point highlights that systems could be in conflict. The implication for emotion systems is that these systems, in line with the Social Constructivist Program, have varied types of interests, whether cooperative, neutral, or conflictive. And in terms of our four-dimensional framework such interests express the abstract capacities or tendencies of the characters that propagate through the population.

The problem with DIT, however, it that it is theoretically rich but empirically poor (Smith 2000, p.32). Moreover, while it captures some of the interrelations regarding the systems of weak social constructivism, its theoretical framework more or less still falls under the strategic stance. The contribution that DIT makes is its population-level orientation—the population genetic system and the population cultural system (populated by cultural "variants")—which assists in distinguishing the Social Constructivist Program from the Cognitivist Program. This is important, since both programs have been labeled as "ecological" theories of emotions. The programs are distinguished by the different types of ecological dynamics giving rise to higher-order emotions: for the Cognitivist Program, emotions are multidimensional environing

---

13 The mathematical tools used in DIT are essentially dynamical systems of the sort capturable by the strategic stance. See, for example, Boyd and Richerson (1985) and Hofbauer and Sigmund (1998).
appraisals of what is significant for the organism (see Griffiths 2004a, 2004b); for the Social Constructivist Program, emotions are population-level entities (roles, groups, etc.), where the emphasis is on tracking emotion-characters and their interrelations with other systems. But while DIT assists in this respect (in distinguishing the two programs), it also turns out that weak social constructivism goes beyond the scope of the strategic stance, and thus goes beyond DIT.

Distinguishing Dual Inheritance Theory (DIT) from Developmental Systems Theory (DST): Creating a Space for Differing Philosophical Frameworks

The notion of covert construction makes clear the added qualitative dimension that DIT cannot accommodate. For while DIT yields representations which cash-out intuitions about the dynamics between systems at the population level, these are not representations intended to also represent their own representations and the ensuing “self-reflective” dynamics. Let me explain.

To revisit the distinction between overt and covert constructs, for the former, revealing the social nature of the construction does not significantly alter the dynamics of the practice which enables the construct; but for the latter, revealing its nature disturbs the practice itself, and thus potentially destroys an emotion-syndrome. Scientific representations provided by DIT—even if they could capture the general dynamics of weak social constructivism—still presume a kind of “dualism” between what the representation (i.e., the sign) is a representation of (i.e., what is signified), and in what capacity the sign relates to the signified. In particular, the capacity is assumed to be one where it is legitimate, “scientifically speaking,” to separate the sign from the signified without much damage to the dynamics of the system studied. But such an assumption is problematized when dealing with covert constructions, since the representational system used, if revealed to the practice studied, might actually subvert the delicate order established through what the emotion terms are tacitly presumed to “refer” to. For dynamic nominalism, the employment of these
terms relies on the covert (dispositional) mechanisms that confer a kind of reality to the terms constructed and insulate the group from other practices.14

As a result of the “recursive difference” made by what a representation is taken to mean, at the level of philosophical frameworks a space is created for genuine disputes over the presuppositions guiding the employment of scientific representations, over what makes them “scientific,” and over their bearing upon experience and praxis. The outer reaches of DST acknowledge precisely this point when we look exclusively at its philosophical framework of “dialectical biology”—a framework emphasizing the multiplicity of developmental organism-environment relations, which includes how the “observer” is intertwined with the study of such relations. From the standpoint of viewing science as heterogeneous, philosophical frameworks about the way in which science “ought” to proceed have different assumptions regarding the roles of representations, whether “dualistic,” “dialectical,” or otherwise. Thus when comparing philosophical frameworks at the meta-level, there is room for dispute about just what the phenomena are—phenomena at what level, for what representation, and in what capacity? Covert construction raises these issues, revealing that the strategic stance, while it can capture a number of ecological approaches to emotions, cannot really handle the meta-level problem of conflicting philosophical frameworks. For the strategic stance already presupposes that types of representations model kinds of behaviors, and the stance then affords a way to bring into relief certain aspects of those behaviors. But if the dispute is over the very capacities of the representations and what the representations are representations of, the strategic stance gives us

14 Additionally, as Hacking suggests, what links various constructivisms together is the idea that once we understand social constructs, we are in a position to see that things need not be as they are; and if we understand how and why social constructs are built, we are also in a position to change things through the building process (1999). Applied to syndromes, building new social scripts is really about new forms of social/group arrangement—a tradition going back at least to Plato (the “noble lie”). (Very briefly, for Plato the noble lie built by the philosopher is actually a covert construct; images are “passively accepted” by the hoi polloi, as their reason is too weak). Thus if Hacking is correct, constructivism is, to some extent, a meta-level position that ought to be distinguished from “object-level” constructs like syndromes.
no genuine assistance; a wider stance is required to accommodate the outer philosophical reaches regarding how, in terms of our four-dimensional framework, abstract capacities and tendencies are taken as capacities and tendencies. *It is at this level of examining conflicting philosophical frameworks that we can have disputes over conceptions of “science” and just what might constitute “science”—especially since these outer philosophical reaches of DIT and DST are centrally bound up with highly “humanistic” concerns.*

6.2: The Semiotic Stance

The semiotic stance provides the means for orienting philosophical disputes over representations. In particular, the above representations deal with interrelations between systems, and the philosophical disputes are about the adequacy of the representational schemes employed. Since the functional stance and the strategic stance apply to certain “lower” levels of investigation—the 4-D subspace “shapes” situated in our four-dimensional space—the last “shape” occurs at the capacity and tendency level; but the added twist that weak social constructivism reveals is that when the dispute is over the philosophical frameworks deployed, we are asking not merely about dispositions (capacities or tendencies), but also about the (meta) capacity in which the dispositions (expressed by models) are adequate representations. The functional stance carves out the base level; the strategic stance carves out the next level; and the semiotic stance carves out the highest level delineating dispositions—both in terms of the “object-language” concerning dispositions, and the “meta-language” concerning a philosophical framework’s presuppositions about dispositions. Thus part of what makes the semiotic stance unusual is that it can be used to inquire into its own highest level by asking about the meta-capacities of dispositional representations. Additionally, the semiotic stance is unusual since it is already presupposed in the functional and strategic stances, which I discuss at the end of this
section. Because the semiotic stance can “borrow” from each of these two stances, it is useful when what is in dispute is the very notion of a representation operating at a certain “level”—that is, when the boundaries are in dispute precisely because certain “emotions” are thought to be inextricably entangled in multiple levels (what will be called “thickets”).

Peirce on Semiotics

I will appropriate C.S. Peirce’s semiotics to represent this stance. Indeed, as two commentators on semiotics note, their favorite (hedged) “definition” of semiotics comes from Peirce: “semiotics is ‘the quasi-necessary, or formal, doctrine of signs.’” Simply put, Peirce is here emphasizing the equivalence he sees between logic and semiotics, where signs are not only the products of inference but raw material for further inference independent of the objects themselves” (Cunningham and Shank). Put in the language of informational stances, (intentional) “agents”/systems, whether functional or strategic, can serve as the raw material for semiotic inference. The way in which semiotics bears upon multiple-level thickets will become more apparent as the bits of Peircean language are put into place.

Inquiry and Dispositions

For Peirce the irritation of doubt spurs us on to inquiry, and through this process the eventual feeling of believing fixes our habits/dispositions as rules of action (Buchler, 1955, p.10). Such rules are essential to meaning; what a thing means is what consequences it involves, which is our idea of a thing’s sensible effects, including the various relations that would tend to be exhibited. Even stronger, the meaning of something is tied to Peirce’s conception of truth. Briefly, truth is that which would be fated to be agreed upon by an ideal community of inquirers; meaning is related to truth in that the total range of dispositions a thing would involve is that thing’s (full) meaning—(full) meaning concerns the totality of the thing’s consequences. An

\[\text{Note that Peirce is a realist, and that relations are real.}\]
implication of the pursuit of truth as it relates to dispositions is that our assessment of a disposition at any stage of inquiry is always subject to revision since no actual community can ever grasp a totality. Additionally, the process of refining our dispositions is a matter of bringing control to our inquiries. And this means not only our “inner” dispositions, but also “external” dispositions, for it is part of Peirce’s realism that the range of consequences which we ought to consider cannot be restricted to merely its potential cash-value for us. Rather it is only when there is no doubt (and thus no reason for further inquiry) that there can be in fact an “absence of control,” because this would be an ideal state of a disposition in its full-blooded reality (there is no need for controlling inquiry since in this ideal case inquiry ends).

A Triadic Framework for Understanding Dispositions: Sign, Signified, and Interpretant

How should we understand dispositions as (consequential) relations? Peirce categorizes relations as they exhibit Firstness, Secondness, or Thirdness. There are a number of ways to characterize these categories, although generally Firstness is “quality,” Secondness is “reaction,” and Thirdness is “meaning.” Quality is an abstract “potentiality” of a wide range of effects that is nonetheless real; reaction is the category of the “mutual struggle” between two things without considering their relation to a third; and meaning involves dispositions as the “third” which mediates between Firstness and Secondness as a (real) relation. Peirce’s work on semiotics lays out the broad framework by which to understand the three categories. Although there are medium and large-sized classificatory schemes that Peirce offers to try and document all classes of signs, the gist of these schemes is that Firstness deals with signs, which are things that stand to some agent for something in some capacity (“potentiality”); Secondness deals with objects, which are what the signs stand for (“mutual struggle”); and Thirdness deals with “interpreters.”

---

16 For Peirce generals are also real.
that relate the objects to the signs and give those relations *significance* ("meaning")—it specifies how a sign is to be taken, relative to what it is a sign of, in its capacity as a sign.\(^{17}\)

**Semiotics as a “Dispositional” Framework**

The general logic that Peirce develops is one of sign action, or semiosis. Signs are broken down into icons (whose qualities are similar in some capacity to the object of representation), indices (natural signs such as smoke being a sign of fire), and symbols (conventional signs, which encompass icons and indices; in particular, symbols as used in logic are divided into terms, propositions, and arguments, where argument-forms represent the major types of scientific inference: inference to the best explanation, induction, and deduction.) What matters is that all of these distinctions are not static, but are conceptual moments of sign action. Furthermore, semiosis centrally involves the notion that information is tied to a conception of infinity: the three *placeholders* of sign, object, and interpretant each "fully constitutes the symbol [the conventional representation employed by the interpreter] and yet all are essential to it. Nor are they the same thing under different points of view but [rather] three things which attain identity when the symbol attains infinite information" (Peirce (Vol.1), p.503). This "infinite information" is the totality of differences that make differences with respect to a symbol; a symbol is identified (in the "limit") with the relevant signs, objects, and interpretants when the total information expressed by a symbol is just the totality of the ways in which sign, object, and interpretant are relevantly configured. Logic, as the study of inferences, is more generally the study of sign action whose reality again is tied to a consequential totality. The importance of semiosis, information, and symbol (as a conventional sign) is that our *representational-inferential models*—whether mathematical, qualitative, conceptual, etc.—as employed by various sciences are, as sign actions, dispositional.

\(^{17}\)For further classifications of signs of signs (e.g. Icon, Index, and Symbol; Rheme, Dicsign, and Argument; etc.) see “Logic as Semiotic: The Theory of Signs” in Buchler (1955).
Inquiry and Semiotics: A General Framework for Projectibility

Putting the above together, the picture is that through the process of inquiry, dispositions acquire further degrees of informativeness—we have a growing web of investigation, or an expanding hierarchy of “dispositional information.” Informally, take “Level 1” in the “universe-as-dispositions” as a starting point for inquiry; other levels then establish relations between objects at Level 1. Continuing in this manner, in principle there can be an infinite hierarchy of Levels that acquires indefinite and massive semiotic significance. It might seem that such a claim is at odds with Peirce’s position that “in the limit” dispositions clarify themselves by giving a complete working out of reality. But it must be remembered that dispositions serve as both statistical results as well as that which are acquired by systems. Allow me to explain.

The upshot for semiotics as it relates to dispositions is that dispositions are sequences of interpretants that specify an evolving range of consequences relative to the process of inquiry. That is, such sequences express an “inheritance of acquired characteristics”—dispositions (probabilistically expressed) inductively self-correct themselves and so ultimately have nothing to do with particular belief-based hypotheses. “Rather [the inductive use of] probability is about relative frequency. Relative probability is not about degrees of belief [per se], but a ‘realism’” (Hacking (1990), p.210). Inferences have “truth-producing virtue.” In other words, truth is what induction gives. His [Peirce’s] theory of probable inference is just a way of producing stable estimates of relative frequencies. But on the other hand the real world is just a set of stabilized relative frequencies whose formal properties are precisely those of Peirce’s estimators [the universe acquires dispositions]. Method and reality do not fit by good fortune or preestablished harmony. Each defines the other (p.213).

All is semiotic activity; dispositions differentially create and are differentially created by other dispositions—the stable (statistical) estimators of relative frequencies also partake in what dispositions are acquired. Thus the potentially infinite hierarchy of semiotic relations is tied to the infinite process of inquiry that approximates truth as it “heads toward the limit.” Semiosis is
indefinite in significance, where significance is primarily a function of the sequence of interpretants that serve as a third relation to draw out informativeness between "chaos and (dull) order" as they operate within and between levels. In brief, this third relation expresses degrees of projectibility (or more generally, resiliency) relative to past successful inductions.

The Semiotic Stance Encompasses the Functional and Strategic Stances

Peirce takes relations seriously, for they permeate his whole system. Furthermore, Peirce's semiotics is all about classifying types of relations. And dispositions are, in short, relation-saturated. On this basis it is at least conceivable that there might be explanations that appeal to complexes which can only be understood with an adequate fund of relations. Thus Peirce's semiotics is clearly applicable to multiple-level thickets, especially when boundaries and levels are in dispute (specifically concerning disputes between philosophical frameworks, as with DST and DIT\(^8\)). Additionally, not only would functions and strategies be accommodated somewhere within the semiotic framework—since the functional stance takes a system (sign) as signifying some "real system" (interpreted via the sign's capacity to obtain successful predictive and explanatory results), and since the strategic stance builds upon the functional stance in a similar semiotic manner—but more generally the indefinite range of relations, bundles of relations, relations between bundles, and so on are capturable by the rather elegant triadic structure of Firstness-Secondness-Thirdness. The semiotic stance, then, yields a powerful tool for framing dispositions and systems. And while we need not wholly subscribe to the mechanism of "acquired characteristics" in accounting for the "evolution" of dispositions—indeed, modern evolutionary theory broadly teaches us otherwise (i.e., Lamarckian mechanisms are taboo, as are global teleological processes)—what Peirce gives us, I submit, is a most general view of relationality, and the manner in which our (epistemic) inquiries are (dispositionally) informative.

\(^{8}\) I discuss this point in section 7.5.
However, observe that there is something like a Lamarckian process in DIT, where representations track particular ways that culture “acquires characteristics” (the heritability of cultural characters); and also the organism-environment mutualism of DST is compatible with Peirce’s notion that dispositions create and are created by other dispositions. *The semiotic stance assists precisely in these outer philosophical reaches of evolutionary theory and their tangled relations to culture through Firstness-Secondness-Thirdness.* Indeed, the semiotic stance is the tool which helps to disclose differing conceptions of “science” at the level of philosophical frameworks.

**The Semiotic Stance and Philosophical Frameworks**

In sum, the semiotic stance creates a space for orienting disputes about differing philosophical frameworks, whether “dialectical” (DST), “dualistic” (DIT), or otherwise. The question at this meta-level concerns the informativeness of inquiries in light of the dispositional representations used—that is, the question is about the meta-capacity in which these representations tend to produce “worthwhile” results. If this meta-level point is not explicitly recognized, disputes can degenerate into mere “talking past” one another, since it wouldn’t be made clear exactly where the differences lie (e.g., DST is “just” philosophical and not scientific, since the framework isn’t falsifiable; DIT is “just” mathematical metaphysics without any empirical traction, and so is not science; and so on). The semiotic stance affords us a way to see what is at stake in multiple-level thickets, either at the object-level of dispositions (and how they relate to aspects highlighted by the functional or strategic stances), or at the meta-level. In terms of our four-dimensional framework, the former corresponds to the 4-D subspace emphasizing

---

19 I discuss some of the connections between evolution and its cultural extensions in Chapter 7. Here I simply make the general point that the semiotic stance may assist in framing the tangled relations between culture and evolution (specifically, the tangled relations between “culture”—whatever that may be—and the “outer reaches” of DIT and DST).

20 Recall from Chapter 2 that the four-dimensional framework can assist in disclosing where disputes are at cross-purposes.
dispositions and research programs, and how these parameters then bear upon the fidelity axis and the two "lower" functional and strategic shapes. The latter corresponds to "fixing" the philosophical framework parameter, and then inquiring into the assumptions about dispositions; when this peculiar 4-D subspace is emphasized, we are actually at the meta-level, asking "to what extent are the philosophical assumptions that guide our object-level representations responsible for disputes about multiple-level thickets?" For just what the "levels" are may be ambiguous, as may likewise be the case with the "systems" and the object-level "dispositions" expressed.

6.3: Overview

By applying the general framing tools provided in Chapter 2, chapters three through six organize the conceptual features of different emotions research programs. I have examined these features at a high level of investigation, and from the viewpoint that we are dealing with heterogeneous "sciences." I save further discussion of this heterogeneous perspective for chapters seven and eight. One of the major reasons for adopting this viewpoint is that most contemporary research programs reject a division between "emotion" and "cognition." As a result, they examine emotive-cognitive systems, which tend to be "heterogeneous" (centralized) systems that still exhibit degrees of tractability. Thus we not only have investigations of heterogeneous emotive-cognitive systems; we also have a "heterogeneous array" of sciences investigating, at multiple levels, these "creatures of mystery" called "emotions."

I have also selected emotions research programs exhibiting degrees of evolutionary import. My reasons for this evolutionary slant will be explained in the following chapter. More specifically, the following chapter further expounds the links between biology, psychology, and the emotions research programs. It should be noted that at this stage of the discussion, I haven't directly addressed whether a science of emotion is possible. Rather, at this stage I have only
attempted to align the major parameters of promising emotions research programs by way of the
four-dimensional space and the three informational stances. Accordingly, an overview of the
argument thus far is expressed in the following diagram:

Figure 10. The Three Informational Stances and the Four-Dimensional Space
Chapter 7: Biology and Psychology: The Link Between Evolution and the Four Emotions Research Programs

Recapitulation

Chapter 3 and Chapter 4 examined "functional" approaches to emotions—"functional" in the (synchronic) reverse-engineering sense, and "functional" in the (diachronic) adaptationist sense, respectively. I further discuss these two senses in this chapter regarding the methodological orientations of evolutionary biology and cognitive psychology. Chapter 5 and Chapter 6 emphasized the developmental-social dimensions of emotions; if there is a sense of "function" applying to these research programs, it concerns the emphasis placed on developmental-social "functional processes." I discuss in this chapter how "functional processes" pertain to developmental views in biology and developmental-social views in psychology—views which differ in emphasis from cognitive psychology, especially given their more prominent human-science orientation.

Prospectus

As mentioned at the end of the previous chapter, I have not addressed the question whether a science of emotion is possible. I suggest that establishing a science of emotion may stem from a complex-systems approach. Chapter 8 will outline a complex-systems perspective, and Chapter 9 will apply that perspective to the question. First what requires discussion are two major assumptions built into chapters two through six; for it is in addressing these assumptions that the need for a complexity framework will be revealed. The first major assumption—the subject of this chapter—concerns the evolutionary perspective adopted in chapters two through
six. The second major assumption concerns the examination of sciences from the standpoint of “heterogeneity.” This assumption is the subject of the following chapter.

The first major assumption is that evolutionary considerations generally situate the (heterogeneous) sciences of emotions. To retrace our steps, in Chapter 2 it was outlined how biology is a model science regarding heterogeneity. From biology-related considerations the broader four-dimensional framework for thinking about “science”—from the standpoint of heterogeneity—was produced. The framework was then applied to four major emotions research programs. But what justifies the application in the first place? Besides appealing to heterogeneity, the application is based on a series of major parallels between “biology” and “psychology” (recall Griffiths’ diagram—Figure 1—given at the beginning of Chapter 2). In this chapter, section 7.1 expands upon the parallel between biological and cognitive psychological levels of investigation. The parallel matters since many of the methods and explanatory resources of biology and cognitive psychology are also employed by the four emotions research programs.

Additionally, in chapter two it was mentioned that there is a significant dissimilarity between biology and psychology. Section 7.2 discusses the dissimilarity, which concerns what biology has—and psychology apparently lacks—namely a consilient framework. I will expound a dispositional interpretation of this framework in accordance with the “received view” of evolution (i.e., the Modern Synthesis). Then in section 7.3 I proceed to link the received view with the functional approaches to emotions.

However, there is a contemporary tension in the evolutionary framework, coming from the “Developmentalist Challenge.” A corresponding link will also be made between this challenge and the developmental emotions research programs of chapters five and six. The challenge is covered in 7.4. The link is discussed in 7.5 (Figure 13), where Griffiths’ diagram in

---

1 I am indebted to Cornelius (1996), who identifies the four programs (although the four programs as presented in chapters three through six differ from his presentation).
7.1—the parallel between the received view and cognitive psychology (Figure 11)—is modified in accordance with developmental biology, and developmental-social views in psychology (which differ in emphasis from cognitive psychology). More broadly, the tension in biology between the “functionalists” (who defend the received view) and the “structuralists” (who defend the developmentalist view) has a philosophical analogue. This analogue, which I discuss in Chapter 8, bears upon the second major assumption and the need for a complex-systems perspective.

7.1: The Parallel Between Biology and Cognitive Psychology

Three Reasons for the Parallel Between Biology and Cognitive Psychology

There are three prominent (and interrelated) reasons for the parallel. The first is that the predominant “method” of cognitive science, laid out in David Marr’s work Vision, employs the computational, algorithmic, and implementational levels, which conjointly delineate a particular brand of reverse engineering (recall section 3.2). (Note that the Griffiths’ diagram employs the ecological, computational, and implementational levels of analysis. These terminological differences were discussed in section 4.2, in relation to the functional stance.)

The second reason is that functional analyses in biology make use of a related—though different—brand of “reverse engineering.” More precisely, biology’s more robust understanding of “function” concerns synchronic and diachronic factors in reverse-engineering putative adaptations. Whereas Marr’s three levels focus mainly on synchronic factors, biological reverse engineering integrates synchronic and diachronic factors in attempting to understand putative adaptations/functions.

And thirdly, the study of function, despite its different uses in biology and in cognitive psychology, serves as the conceptual “hub” organizing the research methodologies for both disciplines (even stronger, the explanatory resources of biology and cognitive psychology issue
from the concern with function). Griffiths' diagram, presented in Chapter 2, was modified to provide a quick illustration of the parallel between biology and cognitive psychology. The following is an accurate representation (1997, p.221):

<table>
<thead>
<tr>
<th>Biological Explanations</th>
<th>(Cognitive) Psychological Explanations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population Dynamic Level</strong></td>
<td><strong>Ecological Level</strong></td>
</tr>
<tr>
<td>Traits classified solely by relative fitness functions. Explanation by &quot;consequence laws.&quot;</td>
<td>(Level of Task Description)</td>
</tr>
<tr>
<td><strong>General Ecological Level</strong></td>
<td></td>
</tr>
<tr>
<td>Traits classified by the adaptive problem that they solve. Explanation involves &quot;source laws.&quot;</td>
<td><strong>Computational Level</strong></td>
</tr>
<tr>
<td><strong>Natural Historical Level</strong></td>
<td></td>
</tr>
<tr>
<td>Traits classified by homology. Explanation by historical narratives.</td>
<td><strong>Implementational Level</strong></td>
</tr>
<tr>
<td><strong>Anatomical Level</strong></td>
<td></td>
</tr>
<tr>
<td>Traits classified by their physical capacities</td>
<td></td>
</tr>
</tbody>
</table>

Figure 11. The Parallel Between the Received View and Cognitive Psychology

The Psychological Side of the Diagram

The first reason for the parallel between biology and cognitive psychology focuses on the psychological side of the diagram. Marr's three levels of analysis generally map to the three levels in the diagram (although the Ecological Level is "richer" than Marr's highest level; see section 4.2, "Terminological Clarifications"). The above "Computational Level" delineates a physical platform's design, relative to some higher-level task (the "Level of Task Description"), and the "Implementational Level" physically instantiates that design. In relation to the emotions research programs discussed in chapters three and four, the functional approaches roughly map to the Computational and Implementational Levels (see sections 3.2 and 4.2). The highest level is the "Ecological Level." This level inquires into the ecological roles that a design plays. As
ecology is the study of organism-environment interactions, this level of “task description”
emphasizes the broader roles of the architectures provided by the computational level.

*Concerning the research programs in chapters five and six, these programs roughly map to the Ecological Level* (although note that I provide an amended diagram in 7.5—Figure 13—since these developmental-social programs do not squarely fall within cognitive psychology). The mapping between the functional approaches and the computational-implementational levels is a matter of emphasis, as is similarly the case concerning the mapping between the developmental-social approaches and the ecological level. For these approaches appeal *differentially* to the three levels, since the modal activation of emotion systems taps differentially into levels of analysis.²

From the perspective of our four-dimensional space, the implementational level corresponds to
the concrete causal-phenomenological level; the computational level corresponds to the semi-concrete level; and the ecological level corresponds to the more abstract level. The three
stances with their 4-D shapes and overlapping regions further reinforces that the three
psychological levels ought not to be interpreted as fixed levels of analysis.

**The Biological Side of the Diagram**

The second reason for the parallel between biology and psychology deals with the biological side of the diagram as it exhibits “reverse engineering.” The first three biological levels—the “Anatomical Level,” the “Natural Historical Level,” and the “General Ecological Level”—are the primary focus of (biological) reverse engineering. In a nutshell, reverse engineering is situated within Mayr’s proximate-ultimate distinction mentioned in Chapter 2 (section 2.1). **Proximate mechanisms** provide answers to questions about how biological processes work; asking why those mechanisms evolved appeals to ultimate causations.

---

² The idea of emotions-as-modes runs through chapters three through six; “modal activation” will be more fully discussed later in this chapter, and also in chapters eight and nine.
The proximate causes of an organism's traits occur within the lifetime of the organism. They involve the expression of the information contained in the organism's genetic material, as mediated by the environment. The ultimate causes occur prior to the lifetime of the organism, within the evolutionary history of the organism's species. They involve the reasons why members of that species have come to have the genetic information that they do (Beatty 1994).

The proximate-ultimate distinction has been used for several purposes in defending the received view of evolution; I discuss some of these uses in the following section. The salience of the distinction as it relates to the three levels is that the anatomical level maps to the investigation of proximate mechanisms (e.g., physiological studies describing how the eye operates), and both the general ecological level and the natural historical level involve the search for ultimate causations. The general ecological level is about particular "source laws" (expressed via models), which provide the sources of relative fitnesses in an abstract population. The point of such models is to find the "inputs" constraining population dynamic models. The "general ecological descriptions are then realized in particular cases by particular lineages of organisms or traits" (Griffiths 1997, p.218); this natural historical level provides further (historical) constraints via the search for homologies (or analogies) and the different adaptive problems they solve:

The flukes of whales, the tails of fish, the feet of seals, and the wings of penguins are different lineages' solutions to the problem of exerting muscular force on a liquid medium. They are cladistically distinct, but functionally identical. Kinds at the natural historical level enter into natural historical narratives [emphasis mine] about the evolution of particular lineages. Source laws [at the general ecological level] and consequence laws [at the higher population dynamic level] concerning relative fitness figure in these narratives, but so do particular historical facts that affect the outcome of the process (ibid).

---

3 I defer discussion of the highest level to section 7.2.
4 The language of "source laws" is Sober's (1984). Source laws describe the "circumstances that produce forces [dispositions]" (p.50); in other words, the ecological parameters describe the abstract circumstances (time-energy budgets/investments, etc.) that frame possible adaptive landscapes. An example of a source law is Fisher's sex-ratio argument for why the sexes tend to be distributed in a roughly 50-50 manner (see Sober 1984, Ch.1). This source law provides "initial conditions" for selective pressures, but it does not describe how evolution occurs within an adaptive landscape (the population dynamic level provides models for this purpose). Thus source laws are distinguished from the highest level, the "Population Dynamic Level," which appeals to "consequence laws" that "describe how forces, once they exist, produce changes in the systems they impinge upon [i.e., how they produce changes in adaptive landscapes]" (p.50).
These evolutionary constraints expressed by the first three levels delineate "layers of function": the general ecological level searches for sources constraining the possible functions which traits serve; the natural historical level searches for historical facts further cashing out abstract ecological functions, resulting in a historical narrative for a particular lineage (comparing homologies within a lineage, or analogies between lineages); and lastly the anatomical level examines physiological functions of current (or extinct) instances of a lineage, and (ideally) the manner in which these particular functions interrelate to the previous two levels. Broadly construed, the three levels are just the moments of "reverse engineering" situated within the received view of evolutionary biology. In other words, comprehending the layers of function—understanding them individually at each level and jointly between levels—is the project of (biological) reverse engineering. ⁵

The Parallel: Two Interrelated Senses of "Function"

The last reason for the parallel between biology and psychology is the concern with function, for the study of function serves as the conceptual hub organizing the research methodologies. The prior two reasons are about the broadly similar—though contextually different—employments of reverse engineering; that is, reverse engineering makes sense for both biology and cognitive psychology because of the search for "function." Specifically, cognitive psychological "function" is situated by the three levels of analysis (see section 3.2 on synchronic reverse engineering; recall also fn.10 on "Cummins function"), and biological "function" is situated by the four levels of analysis (see fn.6 below; see also section 4.2 on the functional stance as applied to synchronic and diachronic/adaptive processes).

⁵ Note that biological reverse engineering places less emphasis how forces evolve (i.e., the population dynamic level); the foci are on the "causal roles" of forces at the anatomical level, the historical constraints on forces at the natural historical level, and the abstract "source" constraints on forces at the general ecological level.
In addition, the search for function matters since these disciplines' claim to autonomy is grounded in the non-reducibility of function. For biology, recall from chapter two that biology's autonomy issues from its historical dimension. More accurately, while proximate explanations may sometimes be reducible to explanations from the physical sciences, ultimate causations above the anatomical level are what make biology a different kind of science. The search for function at the higher levels, *in conjunction with* the search for function at the anatomical level, lend credence to biology's (semi) autonomy. As for cognitive psychology, the search for function occurs at multiple levels; the claim to autonomy is based upon the "multiple realizability" of psychological states (see Chapter 1 on peripheral and centralized systems; see also Chapter 8). That is, states at the implementational (neural) level do not by themselves assist in delineating higher-order functional "kinds" which the computational and ecological levels reveal.

In short, the explanatory and methodological resources of both biology and psychology are based upon the search for function, and the search for function, in turn, supports the autonomy of "special sciences" like these. (I cover the philosophical debate that underscores the importance of function in chapter eight.)

**Relating the Diagram to the Functional Approaches to Emotions**

Hence the parallel between cognitive psychology and biology issues from 1) reverse engineering—situated within the three psychological levels—as the conceptual method for understanding cognitive psychological systems; 2) (biological) reverse engineering—situated within the four biological levels—as the conceptual method for understanding biological systems; and 3) the "similar" concern of biology and psychology with "functional explanation."\(^6\) Applied to our functional emotions research programs, *Ekman's program generally maps to the*

\(^6\) Note that the parallel does not imply that biology and cognitive psychology are "equivalent," for as Griffiths notes, if cognitive psychological "function" is given a narrow philosophical interpretation, it differs markedly from biological "function" (especially concerning biology's historical constraints). See Griffiths (1994); for more on biological function, see also Hull and Ruse (1998).
Computational and Implementational Levels on the side of (cognitive) psychology, since affect programs are Fodorian modules. (To a lesser extent Ekman's program maps to the natural historical and anatomical levels on the side of biology, as the evolutionary dimension of affect programs is acknowledged but placed in the background.) Evolutionary psychology generally straddles the psychological and biological sides; the primary focus is on the Computational Level (producing boxologies) and the Natural Historical Level (relating to the EEA, for example). 7

While the parallel between biology and cognitive psychology is determined by these three prominent reasons, there is a significant dissimilarity between biology and psychology, indicated by the lack of a corresponding fourth level on the side of psychology. The population dynamic level in biology is part of what gives biology its "unified" framework. In spite of the fact that biology is the science of exceptions, its heterogeneous subject matters hang together by virtue of the received view's consilience (thus the oft cited phrase, "nothing in biology makes sense except in the light of evolution"; see Chapter 2 fn.5, and section 7.2 below). This observation is important for our purposes since if there is a genuine possibility for a science of emotion, given that emotions are heterogeneous things varying by feel, operation, (formal) object, time, development, and so on, the question is whether such heterogeneity can likewise "hang together" by means of a consilient framework (thus making sense of the general category of emotion—if it even exists). To properly explore this question, a further investigation of the interrelations between the highest level in biology and the other levels is required, as it will provide crucial clues in the search for a possible science of emotion.

7.2: Evolution and Function

7 To reiterate, I will discuss an amended parallel between developmental biology and developmental views in psychology in 7.5.
To take stock, the autonomy of biology is based upon the proximate-ultimate distinction. This distinction also frames the search for biological function at multiple levels. And since the received view of evolution provides the theoretical background for the functional approaches to emotions, the parallel between biology and cognitive psychology is presupposed by these functional approaches. However, within biology there is another research program, evolutionary developmental biology (alternatively called "EvoDevo," or "Developmental Systems Theory"), which exhibits tangled relations to the received view, since DST challenges the proximate-ultimate distinction, as well as other cornerstones of the Modern Synthesis. While DST is more fully covered in section 7.4, certain aspects will be briefly outlined in the ensuing critique. That is, in this section I examine the interrelations between the highest level (the population dynamic level) and the received view; these topics will be expounded via a critique of the proximate-ultimate distinction, and the responses on Mayr's behalf to the objections.

Ariew's "Reconstruction" of Mayr

The specific critique of Mayr’s proximate-ultimate distinction, provided by André Ariew (2003), aims to reconstruct this distinction. Ariew begins by critiquing the proximate side of the distinction: Mayr’s proximate classifications answer how-questions, which are subsumed under the field of “Functional Biology,” but this actually conflates two distinct causal processes, namely developmental causation (how does something come to be?) and physiological causation (how does something operate?). Ariew’s subsequent amendment—the first part of his “reconstruction”—distinguishes between these two types of causations. However, in Mayr’s defense, Mayr does recognize the different types of questions, and places the former question under functional accounts, emphasizing that developmental questions still require an account of

\[^{8}\] Note that strictly speaking, EvoDevo is the modern practice of evolutionary developmental biology, and "DST" is the inclusive term for the practice (research program) and its philosophical framework. Keeping in mind these distinctions, I use "DST" as shorthand for “EvoDevo.”
what role development plays. That is, cashing out “how something comes to be” is organized around the primary aim of understanding “how something functionally operates,” relative to the anatomical level of investigation.

This response actually obscures what the Developmentalist Challenge claims. More precisely, to claim that developmental causation is subsumed under functional accounts misses the point of the challenge. Mayr, though, is aware of the nature of the challenge; his reason for subsuming developmental causation is not just to highlight function (adaptive or otherwise), whose proximate mechanisms may be multiply realized by developmental processes. Rather, Mayr’s employment of the proximate-ultimate distinction in his later career as a philosopher and historian of biology (Beatty 1994) defends the Modern Synthesis against certain “revolutionary” elements of the Developmentalist Challenge (Amundson 2005). Mayr’s defense is that function renders development largely irrelevant in understanding the architecture of evolutionary thought. The claim is not that evolutionary developmental biology isn’t biology, nor that EvoDevo doesn’t make legitimate contributions to the received (Modern Synthesis) view of evolution. Rather, the contributions of EvoDevo still fall under either proximate causations, or under ultimate causations (or a mix of both). The search for biological function, “properly” understood, leaves the received view intact, and also situates EvoDevo as a branch of the received view.

The Reconstructed Proximate-Ultimate Distinction

Ariew likewise concentrates on the architectural level. However, his critique of the proximate-ultimate distinction attempts to take seriously not only the received view’s population dynamic level, but also the Developmentalist Challenge. Thus his reconstruction of the proximate-ultimate distinction proceeds as follows: first, Ariew claims that the proximate side of the distinction really applies to individual level causal events; he then proposes the distinction between individual level causal events (i.e., developmental and physiological causations) and
statistical level events (Ariew's reconstruction of ultimate explanations). Upon inspection, though, I find that Ariew's exposition of Mayr, and the subsequent objections raised concerning Ariew's aim to reconstruct, do not do full justice to Mayr. The trenchant nature of the proximate-ultimate distinction can still be maintained if one interprets Mayr in a suitably charitable manner. A response to Ariew's critique of the proximate side of the distinction has been discussed above; the other side now requires consideration.

Reconstructing the Ultimate Side of the Distinction

There are four theses Ariew attributes to Mayr which supposedly justify the ultimate side of the distinction. "1) Natural selection is the sole explanation for nature's diversity, hence 'ultimate' refers to natural selection explanations only. 2) Ultimate qua evolutionary explanations are essentially historical. 3) Since ultimate qua evolutionary explanations are essentially historical, they are not purposive. 4) 'Ultimate' like 'proximate' refers to individual level causal processes. I disagree with all of that" (p.557). Ariew disagrees with the first thesis because there are other legitimate evolutionary explanations appealing to drift, migration, etc. (p.558); hence natural selection is not the "sole explanation for nature's diversity." Response: In Mayr's essay "How to Carry Out the Adaptationist Program?" (1988) Mayr notes that there is immense heuristic value in assuming X is an adaptive trait produced by natural selection; the method has generated (and continues to generate) fruitful results (p.153).

Basically, the methodology consists in establishing a tentative correlation between a trait and a feature of the environment, and then to analyze in a comparative study, other organisms exposed to the same feature of the environment and see whether they have acquired the same specialization. There are two possible explanations for a failure of confirmation or correlation. Either the studied feature is not the result of selection force or there are multiple pathways for achieving adaptedness.

---

Note that methodological adaptationism makes no claim to being "universally valid" in ontological or metaphysical terms. That is, Mayr nowhere claims that all characters of organisms (ontologically) are adaptations, nor that anywhere the general evolutionary process operates (metaphysically), adaptations will be produced. For types of "adaptationisms," see Godfrey-Smith (2001).
When the comparative study results in a falsification of the tentative hypothesis, and when other hypotheses lead to ambiguous results, it is time to think of experimental lests. Such lests are not only often possible but indeed are now being made increasingly often, as the current literature reveals. Only when all such specific analyses to determine the possible adaptive value of the respective trait have failed is it time to adopt a more holistic approach and to start thinking about the possible adaptive significance of a larger portion of the phenotype, indeed the possibility of the Bauplan as a whole.

Thus, the student of adaptation has to sail a perilous course between a pseudoexplanatory reductionistic atomism and a stultifying nonexplanatory holism. ...[W]hen we look at Gould and Lewontin's 'alternatives to immediate adaptation,' we find that all of them are ultimately based on natural selection, properly conceived. It is thus evident that the target of their criticism should have been neither natural selection nor the adaptationist program as such, but rather a faulty interpretation of natural selection and an improperly conducted adaptationist program. Gould and Lewontin's proposals [e.g., drift] are not 'alternatives to the adaptationist program,' but simply legitimate forms of it [emphases mine] (pp.154-55).\(^{10}\)

As for the second thesis, Ariew claims that Mayr is overly restrictive concerning the historical domain of evolution. For what about traits currently undergoing evolution? Ariew speculates that the rationale for the second thesis comes from Mayr's rejection of "prediction" in evolution.

Response: While Mayr does claim that a retrospective view is required in order to understand evolution in general, he does not preclude local predictions based on models employing particular statistical-biological "laws." Moreover, the predictions contested are the sharper (usually quantitative) predictions found in the physical sciences. Ariew's second thesis is based on a false target. Regarding the third thesis, Ariew makes a number of points (see pp.559-60) that are puzzling precisely because they are the sorts of claims which Mayr, sympathetically interpreted, would subscribe to. I put on the side an exposition of, and response to, this thesis, and simply refer the interested reader to Mayr's paper "The Multiple Meanings of Teleological" (1988).

Does Evolution Range Over Individual Level Causal Properties?

The fourth thesis lies at the core of the critique, since it serves as the "foil" for Ariew's statistical reconstruction of Mayr's ultimate explanations. Ariew's objection stems from a

\(^{10}\) For what a complete adaptive explanation would require, see Brandon (1990), pp.165-76. I briefly expand on this point in 7.3.
position attributed to Mayr, namely that “evolution is a ‘dynamical’ process, meaning that it ranges over individual level causal properties” (p.560). This would be partially true\(^{\text{11}}\) with respect to Mayr’s claim that the primary target of selection is the organism, which has “causal properties” insofar as it acts-in-the-world. Given that the individual is the target of selection, the process of evolution “ranges over” individuals which differentially act in an environment. But Ariew denies that such causal processes constitute an evolutionary explanation—“evolutionary explanations range over statistical attributes of a population, not dynamical properties of individuals” (p.560). The problem, as Ariew sees it, is that if populations are just aggregations of individuals—a position attributed to Mayr—then evolutionary explanations are reducible in principle to explanations involving proximate causes of how organisms are selected relative to their life-histories. And if this is so, how is an ultimate explanation really different from an enumerated proximate explanation? In other words, just where would the proximate–ultimate distinction lie? To further clarify the problem, since all individuals are different in some way, even if an enumeration of all the individual-proximate causes is given, the question remains, what do some of these individuals have in common in spite of their variegated life-histories? Since apparently Mayr focuses on the individual as the target of selection, the ability to delineate commonalities in spite of particular variations is obscured; it appears as if a better working distinction is required. Ariew’s fix is a statistical appeal. What individuals have in common is identified by “statistical properties of an evolving population. Hence, evolutionary explanations differ in kind from proximate explanations” (p.561). As an example, Ariew claims that drift and trait fitness are two different evolutionary explanations since they “represent different statistical-level events”—that is, “trait fitness is the expected survivability and reproductive rate of success of individuals sharing a common feature. Drift is the deviation of the actual outcome from the..."
expected" (p.563). These explanations appeal to statistical-ensemble patterns which proximate explanations miss. In sum, the accusation is that Mayr's ultimate explanations really boil down to fitness propensities of individuals; and while we might explain such propensities by appealing to mechanisms like trait fitness from an individual-proximate point of view, this would be misleading, as trait fitness is actually a statistical-level pattern. Statistical-level explanations are different kinds of evolutionary explanations—something that Ariew can account for, but which, apparently, Mayr cannot.

Defending Mayr

While I am sympathetic to the tenor of Ariew's critique, since Mayr generally does not give sufficient credit to the use of statistical techniques as they have matured in mathematical biology, the claim that ultimate evolutionary explanations are in-principle reducible to individual-proximate mechanisms (or at least open to the charge of such reducibility) is problematic. For even if statistical models are "in-principle" reducible to their lower level(s), it does not follow that population systems fail to have "emergent" phenomena. That is, aggregations of individuals from a population standpoint—an important part of what Mayr's ultimate explanations emphasize—may have qualitatively different characteristics which cannot be practically explained from the individual level. Furthermore, Ariew's objection that proximate explanations are different in kind from statistical explanations is something Mayr would agree with (recall from section 2.1 that biological concepts cannot be reduced to lower-level explanations); Mayr's disagreement would probably stem from the use of mathematical tools as substitutes for qualitative concepts and the importance he places on (proper) evolutionary narratives (as

---

12 Mayr, though, does recognize the importance of statistical thinking, but only as subsumed under populational thinking. What he criticizes is the use of statistical models (with highly artificial assumptions) that are taken as "real"—what he disparagingly calls "bean-bag genetics."

13 For Mayr, populational thinking is about the variation between actual individuals; statistical tools used to extract "types" in a population are just abstractions. Accordingly, we ought to distinguish the real objects of biological study and the mathematical tools for modeling variation.
conceptual structures). I think a charitable construal of Mayr's views would hold that modern sophisticated employment of statistical models is useful, but that they fall under the broader search for evolutionary narratives and the problems associated with cashing-out narratives. For the natural historical level's employment of narratives is the hub organizing the search for function at the higher and lower levels (of Griffiths' diagram).

An Architectural Overview of the Received View: A Propensity Interpretation of Fitness

It was mentioned above that according to Ariew, the ultimate side of Mayr's distinction "boils down" to fitness propensities of individuals. Although this charge is too strong, there is a sense in which it is correct, since individual organisms are the "ontological stuff" of biology. Accordingly, the next topic discusses how fitness can still be given a propensity interpretation. First, it should be noted that Ariew's statistical-level explanations do not preclude a propensity interpretation. However, the contrast he draws between dynamical fitness propensities of individuals and non-dynamical statistical events (p.564) makes it difficult to see how that interpretation would be compatible. Perhaps by "non-dynamical" he means events that do not depend—at least with respect to coarse-grained notions dealing with averages, deviations, etc.—on the particular life-histories of individuals, which would be compatible with a limiting-frequency interpretation of probability (and so compatible with a type of propensity/dispositional account). But Robert Brandon, in "Adaptation and Evolutionary Theory," argues against such an approach to defining "relative adaptedness," an approach that would define fitness in terms of an...

---

14 Cooper's suggestion (from Chapter 2) that ecological studies should use a plurality of models/theories to construct "borderland" accounts interrelating systems ecology and evolutionary ecology can be applied here. The patchwork of theories in evolutionary biology seeks interrelations between proximate and ultimate mechanisms, whereupon consilience is manifested through the use of narrative structures. (Although note that consilience marks an important difference in emphasis from Cooper's concerns, as there may not be consilience between various ecological borderland accounts.)
average reproductive success of a class.\textsuperscript{15} Briefly, the problem with this approach is that it cannot handle certain cases in which individuals would be assigned a higher fitness value when they intuitively shouldn't be.\textsuperscript{16} So what interpretation ought "fitness" be given? Another approach Brandon discusses appeals to certain properties that are independent of reproductive success (thus avoiding the tautology objection; see 1978), but the problem is that "nature has refused to cooperate. There is no property that will do the trick. Put another way, there is no property that is invariably selected for" (1996, p.49). The third approach advocated is a "propensity" approach, as it is based on a propensity interpretation of probability.

The general idea proceeds as follows (1990). For organism $O$ and environment $E$, let there be a range of possible numbers offspring $Q_1^{OE}, Q_2^{OE}, \ldots, Q_n^{OE}$ such that for each $Q_i^{OE}$ there is a probability assignment $P(Q_i^{OE})$ of $O$ leaving $Q_i$ offspring in $E$. Define the adaptedness of $O$ in $E$ as the expected value $A(O,E) = \Sigma P(Q_i^{OE})Q_i^{OE}$ (think: "$A$" for "Average"/weighted reproductive success). Relative adaptedness (RA) is then characterized as: organism $a$ is better adapted than $b$ in $E$ iff $A(a,E) > A(b,E)$.\textsuperscript{17} The claim is that the underlying physical properties of an organism-in-an-environment are expressive of propensities; the propensity interpretation of probability associates probabilities in the expected value with those properties that are causally responsible for observed frequencies when RA is applied.\textsuperscript{18} That is, RA itself is general and has no biological

\textsuperscript{15}For example, take a population where the individuals can be separated into two similarity classes, and where the two classes are homogeneous with respect to reproduction. Fitness for an individual gets defined \textit{in the long run} by a function that associates the average reproductive success of the relevant class to the individual (Brandon, 1978, p.193).

\textsuperscript{16}The full argument is in Brandon 1978, pp.193-5.

\textsuperscript{17}An additional term is subtracted from the original expected value to account for differential selective advantage as it relates to variance in offspring (i.e., frequency-dependent selection), defined by the function $f(E,r^2)$. For the details see Brandon 1990, pp.18-21. Mills and Beatty (1979) also offer another propensity interpretation.

\textsuperscript{18}A terminological point to note is that Brandon distinguishes "propensities" from "dispositions" (e.g., the disposition of sugar to dissolve in water), since dispositions (for Brandon) tend to exhibit deterministic behavior, whereas propensities in natural selection allow for chance events that would "prevent perfect correlation between adaptedness and fitness" (1996, p.49). I depart from Brandon's usage, since
empirical content—it is what Brandon calls a “schematic definition.” Empirical content comes from particular exemplifications of RA where “we fix the value of the environmental parameter E, and limit the range of the individual variables a and b to a particular population [emphases mine] of organisms living in E” (1990, p.21). Then certain kinds of data are gathered to find estimates of the terms in the expected value. Why all this matters is that the core of Darwinian theory is the principle of natural selection, which relates relative adaptedness to reproductive success: “(Probably) If a is better adapted [relatively] than b to their mutual environment E, then a will have greater reproductive success than b (in E)” (1996, p.47). The upshot of the above discussion is that the structure of evolutionary theory has at its core this principle, which has no empirical biological content yet still is the most important part of the theory. The two presuppositions of the principle’s applicability are biological [i.e., (1) biological entities are chance set-ups with respect to reproduction, and (2) some biological entities differ in their adaptedness to their common environment, this difference having its basis in differences in some traits of the entities (1990, Ch.3)]. They [(1) and (2)] form the empirical biological core. Slightly less central is the third empirical biological claim [i.e., (3) adaptedness is to a degree heritable, or equivalently, the causal bases of adaptedness values are to a degree heritable], not a presupposition of the applicability of the principle, but rather a necessary condition of natural selection having evolutionary effect. At the periphery of the theory are the instantiations of the principle of natural selection. They encode the various low-level theories about the evolution of specific populations under specific environmental conditions (1996, p.56).

So it would appear that the structure of evolutionary theory is profoundly “dispositional,” and that the individual, to which dispositional properties are attributed, is the target of selection.

**Summary**

dispositions include capacities and tendencies, both of which are neutral concerning the exhibition of deterministic behavior.

19 Note that a and b occur relative to a population, where the population can be modeled statistically.

20 The two types of data and their associated methods are discussed in 1990, pp.22-23.

21 Note that RA is a schematic definition capturing certain pre-existing intuitions, where instantiations of RA further refine those intuitions (e.g., the continuum between R-selection and r-selection at the general ecological level). The same applies to the principle of natural selection: the propensity interpretation is a placeholder for capturing certain intuitions, and for further refining those intuitions when examining properties that are causally responsible for observed frequencies. The principle of natural selection is an organizational principle whose propensity formulation discloses how that organization occurs.
Combining this dispositional interpretation with Ariew's critique (and the responses to the critique), we have the following characteristics of the interrelations between highest level of the diagram in 7.1 (the population dynamic level) and the received view: 1) explaining function at its multiple levels is organized around the search for evolutionary narratives; and 2) narratives “feed up” into the higher levels—the population dynamic level and the general ecological level—through the use of statistical models that model individual propensities in a population. In conjunction with these two points, the overall picture of the received view additionally holds that 3) narratives “feed down” to the anatomical level through the examination of individual causal level events (functional and developmental causations); 4) the proximate-ultimate distinction conceptually encompasses the prior three points; 5) narratives, at the philosophical level concerning the architecture of evolutionary theory, are framed by the propensity interpretation of natural selection and the (empirical) presuppositions mentioned in the previous passage; and 6) also at the philosophical level, narratives utilize the proximate-ultimate distinction to pinpoint what is distinctive about biology as an autonomous, historical science.

7.3: The Received View and Functional Approaches to Emotions

A Complete Account of an Adaptive Explanation

Since the received view places narratives at the conceptual center of its search for function, it is natural to ask what constitutes a complete account of a narrative about function. In other words, what constitutes a complete account of an adaptive explanation? Brandon (1990) lists five criteria (see pp. 165-76 concerning the justification for, and exposition of, each criterion):

an ideally complete adaptation explanation requires five kinds of information: (1) Evidence that selection has occurred, that is, that some types are better adapted than others in the relevant selective environment (and that this has resulted in differential
reproduction); (2) an ecological explanation of the fact that some types are better adapted than others; (3) evidence that the traits in question are heritable; (4) information about the structure of the population from both a genetic and a selective point of view, that is, information about patterns of gene flow and patterns of selective environments; and (5) phylogenetic information concerning what has evolved from what, that is, which character states are primitive and which are derived (p.165).

Note that points (1) and (2) cohere with the natural historical level and the general ecological level. Point (4) maps to the general ecological level and the population dynamic level. Point (3) may appeal to the anatomical level in explaining the “mechanics” of heritability. Lastly, (5) is perhaps the most important criterion, since it emphasizes the natural historical level and the constraints imposed upon constructed narratives.22

Linking the Parallel Between the Received View and Cognitive Psychology with the Functional Approaches to Emotions

A fair amount of time has been spent on the structure of the received view of evolution. With that machinery in place, I can now draw implications for the functional approaches to emotions. As noted in section 7.1, Ekman’s program maps to the computational and implementational levels for cognitive psychology, and to a lesser extent it maps to the natural historical level and anatomical level for biology. Evolutionary psychology straddles the psychological and biological sides of the diagram; the primary emphases are on the computational level (producing boxologies) and the natural historical level (relating to the EEA, for example). In chapters three and four it was argued that what classifies Ekman’s program and evolutionary psychology as functional approaches to emotions is the search for FIIPS. Yet in spite of the common search for FIIPS, the question whether there can be a science of emotion has different answers for each program.

Ekman’s Program

22 This criterion expresses the distinction between homologous traits and analogous (“convergent”) traits. What might be taken for convergent evolution might actually be a result of traits stemming from the same primitive characters. Without examining the phylogenetic constraints for “how evolution tinkers,” one cannot properly construct plausible narratives.
Regarding Ekman's research program, affect programs technically define what "the emotions" are. On a narrow interpretation, this means not only that a science of emotion is possible, but it has already been outlined. Affect programs (currently) yield the best scientific account of emotions, and perhaps this account will stand as the only viable science of emotions. Thus the general category of emotion is just the enumeration of affect programs. All other colloquial or higher-order employments of "emotions" are the result of display rules; the result of display rules juxtaposed with particular affect programs; the result of display rules juxtaposed with combinations of affect programs; and so on. The point is that affect programs, as a (projectible) scientific concept of emotions, are partitioned from other senses of "emotions." If this narrow picture is correct, then we have a rather flat answer to our question whether there can be a science of emotion. This would be an answer by technical default.

However, on a less narrow interpretation, the scope of emotions-as-affect-programs is merely the restriction of these modules to their Fodorian characteristics; the scope exhausts and maps emotive phenomena in so far as we search for Fodorian characteristics of affect programs. But such a qualification suggests that affect programs do not exhaust the general category of emotion, as there are other possible scopes (beyond the concern with Fodorian characteristics) which future sciences of emotions may provide. And this possibility would suggest that one cannot "define away" the problem by restricting emotions to the narrow characteristics of affect programs. Indeed, evolutionary psychology's extended modular approach and the general language of FIIPS indicate that the category of emotion is wider than the restriction to affect programs.

Evolutionary Psychology

Evolutionary psychologists generally subscribe to the received view of evolution. The new move on their part (the Santa Barbara School in particular) is extending the search for
function to psychological function. Normal evolutionary functions have anatomical or behavioral counterparts. While psychological functions exhibit aspects of these counterparts as well, the difficulty is that psychological functions occur in the "realm of the mind." Unlike Ekman's program, which emphasizes the computational and implementational levels in cognitive psychology, evolutionary psychologists emphasize boxologies straddling the natural historical level and the computational level. This mid-level orientation is appropriate since psychological functions are multiply realizable by the lower anatomical and implementational levels (thus Griffiths has called these functions "virtual modules"). Additionally, the mid-level orientation stems from biology's emphasis on narrative structures at the natural historical level. In other words, regarding the computational level, the focus on boxologies is just the first empirical step to see if functional regularities "exist" (i.e., there are robust correlations indicating genuine phenomena). The hard part is to find the relevant evolutionary constraints at the natural historical level that provide a plausible range of historical narratives for why a FLIPS may have evolved. It is this latter move that many critics of evolutionary psychology think is problematic, if not impossible in principle.

In my estimation, three responses need to be given. Firstly, the charge is a "blanket charge." The proper charge would be to indicate which of the five criteria for an ideal adaptive-functional narrative are lacking in evolutionary psychology, *keeping in mind that this is a young research program*. The second response is that indicating which of the five criteria need further investigation is *not* the same as claiming that evolutionary psychology in-principle cannot meet such an ideal. Evolutionary narratives in general do not live up to the ideal, but to claim that this is the fundamental problem with evolutionary theorizing is the wrong attitude to adopt. Such an attitude, if anything, indicates that one does not properly understand why biology is a different
kind of science. (And no matter what critics may say, the starting point for serious inquiry into this topic is that biology is a science; no charitable interpretation assumes less.)

Thirdly, in tandem with the recognition that evolutionary psychology is a young research program, the real issue is the extent to which evolutionary psychology currently meets each of the five criteria in its search for psychological function. A strong critic might claim that none of the five criteria can be fully met, making irrelevant the issue of what the program currently does. This charge would be unsafe. Witness, in general, the spotty record of philosophical prognostications about what science cannot do. Additionally, the charge would be beside the point, since many evolutionary explanations don’t meet the ideal and yet are considered to be plausible working hypotheses. A more tempered critique is required. A weak critic could challenge the degree to which any of the criteria are “reasonably” approximated by the current state of evolutionary psychology. One response is that such criticism would simply be legitimate criticism in the arena of scientific dispute. (Recall Mayr’s warning: the “just-so” charge should be used with caution, and should be used to discount improperly conducted adaptive explanations, that is, explanations that hastily preclude other levels of function and the interrelations between levels.) The other response is that as a young research program, we have yet to see the degree to which evolutionary psychology will meet the challenge. “Tempered” criticism is fine, but don’t squelch the program and what it might have to say and offer.

Thus as a young research program, evolutionary psychology’s search for FIIPS is just getting underway. The extent to which FIIPS can account for all (or very many) centralized processes is unknown. And while evolutionary psychology assumes that the mind consists mainly of FIIPS, this assumption is really a starting point for investigating the “frontiers” of central processes. These processes make no hard divisions between “emotive” and “cognitive” processes, which implies that classifying FIIPS as emotion-systems or as cognitive-systems
employs misleading conceptual categories. For remember that emotions, indeed all mental processes, are *modes of operation*—modes that are understood through the conceptual features of FIIPS. Regarding our question whether there can be a science of emotion, this also implies that such a question is not judiciously phrased; it ought to be *subsumed* under the question, is the mind “nothing but” FIIPS? If it is, then a *corollary* would be that a science of emotion is possible. In addition, there would be no need for a highest level in psychology (corresponding to the population dynamic level in biology) since psychology would actually be an extended branch of biology. The category of emotion would “hang together” through the consilience of function within and between biological levels as applied to “psychological function.” However, the general problem, to repeat, is that the bounds of this approach are currently not known.

**Summary**

Taking stock, this section provides a sketch of the relations between functional emotion research programs and evolutionary considerations. I will address further implications for a science of emotion in the final chapter. The diagram in 7.1 that frames the interrelations between biology and cognitive psychology has been used thus far to cash-out the interrelations between the *received view* of evolution and the functional approaches to emotions. The diagram will also be of assistance in organizing the relations between evolution and the developmental-social approaches to emotions. More fully, the following section examines the Developmentalist Challenge to the received view; section 7.5 then examines the parallel between evolutionary developmental biology and the developmental-social approaches to emotions.

**7.4: Function Versus Structure**

It is important to note that the “Structuralist”/Developmentalist Challenge to the Modern Synthesis is *still* an evolutionary view. Philosophically, *Developmental Systems Theory*
challenges the proximate-ultimate distinction and the understanding of function following from the distinction. There are two particular theses that challenge the received view: 1) natural selection might not be the most important mechanism in the evolutionary process; and 2) what evolves are not populations (as the Modern Synthesis holds), but rather ontogenies (Amundson 2005). In this section I will outline from an “architectural” perspective the nature of the challenge, and the tangled relations between the received view and evolutionary developmental biology.

Developmental Biology and the Modern Synthesis

Recall from chapter four that DST's conceptual structure has two coupled “differential equations” which track the changes on the organism and the changes on the environment. Adding to this picture as it relates to the received view of evolution, as a first approximation, developmental biology is situated between “Functional Biology” and evolutionary biology, where “Functional Biology” refers to the field which studies the proximate mechanisms pertaining to physiology and the like (the anatomical level), and evolutionary biology studies the ultimate causes pertaining to the three higher levels (the natural historical level, the general ecological level, and the population dynamic level). Historically, developmental biology was “absorbed” by the Modern Synthesis, as its concepts and tools were subsumed under the study of proximate mechanisms. More precisely, two things happened: the significantly different concepts and tools of developmental biology (e.g., “morphogenetic field,” “homology,” and embryological studies in general) were jettisoned and then restructured in accordance with “proper” genetic studies; and secondly, the restructuring was interpreted at the philosophical level as the appropriate manner by which to view developmental biology—it is a branch of biology investigating proximate causes. Informally, developmental biology was “hijacked” by genetics (embryological studies are really studies about gene expression), and history was then rewritten by the Modern Synthesis “victors”
(developmental biology is *just* a branch of the received view of evolution, tucked away under the “Department of Proximate Investigations”).

**EvoDevo’s Architecture**

As Amundson (2005) argues, EvoDevo—the modern extension of developmental biology—is no longer merely a “theory”; it is an established research program with a robust body of experimental results. These results come from *actual* genetic and cellular studies made salient by technological advances within the past twenty years or so (which is ironic given the contrast to population “genetic” studies of earlier Modern Synthesis proprietors, for whom “genes” to a significant extent were quasi-mathematical *abstractions*). However, as mentioned in chapter five, DST (see fn.8 above) is also a highly philosophical framework whose conceptual categories overstep its body of (strongly causal) experimental evidence. Yet in spite of its highly philosophical perspective, DST is still included within the evolutionary framework, broadly construed. Since the Developmentalist Challenge challenges the received view of evolution, it needs to be clarified in what sense DST is still included within “the” evolutionary framework. Revising Mayr’s separation of biology into “Functional Biology” (proximate causes) and evolutionary biology (ultimate causes), developmental biology is now situated between the two categories; DST subsequently encompasses developmental biology and a reconstituted sense of “Evolutionary Biology.” DST’s architecture is given below (Gilbert et al. 1996):

| Functional Biology = anatomy, physiology, cell biology, gene expression |
| Developmental Biology = δ [functional biology] / δt |
| Evolutionary Biology = δ [developmental biology] / δt |

Figure 12. The Architecture of EvoDevo

Notice that evolutionary biology is reconceptualized as the study tracking the changes in developmental resources. By comparison, the received view would claim that developmental
biology—whether a part of “Functional Biology” or separate from it—still falls under the search for proximate causes, and as such, is significantly different from the study of ultimate causes. For the study of ultimate causes is not merely the study of changes in proximate mechanisms.

**Differentiating DST and the Received View**

This last point is important since it expresses the key difference between the received view and DST. For the received view, ultimate causations track functional regularities at the natural historical level, whose narratives ought to cohere with the other levels above and below this level. Remember that for Mayr, developmental changes are largely irrelevant in constructing narratives about function. The reason they are mostly irrelevant is because natural selection is the primary driving force in evolution. Developmental changes may produce various phenotypes, but what matters is that the functions that phenotypes serve within a population lead to differential rates of survival. That is, the primary target of selection is the individual organism, and the disclosure of evolutionary trends occurs at the population level. Population dynamic models delineate how these large-scale trends occur, where *epistemically speaking* populations evolve, not individuals (cf. Amundson 2005, p.255). And in conjunction with this epistemic point, *ontologically speaking* individuals are actually selected—they are the main targets of selection. Both of these points, to reemphasize from section 7.2, are just the dimensions of a proper understanding of evolutionary narratives and the search for function.

The developmentalist view of evolution places less emphasis on function as understood by the received view; by contrast, the focus of DST is the “morphological field.” The two general theses DST defends are (to repeat) that 1) natural selection might not be the most important mechanism in the evolutionary process, and 2) ontogenies, not populations, evolve.

**Ontogenies, not Populations, Evolve**
Expanding on the second point, ontogenies are the developmental trajectories of individual lifecycles. This does not mean that development occurs in “equipotential” ways, since DST claims that the differential changes in organism-environment relations still proceed in “modular” fashion. The unit of DST is the morphological field, which is characterized as a modular “gradient” field. In the process of embryological development, certain clusters of cells are identified as morphogenetic fields (e.g., the limb field, the eye field, etc.) that develop into the relevant “characters.” These characters, such as limbs, and their developmental stages of construction, are “modular” in two senses: the characters are individuated and serve certain causal-functional roles; and the process of construction taps into basic developmental resources that have been conserved in evolution (i.e., they are utilized through a wide range of species). Modularity also characterizes fields, which are formed by gene products and gene interactions. They are modular in the sense of being limited in space (or form): “fields can be limited by diffusion [e.g., the range of influence of morphogens], competence [ability of a cell to respond to inductors], gap junctions [proteins arranged to form channels for exchanging ions and other small molecules between cells], or cell adhesion molecules [molecules whose “stickiness” gives rise to more complex structures]” (Gilbert et al. 1996, p.366). Thus morphogenetic fields are modular due to (1) the partitioned embryological locations of the fields, (2) the individuability of characters, (3) the evolutionary conservation of developmental resources across species, and (4) spatial or formal limitations.

These four characteristics imply that changes in developmental types (by way of changes in morphological fields) are not “equipotential” changes. To further clarify, first observe that DST is conceptually framed by two coupled differential equations (section 5.2), one which tracks changes on the organism as a function of the organism’s relation to the environment, and the
other which tracks changes to the environment as a function of the environment’s relation to the organism:

\[
\frac{dO}{dt} = f(O, E) \\
\frac{dE}{dt} = g(O, E).
\]

Conceptually speaking, the function \( f \) indicates that development is structurally constrained. In particular, developmental processes occur through modular characteristics that structurally constrain the resultant developmental types. As “generators,” morphological fields are “invariant transformations” that serve as structural potentials for constructing characters; the function \( f \) expresses such modular invariance. But morphological fields also give rise to diverse characters, relative to environmental affordances; the solution set to the coupled equations conceptually represents such diversity. Thus the modularity of morphogenetic fields applies not only to the entities constructed at multiple levels, it also applies to the dynamic processes giving rise to the diverse types. As a “discrete solution set” to the coupled equations, morphogenetic fields are dynamically modular; the notion of equipotential change is not empirically defensible.

Perhaps the best example illustrating the modularity of morphogenetic fields (and point (3) above) is the homeobox, which is basically a set of master regulatory genes that guide the anatomical construction of a wide range of organisms. The homeobox has been conserved through evolution, and serves as a “universal tool kit” that, when certain parameters are tweaked (temporally, gradient-wise, via suppression, etc.), results in insect segmentation or in vertebrate segmentation, for example. To illustrate further, consider the eye (Carroll 2005), which has evolved on numerous phyla. The standard evolutionary story is that the eye is a classic case of “convergent evolution”—the eye evolved independently on separate phyla due to similar adaptive pressures. So although there are eyes for insects and eyes for vertebrates, in spite of their variegated machinery they serve the same adaptive purpose of processing relevant “visual”
information. The developmental story is rather different. Genetic studies indicate that the developmental resources for constructing eyes probably evolved only once; the eye is not merely a case of convergent evolution. The same developmental resources were evolutionarily conserved across phyla, and this universal tool kit (as an invariant developmental resource) has been deployed to produce (via changes in morphological fields) different sorts of eyes. Thus perhaps many putative examples of convergent evolution actually represent the modification and replication of morphogenetic fields. What this means for DST is that ontogenies—the developmental trajectories of morphological fields—evolve, where diverse types are effects of the modification-and-replication of these fields.

**Natural Selection Might not be the Most Important Mechanism in the Evolutionary Process**

It is crucial to recognize that this picture of evolution is quite different from the received view. The received view holds that populations—not individuals—evolve, and that “species are the effects of the evolution of populations” (Amundson 2005, p.249). By contrast, for DST ontogenies—not individuals—evolve, where “characters [the ontogenetic analogue of traits for population genetics] are effects of the evolution of ontogenies” (*ibid*). To clarify the difference, note that when DST speaks of the morphological field as the “unit” of ontogenetic and phylogenetic change, such a claim is distinct from the received view’s language concerning the “unit of selection” (or the “target of selection”), which refers to the individual. For the received view, *ontologically speaking* the individual is the target of selective pressures, but *epistemically speaking*, populational thinking must be employed to understand evolutionary changes. For DST, by contrast, the usage of “unit” pertains to ontogenetic and phylogenetic *change*; the morphological field simultaneously plays the roles of being the unit of evolutionary (phylogenetic) change as well as being the ontological stuff of ontogenetic change in an individual.
We have a shift in the importance placed on the primary mover of evolutionary change. On the one hand, natural selection is the primary force of evolutionary change for the received view. To repeat, the individual is the (ontological) "unit" of selection since selection operates on particular organisms (those that differentially reproduce versus those that do not), and seeing the emergent effects of these units is the job of populational thinking (it is in this sense that "populations evolve" and not individuals). On the other hand, modular changes in morphological fields are the primary force of evolutionary change for DST. While natural selection is important as well, it is "merely [emphasis mine] a filter for unsuccessful morphologies generated by development" (Gilbert et al. 1996, p.368). To repeat, ontogenies are the primary "units" of evolutionary change; universal toolkits like the homeobox construct particular ontogenies, and macroevolutionary change concerns the replication and modification of ontogenies. Thus the difference between the received view and DST is a difference over the primary agent of evolution. Informally, the former holds that natural selection is the primary agent: it doesn't much matter how variation is ("proximately") generated, just so as long as it is; these generative mechanisms are black-boxed, since what really matters is the cutting away of "less apt" designs. DST instead holds that the replication and modification of ontogenetic types is the primary agent. It matters how variation is produced. Firstly, the notion of "invariant" developmental resources that generate diverse morphologies is "incommensurable" with the population genetic models of the received view; the incommensurability stems from the epigenetic effects of regulatory genes (Schaffner 1998, Griffiths and Knight 1998, Amundson 2005). And secondly, "development constrains [emphasis mine] selection in its ability to produce new phenotypes" (Gilbert et al. 1996, p.362). In sum, the received view says that the weeding process accounts for most evolutionary change. DST says the production of morphological types (which impose constraints on selection) accounts for most evolutionary change. For the former, development is largely
irrelevant—how something develops only matters as it bears upon function. For the latter, development is everything—developmental accounts further assist in distinguishing homologies from analogies, by way of focusing on characters and their macroevolutionary roles.

7.5: DST and Developmental-Social Approaches to Emotions

Overview of the Argument

Given that DST differs from the received view, and given also that the received view primarily applies to the functional approaches to emotions, a revised diagram is required to examine the link between DST and the developmental-social approaches to emotions covered in Chapter 5 and Chapter 6. Concerning the former, the new organizing concept for DST is the morphological field—not function. (It is important to note, though, that the morphological field is modular, and pertains to dynamically causal-functional processes; the contrast is with the received view's understanding of "function."’) And concerning the latter, the developmental-social emotions research programs emphasize various “functional processes”—that is, “function” is understood in dynamical terms. First I amend Figure 11 given in 7.1; the new diagram concerns parallels between DST and developmental views in psychology. The biological side of the diagram (DST) encapsulates the discussion from the previous section; the psychological levels remain (roughly) the same, although temporal considerations—not reverse engineering—are highlighted. The ensuing exposition emphasizes the “Ecological Level” of psychology, since this level links DST with the developmental-social approaches to emotions. In particular, it was argued in chapters five and six that DST frames—in differing capacities—the developmental-social approaches to emotions. I build upon that argument by mapping the morphological field concept to the Cognitivist Program and aspects of the Social Constructivist Program. In the
course of the exposition I draw implications for a possible science of emotion (as I likewise did in section 7.3).

Amending the Original Parallel

The amended diagram expresses the parallel between developmental biology (in particular EvoDevo) and developmental views in psychology. The biological side now has three levels; and while the psychological side has (roughly) the same three levels, the important difference from Figure 11 is the emphasis placed on temporality. The Ecological Level is highlighted, and the parallel primarily concerns the emphasis placed on temporal/developmental factors:

<table>
<thead>
<tr>
<th>DST Explanations</th>
<th>(Developmental) Psychological Explanations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population Level</strong></td>
<td><strong>Ecological Level</strong></td>
</tr>
<tr>
<td>&quot;Characters&quot; (not traits) classified</td>
<td>(Level of Task Description)</td>
</tr>
<tr>
<td>relative to phylogenetic changes in</td>
<td>What does the &quot;character&quot;</td>
</tr>
<tr>
<td>ontogenies.</td>
<td><em>dynamically</em> do for the organism?</td>
</tr>
<tr>
<td>([Developmental Biology/86])</td>
<td></td>
</tr>
<tr>
<td><strong>Developmental Level</strong></td>
<td><strong>&quot;Computational&quot; Level</strong></td>
</tr>
<tr>
<td>The diversity of characters</td>
<td>How is information <em>dynamically</em></td>
</tr>
<tr>
<td>characterized by the <em>changes</em> in</td>
<td>processed to accomplish the task?</td>
</tr>
<tr>
<td>&quot;invariant&quot; morphological fields.</td>
<td></td>
</tr>
<tr>
<td>([Functional Biology/86])</td>
<td></td>
</tr>
<tr>
<td><strong>Anatomical Level</strong></td>
<td><strong>Implementational Level</strong></td>
</tr>
<tr>
<td>Level of &quot;Functional Biology.&quot;</td>
<td>How are &quot;computations&quot; physically</td>
</tr>
<tr>
<td>(E.g., physiological studies; functional morphology;</td>
<td>implemented?</td>
</tr>
<tr>
<td>cellular biology; etc.)</td>
<td></td>
</tr>
</tbody>
</table>

Figure 13. The Parallel Between DST and Developmental Orientations in Psychology

Some Initial Qualifications

First observe that the above "Population Level" differs from the received view's "Population Dynamic Level," since the former level is *yet to be given*. That is, DST suggests that future population models ought to track epigenetic changes on characters, in contrast with
population dynamic models that track “1-1” linkages between abstract genetic-points and
statistical-level traits (linkages expressed by dominant and recessive alleles, for example). The
future analogue of population genetics ought to be population regulatory genetics (Gilbert et al.
1996). Secondly, observe that the “Ecological Level” in psychology parallels the
“Developmental Level” and the “Population Level” in DST. As the diagram pertains to our
emotions research programs, the Cognitivist Program roughly maps to the Developmental Level,
and aspects of the Social Constructivist Program map to the Population Level (I discuss these
points below). Thirdly, it should be noted that narratives and function are not dismissed by DST,
since developmentalists still acknowledge the importance of natural selection. The revised
diagram above simply highlights the different architecture of DST.23

The Psychological Side of the Diagram

Concentrating first on the psychology side of the diagram, recall that the functional
approaches to emotions generally emphasized function in relation to the computational level. In
other words, figuring out how an emotive-cognitive system functions is the task of reverse
engineering, and reverse engineering begins with a mid-level orientation afforded by the
computational level. Now that we are switching focus to the ecological level and the
developmental approaches to emotions, reverse-engineered “function” is less important. For the
Cognitivist Program, emphasis is placed on the modes of activation within a multidimensional
activation space. While functional regularities (which are types of modes) may manifest
themselves, the real question concerns how modes temporally arise. In a similar manner, the
Social Constructivist Program emphasizes emotion syndromes as modes of operation; these
modes are situated by the vertical and horizontal relations within and between systems (the

23 Perhaps EvoDevo and the received view are “incommensurable,” as Amundson (2005) suggests.
However, as I’ve argued in chapter two, caution is required when using the term “incommensurable.”
Additionally, Chapter 8 proposes that a complex-systems approach may be a more fruitful way of
proceeding, thus avoiding the commensurable-incommensurable issue.
biological, psychological, and social systems). While each system performs certain "functions,"
the real question concerns the manner in which emotion syndromes are "dynamically
nominalistic" (see section 6.1). Thus for both programs, emotions are dynamic modes of
operation. If there is a sense in which emotions are still "functional," it would be that emotions
are "dynamically functional processes."\(^{24}\)

The Biological Side of the Diagram

Secondly, on the DST side of the diagram the main organizational concept is the
morphological field. These fields are individuable and "invariant," and also give rise to a diverse
but constrained range of characters. As individuable and invariant, the fields describe structural
potentials that are modular in form, where such modularity is primarily processual. In contrast to
functional traits—"entities" that on the received view are products of (marginalized)
developmental processes—morphological gradient fields are processual modules which
emphasize a) the differential changes to modular parts in the process of development, and b) how
changes in the rate and timing of development give rise to different types of characters.
Additionally, evolutionary narratives are important for DST, but narratives are revised in light of
the focus on morphological fields. Since natural selection is no longer viewed as the driving
force of evolution, the individual is not the (ontological) target of selection; rather the target of
"selection" (replication) is the organism's lifecycle, or its ontogeny. Also populations are no
longer the locus of evolutionary trends; rather for DST what evolves are ontogenies. More
accurately, macroevolution concerns the replication-and-modification of ontogenies through
evolutionary time. Narratives that weave together the different levels of DST stem from the same
natural historical level as the received view, but with the important amendment that historical

\(^{24}\) To use an analogy, the biological concept of homeostasis is a dynamical process, whose causal-functional
role, roughly speaking, is to actively maintain a balance within the body. Hence homeostasis is not a
functional entity; it is more accurately described as a dynamically functional process.
constraints are viewed through the lens of developmental biology (e.g., the developmental challenge to the homology-analogy distinction of the received view; the focus on macroevolution; etc.).

The Parallel: Dynamically “Functional” Processes

Lastly, I highlight the developmental level and the ecological level as they pertain to the parallel between DST and psychology. Both levels emphasize dynamics: for DST the conceptual hub is dynamic modularity (of morphological fields); and for the ecological level outlining the dynamical roles of “psychological characters” (at this “level of task description”) is paramount. As the ecological level relates to the developmental approaches to emotions, the key idea concerns the modes of operation for the Cognitivist Program and the Social Constructivist Program. The link between DST and these programs concentrates on the link between dynamic modularity and modes of operation—both emphasize dynamically functional processes. Chapters five and six argued that the general conceptual features of both programs are capturable by DST’s framework; this chapter gives further contextualization of DST so as to properly discuss the above link and its implications for each research program.

The Cognitivist Program

Beginning with the Cognitivist Program, in chapter five it was mentioned that this program could benefit from some of DST’s additional machinery—more precisely, from the notion of a set of coupled differential equations. The structural element of morphological fields stems from their modular characteristics. Accordingly, a set of coupled differential equations conceptually captures the structural invariance of fields and the dynamic unfolding of (diverse) types. The Cognitivist Program’s analogue of a morphological field is a dimensional activation space. Both concepts emphasize continuous change (“gradient fields”) and the causal mechanisms of change. And although a dimensional activation space allows for in-principle
infinite continuous variation, cognitivists claim that regions of attraction in the activation space will (probably) tend to be exhibited. These regions of attraction would be analogous to the modular characteristics of morphological fields. But what the Cognitivist Program lacks is the structural-causal tools that track how regions of attraction form. The suggestion that the program could benefit from the conceptual use of a set of coupled differential equations is just the suggestion to use these tools.

As the Cognitivist Program bears upon the question whether there can be a science of emotion, if emotions are essentially judgments—i.e., modes of activation in a dimensional activation space—then a science of emotion may be possible in the following speculative sense: assuming a total dimensional space (ranging over all dimensional activation spaces and their modes) could be given, and assuming all “sets of differential equations” governing the modes of activation in each space have appropriate “boundary and initial conditions,” then a “science” of emotion would be possible—relative to taking the Cognitivist Program “to the limit”—if a “total solution set” exists. The problems with this suggestion are formidable; let me mention just two. First, is it possible to find all “sets of differential equations” governing the behaviors of modes in any activation space? Current models (in computational neuroscience, for example), while increasing in sophistication and while intending to refer to actual physiological processes, are still artificial models of these causal processes. To what extent can “differential equations” capture the processes that current models intend to refer to (a question with greater urgency concerning all actual processes)? (Remember that the Cognitivist Program—as well as DST—employs a strong philosophical framework whose models have a prominent abstract flavor; see Chapter 5.) Secondly, even if a “wide enough” range of differential equations could be given, along with a total dimensional activation space, how do we know that there will be a global solution set? Maybe all we will have are the conceptual analogues of localized solutions and approximative
techniques. Chapter nine discusses related types of issues; for now I just mention a few of the specific problems concerning a cognitivist response to the question whether there can be a science of emotion.

The Social Constructivist Program, Dual Inheritance Theory (DIT), and Developmental Systems Theory (DST)

Moving to the Social Constructivist Program, in chapter six it was mentioned that a portion of this program is capturable by Dual Inheritance Theory (see section 6.1, "Dual Inheritance Theory (DIT): Distinguishing the Cognitivist Program from the Social Constructivist Program"), but that the program has a qualitative dimension not capturable by DIT. It was also suggested that DST generally encompasses DIT (with respect to weak social constructivism), and that DST can account for the qualitative dimension (see section 6.1, "Distinguishing Dual Inheritance Theory (DIT) from Developmental Systems Theory (DST): Creating a Space for Differing Philosophical Frameworks"). Because of these considerations, the link between biology and the Social Constructivist Program requires clarification. A potential confusion regards the relation between DIT and DST, for it appears that there is an inconsistency: first, DST and DIT are at odds, since DIT issues from the received view (i.e., DIT is a cultural extension of the received view), and DST challenges the received view. Secondly, in chapter six it was suggested that most generally speaking, DST encompasses DIT (again, with respect to weak social constructivism). Contrary to the apparent inconsistency, I claim that both positions are essentially correct. The manner in which DST encompasses DIT from a general conceptual level concerns DIT's employment of differential equations in models tracking changes on population "traits" ("dO/dt") that are situated in relation to one another ("dE/dt"). Thus philosophically speaking (i.e., from the perspective of DST's philosophical framework), differential equations

---

25 Since both a population genetic system and a cultural system track (their respective) types of "traits," a "trait" is conceptually representable as an "organism" O. The environment E is just the relation between a "trait" and other "traits."
are *framing* concepts, and in this sense DST encompasses DIT. However, the *particular* assumptions built into DIT’s neo-Darwinian models are challenged by DST, since these assumptions black-box developmental processes at the population dynamic level. Thus at the *philosophical* level DST encompasses DIT (concerning DST as a *philosophical framework*), but at the “lower levels” which matter for *practitioners* of EvoDevo (concerning DST as a *research program*), DST challenges the assumptions built into the modeling practices of DIT.

Keeping in mind this distinction between levels, we can further clarify the link between biology and the Social Constructivist Program. The portion of the Social Constructivist Program captured by DIT emphasizes the population dynamic level of the received view, as DIT applies neo-Darwinian methods to tracking (dual) genetic-cultural “traits” in populations (see section 6.1, “Dual Inheritance Theory (DIT): Distinguishing the Cognitivist Program from the Social Constructivist Program”). In contrast to the received view, DST’s population level replaces neo-Darwinian “traits” with (the broader notion of) “characters”; i.e., the emphasis is not on narrow 1-1 correlations between abstract genetic-points and statistical “traits,” but rather on mid-level regulatory mechanisms that developmentally cause a *range* of ontogenetic “characters.”

Currently, since for DST there is no actual analogue of the (received view’s) population dynamic level—only the *recommendation* for a future population regulatory genetics—there is no further application of “population developmental biology” to the cultural realm. Thus while the received view has been extended into the cultural realm via DIT, DST is still too young a field; it has yet to first provide an explicit population level theory.

**The Social Constructivist Program and DIT: The “Non-Qualitative” Dimension**

Because the Social Constructivist Program has links to both the received view and to DST, the question whether there can be a science of emotion has two separate potential answers. The first (speculative) answer regards the portion of the Social Constructivist Program
encompassed by DIT. Since DIT is as an extension of the received view, and applies broadly to the (dual) transmission of genetic-cultural “traits” at the population dynamic level, emotions would be a particular class of “traits” within this general level. So if DIT proves to be a viable “science” of the cultural realm, it would then be possible to have a science of emotion, even if the details of the science remain empirically “scattered.” For just as the received view exhibits empirically “scattered” results that nevertheless are consilient, thus making possible a science of life—an analogue of the general category of emotion—likewise, if DIT proves to be a viable science of culture, as a corollary a science of emotion would be possible. Recall from Chapter 6, though, that DIT’s philosophical framework assumes that the relation between representation and represented is non-problematic (i.e., the framework is “dualistic”). This “non-qualitative” dimension of DIT frames the above speculative possibility of a “science” of emotion, which raises the further problem: just what conception of “science” would we be operating with concerning the possible range of objects it purports to study?

The Social Constructivist Program and DST: The “Qualitative” Dimension

The second (speculative) answer focuses on DST as a philosophical framework and as a research program. It should be reemphasized that DST is still a relatively new research program with a highly speculative framework (e.g. the extent to which regulatory genetic networks are responsible for most evolutionary change has been contested; see Coyne 2005). If it turns out that a limited set of morphological fields and their epigenetic modifications account for most evolutionary change, and if a similar mapping could be given between the Population Level and the Social Constructivist Program, then finding the relevant (population level) “morphological emotion fields” might establish the possibility of a science of emotion. Since these fields would be “generators” of emotions, the general category of emotion would be bounded by a limited set of generators.
This argument is akin to the argument made above regarding the Cognitivist Program and individual dimensional activation spaces, although here the Social Constructivist Program shifts perspective to communicative (population level) activation spaces. The same sorts of problems raised for the Cognitivist Program also apply. Additionally, the argument is on even shakier ground since a) population developmental biology does not yet exist, b) if such a field of study were to be established, it isn’t clear that an extension similar to DIT could be given, and c) even if such an extension were to be provided, it is unknown what mapping would exist between this extension and the Social Constructivist Program.

But what is most problematic is the qualitative dimension of the Social Constructivist Program covered in Chapter 6. Recall that the dynamical nominalism of syndromes discloses the dialectical, “qualitative” dimension of representations. In other words, syndromes reveal that it is not legitimate to assume a separation between representation and represented. The dialectical aspect of DST’s philosophical framework accommodating this qualitative dimension offers a quite different conception of “science.” So even if “morphological emotion fields” could be given that were “generators” of various emotion-characters, we still would have the problem: what happens when the “dialectics” of these representations are taken into account? Would it even make sense to talk about a “science” of “emotion,” when (assuming that general category even exists) its putative boundaries undergo indefinite change in response to the very study of “emotions”?

7.6: Concluding Remarks

In conclusion, each of the four emotions research programs has salient links to the (respective) diagrams from 7.1 and 7.5. And each program individually encounters problems concerning whether there can be a science of emotion. Is there a way to “synthesize” all four
programs? Perhaps doing so would provide clues to establishing if a science of emotion is possible. However, the synthesis cannot consist merely of picking the best of each program and then “mixing” these elements, since the synthesis must exhibit coherence between programs that is not based solely on “ad hoc mixing.” The problem lies in providing a synthetic framework for conceptualizing the relevant relations between the programs. The framework ought to be tractable enough to handle (heterogeneous) variation, and yet at the same time provide coherent and “elegant” organization for such variation. Chapter 8 addresses these issues, and offers a complexity framework for potentially framing the heterogeneous nature of the category of emotion.
Chapter 8: Beyond Reduction and Anti-Reduction: A Complex-Systems Approach

Recapitulation

Chapter 7 discussed the first major assumption running through chapters two through six—namely the broad “evolutionary” perspective adopted in analyzing the four major emotions research programs. Two parallels were discussed: one between the received view of evolution and cognitive psychology, and the other between developmental biology and developmental views in psychology. The two parallels stem from similar methodological orientations to understanding biological and psychological systems. Specifically, the first parallel concerns “function” understood via reverse engineering and via (methodological) adaptationism; the second parallel concerns dynamically “functional processes.” These two senses of “function” help to organize the heterogeneous subject matters that biology and psychology study.

Prospectus

The tension in biology between the “functionalists” (the received view) and the “structuralists” (DST) has a broad philosophical analogue, expressed through a particular debate within philosophy of mind between anti-reductionists and reductionists. The analogue concerns the ideas that a) functionalists emphasize the natural historical level when providing accounts of biological function, where ultimate explanations are “anti-reductionistic,” and b) structuralists emphasize a morphological field’s interactions, whose epigenetic regulatory mechanisms appear more “reductionistic.” However, it would be inappropriate to flatly claim that functionalists are anti-reductionists and that structuralists are reductionists. For although Mayr claims that the historical dimension of function is irreducible to lower levels of explanation, biological function also involves proximate causations that on occasion allow for reduction (e.g., explanations at the
level of biochemistry); and regarding DST, the emphases on epigenetic processes and developmental modularity require appeals to higher-order, causal-functional levels that are comparatively "anti-reductionistic." It appears as if the philosophical analogue is too broad to be of any use in explicating the particular tension between the functionalists and the structuralists. In spite of this, the analogue will be used to situate the need for a complex-systems approach, an approach that will have bearing upon the tension. Additionally, the particular philosophical debate between reductionists and anti-reductionists which I shall focus on stems from considerations about "heterogeneity"—the second major assumption mentioned at the beginning of Chapter 7.

The second major assumption built into chapters three through six is the assumption that emotions research programs are "heterogeneous" sciences. Recall from Chapter 2 that there are two senses of "heterogeneity": the diversity of subject matters of the special sciences (heterogeneity₁), and the diversity of physical implementers (heterogeneity₂) that arguably implement a higher-order property of a special science. Concerning the former sense, the two major types of approaches—the functional approaches and the developmental-social approaches—exhibit varying degrees of heterogeneity₁. That is, the functional approaches search for FLOPs whose degree of heterogeneity₁ corresponds to the degree to which the systems are centralized; developmental-social approaches tend to exhibit a high degree of heterogeneity₁ because of the nature of dynamic considerations. What about heterogeneity₂? In Chapter 1 it was discussed whether cognitive science is really a science—an issue whose problems apply to the question whether there can be a science of emotion. The reduction/anti-reduction debate over the status of special sciences like cognitive science concerns whether such "sciences" are deserving of the name. A fair number of special-science researchers not only think that they are doing science; they in part think they are doing science because of the relative autonomy of their
disciplines.¹ A philosophical justification for this view appeals to intuitions about heterogeneity, which I will cover in 8.1.

Section 8.1 gives an overview of a particular reduction/anti-reduction debate as it relates to the heterogeneity appeal. In section 8.2 I then sketch an example that intersects with physics, optics, and perception. It describes how both wave and ray theories of light are needed to physically account for the appearance of a rainbow—something that exists only in the “eye of the beholder.” The model exhibits what Robert Batterman calls “asymptotic reasoning,” which explores the borderland between theories, and where such reasoning shows (as an “existence proof”) that it is inappropriate to speak of reduction or anti-reduction in describing borderland accounts. Section 8.3 discusses how this example may help to inform debates over the relative autonomy of the special sciences. Section 8.4 subsequently suggests that an indirect approach, focusing on systems, could profitably shift the debate; the disclosure is that the focus on systems is actually about complexity and the conceptual categories that frame complex-systems thinking. Lastly, section 8.5 summarizes this chapter and chapter seven; in particular, I discuss the link between evolution and complexity. The link is important since it suggests that a complex-systems approach to the four emotions research programs has significant bearing on whether a science of emotion is possible.

8.1 Reduction and Heterogeneity

The two authors I utilize to represent a particular reduction/anti-reduction debate are Jaegwon Kim (1992, 1999) and Jerry Fodor (1974, 1997). (This debate centers on issues in philosophy of mind, as “psychology” is taken to be the paradigmatic example of a special science.) Fodor’s basic argument for heterogeneity (from which claims to autonomy issue) stems

¹ Autonomy is relative in that functionalists are still token physicalists. I discuss this point in section 8.1.
from his denial of a reductive view: to reduce predicates in a higher-order theory to the predicates in a "reduction base" (or basal theory), the kinds in the former are mapped via appropriate laws to (natural) kinds in the latter. In general, the reduction assumes that physicalism (of some sort) is correct, and that the reducing base contains physical laws. The schema of a physical law is of the form $P_1 x \rightarrow P_2 y$ where a (natural) kind is determined by the predicates of such (proper) laws whose variables are bound by the relevant quantificational scope(s). A "law" of a higher-order science is given likewise by $S_1 x \rightarrow S_2 y$ (S is a predicate of a special science) with its respective (putative) kinds. The bridge laws are of the form $P_1 x \leftrightarrow S_1 x$ and $P_2 x \leftrightarrow S_2 x$, from which it follows that higher-order predicates are reduced to predicates in the reduction base, and so nomologically coextensive kinds are expressed in the physical base. A picture of this reductive view is presented below:

**Figure 14. A Reductive View**

2 Fodor (1974). $P_1$ is a physical predicate, and the schema is read as "all events which consist of x's being $P_1$ bring about events which consist of y's being $P_2$." The different variables may range over different domains of quantification, depending on how one conceives of events. Also the connective "\(\rightarrow\)" captures, at a minimum, transitivity.

2 The variable is the same for two reasons. On a strong reductionist reading, it is presumed that the higher-level ontology is identical to the lower-level (physical) ontology. On a weaker reading, bridge laws don't express identity statements; rather "x's satisfaction of a P predicate and x's satisfaction of a S predicate are causally correlated" (Fodor 1974).
Fodor's Anti-Reductionism

What Fodor denies is that the (special) higher-order sciences, where there apparently are kinds, are reducible in such a manner. A paradigm of the non-reducibility of the special sciences is illustrated by the predominant computationalist/functionalist attitude in cognitive science.¹

Take, for example, a Turing machine, whose processes are realizable in any number of mechanisms, real or imagined. Suppose that there are physical laws that can capture some of the features of Turing-computable processes. The question raised is how, if such computation is multiply realizable (think of all the ways a calculator can be realized), there can be bridge laws that properly qualify the predicates in the reduction base(s) as a kind. For even if each (higher-order) realization is instantiated by a lower-order kind $K_p$, Fodor rejects the claim that the disjunction $K_1 \lor \ldots \lor K_n \lor \ldots$ (think: using an abacus, or counting with fingers, or using a slide rule, or...) of such kinds is itself a proper kind. How long would such a disjunction be? (Answer: potentially infinite.) What systematic connections would hold among disjuncts other than satisfying the bare criteria for being Turing computable? (Answer: it is difficult to discern any systematic pattern that would make a set of disjuncts a plausible candidate for being a proper kind.) Thus the general intuition that reduction must fail due to the heterogeneity of the physical realizers/kinds (in the reduction base(s)) that instantiate a higher-order property.

To further cash out this general intuition, let us expand on the second answer. Fodor claims that for functionalists, there are higher-order functional laws. At the same time, functionalists are also token physicalists—that is, they believe that each mental event, say, is causally and nomically connected to some physical event, and that a mental token occurs only under appropriate physical conditions. What they deny is type physicalism (which is stronger

¹ Note that this view differs from biology's understanding of function, since Fodor's notion of function exhibits only the minimal feature of multiple-realizability.
than token physicalism)—i.e., they deny that a mental state can be defined as the disjunction of its possible realizers (the collection of which would form a type); for while a bona fide functional predicate is projectible, its disjunction of “physically realizing” predicates is not, and so there are higher-order kinds that aren’t reducible to their physical realizers (Fodor 1997). Type physicalism is really about the status of bridge laws. Type physicalism would allow for reduction since bridge laws get construed as genuine laws (at the physical level) linking physical types with mental types, and whose predicates are projectible since they would delimit proper kinds (i.e., the variables of the predicates are bound by the appropriate quantificational scope(s)). In this case, apparently, there are genuine/projectible disjunctive laws. But functionalists resist the notion of projectible disjunctive laws, as they resist the “gerrymandering” (p.156) of predicates that are used to cover higher-order functions.

There are two claims. First, if the disjunction is presumed to be exhaustive, as in the claim that “jade is green” is really the statement that “jade is (chemically) nephrite or jadeite,” then such closed disjunctions are not laws at all, but are properly reduced to laws about jadeite and laws about nephrite. Second, if disjunctions are open, they also aren’t properly construed as laws since such openness suggests “missed generalizations” (p.158) at the physical level. So in the first case, predicates are not gerrymandered; rather the truism “jade is green” is mistaken for a law when in fact its disjunctive formulation reveals that it ought to be construed as two separate (non-disjunctive) laws. In the second case, open disjunctions do not avoid the charge of having predicates that are gerrymandered since they fail to satisfy the “intuitions” of the projectibility requirement.⁵

---

⁵ Intuitions on this requirement differ. For Kim (1992) it amounts to not sanctioning generalizations where any law could be made when one of the disjuncts turns out to be satisfied. The intuition for Kim is that any confirmation procedure which allowed for such disjunctive laws would be “cheap and illegitimate” (p.12). Fodor’s focus isn’t on confirmation (1997, p.152); rather he targets “independently certified” (i.e., non-gerrymandered) bridge laws that are putatively projectible laws (p.156)—laws that aren’t used merely to
Kim’s Reductionism

While Fodor takes an anti-reductionist position, Kim offers a revised account of reduction. The two principles Kim uses are the “Principle of Causal Individuation” and the “Causal Inheritance Principle.” The former states that “kinds in science are individuated on the basis of causal powers; that is, objects and events fall under a kind, or share in a property, insofar as they have similar causal powers” (1992, p.17). The latter holds that “if a mental property M is realized in a system at t in virtue of physical realization base P, the causal powers of this instance of M are identical with the causal powers of P” (p.18). What Kim proposes is a local reduction of higher-order properties. The picture is the same as in Fodor’s presentation of reduction, with the important exception that higher-order “laws” are distinguished relative to what Kim calls “species” (p.19); the general idea is that given the Principle of Causal Individuation—which Fodor also subscribes to—talk of “species” makes sense in light of different causal powers. But Kim parts with Fodor regarding how species are employed: first, a higher-order property “E must be functionalized”—that is, E must be construed, or reconstrued, as a property defined by its causal/nomic relations to…properties in the reduction base B” (1999, p.10). The second step is to “find the realizers of E in B. If the reduction, or reductive explanation, of a particular instance of E in a given system is wanted, find the particular realizing property P in virtue of which E is instantiated on this occasion in this system; similarly, for classes of systems belonging to the same species [emphases mine] or structure types” (p.11). Since E is multiply realizable, E is restricted to a set S (for “species”) of higher-order types. The last step is to then “find a theory (at the level of B) that explains how realizers of E perform the causal task that is constitutive of E”
(ibid), where the theory employs bridge laws of the form $S \rightarrow (E \leftrightarrow P)$, for species $S$, higher-order property $E$, and reductive property $P$. The point of Kim’s local reduction is to preserve the homogeneity of “similar” causal powers, and thus (he thinks) the homogeneity of functional properties. What functionalists deny, of course, is the Causal Inheritance Principle when it is wedded to type physicalism.\(^6\) So even if token physicalism—which functionalists subscribe to—in general is compatible with the Causal Inheritance Principle, what is left open is the scope of the bridge laws (genuine versus mere) that are supposed to capture the (causal) identity conditions for higher-order and lower-order properties.

Kim claims that if a functionalist accepts token physicalism and the Causal Inheritance Principle, then higher-order “kinds” cannot satisfy the Causal Individuation Principle since each kind would have to be multiply realizable and so would not be a causal kind, as each causal kind’s scope of quantification needs to be distinct from other causal kinds’ scopes (1992, p.18)—i.e., each kind is homogenous, and given the apparently wild heterogeneous nature of what can realize a functional property (again, think of all the ways in which a calculator can be built), such a property can’t share “similar causal powers.” Since Kim holds that proper scientific kinds are causal kinds, functionalism as exhibited in psychology (or biology), for example, cannot be a part of science proper; “psychology as a science with disciplinary unity turns out to be an impossible project” (p.18). Functionalists are left with apparently two unsatisfactory choices: either the special sciences which require higher-order properties are not genuine sciences, or reduction (via species) is possible, in which case the special sciences are not autonomous.

8.2: Asymptotic Reasoning

\(^6\) Unlike Fodor’s earlier position (1974), it is now no longer clear to what extent Fodor is a functionalist (see Fodor 1997, p.155, where he seems to be leaning toward a more “reductive” stance).
The murkiness of kinds is interrelated with the similarly murky notions of law and theory (Fodor 1974). In a sense, the debate over reduction and heterogeneity amounts to differing intuitions about what an acceptable policy is for admitting kinds into scientific discourse. Kim seems to hold that the mystery of how functionalism can work without homogeneity (of cause) is too great to warrant the inclusion of the special sciences in science proper (Fodor 1997). On the other side, “functionalists” within numerous disciplines (biology, psychology, etc.) continually expound on “kinds” that have enough traction to justify each discipline’s claim to relative autonomy. The terms used in the debate between Fodor and Kim—“properties,” “laws,” “causal powers,” etc.—do not explicitly focus on concerns about theories. But if Fodor is correct in his assessment that law, kind, and theory (at a minimum) are bound together in what seems like an inescapable circle, perhaps an examination of actual scientific theories would be helpful in moving beyond the reduction/anti-reduction debate—a proposal I will pursue in what follows.

**Universality**

Robert Batterman, in *The Devil in the Details*, distinguishes between two types of why-question: one that asks for an “explanation of why a given instance of a pattern obtained,” and another that asks “why, in general, patterns of a given type can be expected to obtain” (p.70). The former question can be answered in any number of ways, and to an extent satisfies some explanatory needs. For example, if I am asked why a particular formal derivation has the form that it does, I might respond by appealing to specifics of the conclusion and premises (which may include retracing the steps of the derivation and the rules justifying each step). But if the question is answered according to the second type, I might appeal to a meta-theorem of the system which claims that derivations of a certain pattern arise from the “interaction” of particular system rules. What is important about the second type of why-question is that there are patterns exhibiting what Batterman calls “universality”: “1. The details of the system (those details that would feature in a
complete causal-mechanical explanation of the system's behavior) are largely irrelevant for describing the behavior of interest. 2. Many different systems with completely different 'micro' details will exhibit the identical behavior” (p.73). A particular example of a pattern of universality involves the wave-ray account of a rainbow.7

The Rainbow Example

Suppose we have a spherical raindrop where a ray of light is first refracted at the point of entry, then reflected off of the inner “back wall” of the sphere, and then again refracted when coming out of the raindrop. With a set of parallel incoming rays to this sphere (given the appropriate scattering angle), some of the outputted rays form the primary bow of the rainbow, whose curve is known as a “fold caustic” (imagine, for those incoming parallel rays, a set of outputted lines that form “tangents” to define a curved line). What a geometrical (ray) theory of optics gives us is the ability to predict the shape of the bow(s). However, the theory cannot account for the “presence of so-called supernumerary bows. These are fainter arcs appearing on the lit side of the main rainbow arc. They are visible in the sky at times as faint alternating pink and green arcs [and are different from a second rainbow]. Furthermore, since actual rainbows do not exhibit light of infinite intensity, [this] singularity [when on the fold caustic/bow] predicted by the ray theory cannot be realistic” (p.82-3). To cover the former gap, wave theory employs the “Airy integral” (along with additional tools from the wave theory of light) which yields the interference patterns that account for supernumerary bows. So it would seem that ray theory and wave theory are all that is needed. But the question remains, how do ray theory and the Airy integral interrelate? For clearly in providing a proper account of a rainbow, it would be theoretically troublesome to simply patch together theories, appealing to one theory where another fails in order to cover the phenomena.

7 Batterman (2002), Ch.6.
The misleading metaphor that the Airy integral gives "the wave flesh that decorates the ray theoretic bones" (p.84) makes it seem as if in the limit the two theories form a smooth theoretical patchwork. Accordingly, Batterman raises the question, just how much can the wave theory really capture? The answer turns out to be surprising: at the very place in which one would hope that the wave theory transitions smoothly into what the ray theory yields, it actually falls apart. A qualitative sketch of the idea goes as follows. Take an arbitrary "wavefront" (as a simple illustration, a concentric light-sphere emanating from a candle). The coordinates on the wavefront can then be used to define rays as lines perpendicular to the surface (the light-rays emanating from the candle are perpendicular to the spherical wavefront). Next, take a patch around a ray and project it to get a "ray tube" (also the patch is used to define an energy flux—the rate at which the patch changes). Such a tube can be used to approximate the wave intensity, and the hope is that the tube and the wave intensity will (everywhere) smoothly interrelate to one another. As this pertains to the wave-ray account of rainbows, unfortunately what happens is that the energy flux on the light-surface/wavefront and the energy flux at the projected point (on the caustic/bow) don't smoothly interrelate, for when the patch at the focal point shrinks to zero, the function blows up. The "interfering ray sum [which is intended to capture the wave-ray patchwork] cannot describe the nature of the wave near and on a caustic. In other words, it breaks down exactly where we need it most. ...Caustics and focal points are the primary objects of study in optics, but the interfering ray sum fails exactly at those places" (p.88).

A further weakness of the wave-ray patchwork is that one of the functions in the "interfering ray sum" depends on the shape of the raindrop. If raindrops deviate from the spherical shape, different solutions are generated that also deviate from actual rainbow phenomena. But "we observe, as a matter of empirical fact, that despite the fact that raindrops have different shapes, the patterns in intensity and in fringe spacings that decorate the caustics are
similar. *These patterns are universal*" (p.90). The structures that capture such universal patterns are the caustics, which (technically) are "catastrophes." A better representation is given in order to understand the limiting, asymptotic borderland between the ray and wave theories—namely "catastrophe optics." The key new idea is that a catastrophe defines an "equivalence class" over a range of perturbations (different raindrops), sort of like the way a calculator defines a heterogeneous class of functional "equivalents."

**Asymptotic Reasoning**

Asymptotic reasoning occurs in the following way. First, the interfering ray sum is put into a "generalized wave function" form, and then another generalized equation is given for the "fold caustic catastrophe" (the catastrophe type of relevance to the shape of rainbows). Finally, a generalized (asymptotic) equation defines a "scaling law" that discloses a (self-similar) pattern which is universal across various perturbations. This scaling law relates the two other generalized equations, one wavelength dependent (and so can give the intensities and fringe spacings needed to account for supernumery bows) and the other wavelength independent (since it describes the stable shape of the rainbow). Taking the wavelength to zero, the fringes and intensities don't exist at the limit and the caustic takes over (we get a "raw, unadorned bow"); when the wavelength is too large the "noise" drowns the caustic pattern (there is no bow—just "colors"); and if the wavelength is perturbed within certain bounds, both the caustic and the interference patterns yield the bowed, colorful rainbow phenomena.

What we have is a mediating law (the scaling law) that does not belong to either the ray or wave theories. The first implication is that the law is not appropriately classified as a bridge law, since it doesn't really have the form of a bridge law, and additionally the objects that a bridge law would quantify over are supposed to be "the same" (whether identical, or causally correlated; see fn.3)—but the singularity of the caustic is different from the wave approximations.
The second and most crucial implication is that the scaling law depends on a higher-level orientation in order to reveal and delimit the equivalence class of microstates captured by the generalized caustic equation. In particular, rainbow phenomena are oriented by systems of raindrops and the (scaling) interrelations they have with perception, optics, and wave physics. These systems are only revealed by the higher-level orientation/perspective adopted—a perspective that affords the qualitative distinction between macrolevel and microlevel.

Another related implication of the rainbow example is that it provides an illustration of a high-fidelity disposition, further emphasizing the importance of perspective. (In chapter two it was mentioned that it is possible to have a "high-fidelity disposition"—an apparent oxymoron.) First, suppose that rainbow phenomena are expressive of a capacity (concerning a raindrop-sunlight system). While such an assumption is not unproblematic, it seems to accord with the characterization of a capacity as a "deep structure" with a "changing mix of different causes, coming and going; a stable pattern of association can emerge only when the mix is pinned down over some period or in some place" (Cartwright 1989, p.182). The deep structure would be given by the generalized scaling law; the changing mix of different causes would be the wave and ray interactions over heterogeneous raindrops; the stable patterns that emerge are the rainbow phenomena (associated with certain rainy conditions); and such phenomena are "pinned down" relative to what tolerances are conducive to the generalized wave equation. (Note that this is compatible with Cartwright’s claim (1999, pp.59-64) that capacities can do many sorts of things, for the generalized scaling law also can accommodate the manifestation of non-rainbow-like phenomena when the wavelength varies.) While the model is simplified (but not simplistic), it is important to remember that to an extent the details don’t matter regarding the emergence of universal patterns. Light is still represented as having a wave-ray "duality"; raindrops can come in a variety of shapes and sizes; and rainbow phenomena are saved. To all appearances, then, it
looks like this is a case of a capacity with high fidelity. Hence dispositions are not necessarily “vague”—they may be abstract, but they still can exhibit high fidelity. In other words, to see emergent, high-fidelity rainbow phenomena—the relevant objects of interest—a higher-level perspective must be adopted. The upshot is that such a perspective is not merely “perspectival”; rather, the systems-orientation adopted has genuine “ontological” (as well as epistemic) import. For dispositions form the “ontological ground” disclosing rainbow phenomena—they are required presuppositions for revealing system dynamics in the first place (see Chapter 2, “Capacities and Tendencies”).

8.3: Autonomy of the Special Sciences?

What is interesting about asymptotic reasoning is that the borderland demands, at least from the theory standpoint, a reexamination of the heterogeneity/reduction debate. Batterman notes that there is a sense in which the emergent caustics are “contained in” the wave equation (p.96). But the relation between these two is not deductive since at the limit we get infinite intensities, and the relation is also not one of philosophical reduction, where the coarser caustic theory gets absorbed by the finer wave theory. Rather if there is an operative sense of “reduction” it is the (inverted) sense that physicists use when talking about a finer-grained theory reducing to a coarser theory by letting certain key parameters go to the appropriate limit(s). In our example, when the wavelength goes to zero some of the ray theoretic structures appear near the limit (of course not at the limit, since the function blows up). In this hedged sense it is appropriate to say that the emergent caustics are “contained in” the wave equation. But certainly it is not appropriate to say that the scaling law is predictable from the (“fundamental”) generalized wave equation, as the stability of the patterns over different raindrops requires structures not present in that equation. Borderland reasoning doesn’t have the same
preoccupation with kinds, "laws," causal powers, etc. Accordingly, the unsatisfactory choices left to the functionalists by Kim are open to dispute if one doesn’t agree to the terms and conditions of the debate. Even stronger, Batterman’s focus on actual scientific theories at least shows the possibility that there is a different way to conceive of emergence without invoking the reductive/anti-reductive language of Kim and Fodor.

But even if we acknowledge such a possibility, does asymptotic reasoning really relate to what matters to functionalists (among others)? For what matters to functionalists is the status of higher-level phenomena that do not lend themselves so readily to mathematical analysis. Batterman acknowledges as much when he observes that “folk psychology by its very nature will likely never be expressible in a sufficiently formal language for this [mathematical] analysis to apply. Nevertheless, to me, it doesn’t seem unreasonable to expect that the emergence of certain mental phenomena, if they are genuinely emergent, will depend on some sort of limit—for instance, the limit of increasing complexity” (p.128-9). Suppose for the moment that Batterman has the right intuitions. It suggests that the question of the autonomy of the special sciences is the wrong one to ask, since if there are “limits” of increasing complexity, or more appropriately, thresholds of complexity, then one would expect to find universal patterns by examining the borderland surrounding each threshold, where such patterns might be “autonomous” from the patterns exhibited in other sciences, but also might reveal similar patterns (and perhaps identities) across any number of sciences.

Beyond Kim’s Reduction and Fodor’s Anti-Reduction

---

8 See his Ch.7 on the borderland between classical and quantum mechanics, and Ch.8 (p.121-6) on how statistical mechanics does not “reduce” to classical mechanics; rather statistical mechanics is more appropriately viewed as a borderland account. See also Auyang (1998) and Stevens (2003) on statistical mechanics and population genetics.

9 For example, Stuart Kauffman’s Boolean (NK) network can be transformed into a spin-glass model, which has integral relations (via Ising models) to statistical mechanics and to quantum field theory. It can also be transformed into neural net models, and into adaptive landscape models. See Kauffman (1993, 1995).
The problem with the autonomy question as it relates to multiple realizability is that it is too narrow, and may be too wide. It is too "narrow" in the sense that the heterogeneity intuition appeals to a rather artificial view of laws and kinds that emphasizes logical form and forgets about physical theories and the importance of structures. And it might be too wide in that claims to autonomy may not be justified. Suppose I am a local reductionist (following Kim) and I hold that certain causal powers are homogeneous and can be used to define second-order properties. Suppose further that I am a functionalist (following Fodor) who holds there are other causal properties that are not homogenous. The intuition in this case would be that the laws which describe non-homogeneous causal properties are projectible and that the related functional laws are "tokened" in each instance (as would be given by the relevant reduction) but the functional phenomena are neither reduced to the non-homogeneous laws nor are such phenomena autonomous. How so? Substitute "laws which describe non-homogeneous causal properties" with "scaling laws" to get the desired outcome.

Consider the rainbow example again. First, given that the concept of "causal powers" (cf. Kim's "Causal Inheritance Principle" and the "Principle of Causal Individuation") is opaque, it is not unreasonable to suppose, for the sake of the argument, that the scaling law expresses certain "causal powers." To capture the intuition that causal powers are homogeneous, observe that the scaling law can express "homogeneous" wave powers and homogeneous ray powers. These powers instantiate second-order properties having to do with the fringes and spacings for the former, and the caustics for the latter. Second, suppose that the scaling law also expresses non-homogeneous causal properties, which is plausible since the catastrophe defines an equivalence class over "heterogeneous" systems of raindrops.10 Third, given the respectable

---

10 It could be objected that systems of raindrops are not wild enough to count as heterogeneous. But if a potentially infinite array of systems of water droplets (with a correspondingly potentially infinite array of different shapes) isn't heterogeneous at least with respect to shape, it seems to me that the burden is placed
status of optics and wave physics, each has some claim to projectibility, and certainly the scaling law has at least the same claim to projectibility. Fourth, the multiply realizable phenomena—e.g., the stable manifestation of rainbows—have to do in part with the appearances of rainbows, and such appearances are not reduced to the scaling law. And finally, even if I am a functionalist who buys into higher-order functional laws of perception (that are, let's suppose, projectible), such laws are not properly characterized as autonomous since, as mentioned in the previous section, rainbow phenomena are oriented by systems of raindrops and the interrelations they have with perception, optics, and wave physics. The point is that claims to relative autonomy (remember that functionalists tend to subscribe to token physicalism\(^1\)) might end up being better understood in terms of Batterman's notion of universality—a more contextualized account of "multiple realizability."

8.4: Systems and Complexity

As mentioned above, the problem with asymptotic reasoning is that it doesn't clearly extend to those sciences where mathematical analysis is not so easily applied. The suggestion to look at the thresholds of complexity provides a way of preserving the spirit of asymptotic reasoning when perhaps only partial mathematical representation is possible. The questions I further address in Chapter 9 include: how are thresholds to be conceived? and how do we understand emergence? The first question taken up here is whether scaling makes sense in the special sciences. The answer depends on the degree to which mathematical models can be

---

\(^1\) To claim that the token in this case has to do with brain states, say, misses the point. Even if "tokens" are taken from brain states, from "states" of the rainbow, and so on, the relevant interrelations will not be disclosed, as many of the interrelations have to do centrally with patterns. And while it could be argued that token physicalism and functionalism are compatible with such patterns, the issue is not about mere compatibility, for compatibility does not help in revealing what patterns are appropriate.
provided that actually invoke the use of scaling laws involving catastrophes, fractals, etc. This is an "empirical" matter that I do not further discuss. Rather what I seek to do is extract some of the general conceptual features of asymptotic/borderland reasoning and apply them to understanding emotions.

"Emotions" as "Affects": Towards Borderland Accounts in Understanding Centralized Processes

At this point a confession is in order. As alluded to in the Preface, my focus on emotions is really about the general problem of understanding some of the dimensions of affective processes—specifically, emotive-cognitive systems. More broadly speaking, the range of cases that qualify as affective is wide indeed, including diverse phenomena like tending to a minimal energy state, the dynamics of fear, and the dispositional nature of evolution. I use "emotion" as a term of art for the simple reason that, in spite of its numerous senses and uses, it is also the term predominantly employed by (emotions) research programs. The "borderland accounts" I have in mind build upon the valuable resources that these differing programs provide. More fully, the three informational stances situate the emotions research programs (of chapters three through six) within the four-dimensional space; borderland accounts may examine the borders between the 4-D subspace "shapes" (as outlined by each of the informational stances) of each program. These programs, for the most part, examine emotive-cognitive systems. In other words, the new sciences of the mind are investigating increasingly centralized processes, and precisely because nobody knows just what a complete "science of mind" would look like—if it can be provided at all—these sciences probably would benefit from adopting borderland perspectives. Since a significant amount of fairly new language has been put in place in order to discuss the "borderland spirit" of understanding "emotions," I hope the reader—whose patience I no doubt

---

12 As one example of a putative scaling law in biology, see "Life on the Scales" at http://www.sciencenews.org (Feb. 12, 2005; Vol. 167, No. 7).
13 Chapter 9 discusses the recent terminological shift to "affect," as indicated by the new interdisciplinary field of "affective sciences" (Davidson et al. 2003).
have stretched — will further bear in mind that this work is, in effect, one long argument seeking to configure a number of scientific senses of “emotions,” and to provide a novel framework. Part of what needs to be included in this new language is talk of “levels,” since with appropriate caveats the use of “levels” will prove to be far more fruitful than using talk of scaling (especially when there is no obvious recourse to any mathematical representation).  

**Emotions as Systems**

To make sense of a borderland account, a systems approach is required. In general I conceive of emotions as systems — they are individuals whose constituents are situated in relation to each other and to that system as a whole. In particular, emotions are (1) modes of operation, where each mode, with the proper perspective adopted, is (2) just a system. Regarding (1), to revisit each of the research programs, Ekman’s program holds that affect programs are modes of operation that accord with the features of Fodorian modules (see Chapter 3); evolutionary psychology holds that FIIPS are modes of operation within the cognitive system (FIIPS are superordinate programs; see Chapter 4); the Cognitivist Program holds that appraisals are modes of activation within dimensional activation spaces (see Chapter 5); and lastly, the Social Constructivist Program holds that emotions are dynamical syndromes situated by the vertical and horizontal relations within and between systems (i.e., syndromes are dynamical social modes oriented by the biological, psychological, and social systems; see Chapter 6). Regarding (2), all of these modes are types of systems: Fodorian modules, and FIIPS more generally, are boxologies whose constituents are situated in relation to each other and to the “boundary” of each boxology.

---

14 Batterman rejects a particular sense of “levels” in understanding emergence as it relates to the rainbow example — a sense stemming from a discussion of emergence as explicated by Kim (1999). My use of “levels” is different, as it is extracted from a complex-systems approach. For example, even if two theories operate at the same “level” of investigation, from a complex-systems perspective we can still talk about differing scales for a system at this particular “level”; that is, from a complex-systems perspective, these scales within a system may be viewed as “levels” (additionally, from this perspective we can also talk about different systems as “levels”). See fn.15 and fn.17 below.
(recall that sections 3.2 and 4.2 discuss how affect programs and FIIPS are systems). The
dynamical modes of the Cognitivist Program and Social Constructivist Program are dynamical
structures whose parameters are situated in relation to each other and to the behavior manifested
by the mode (recall that sections 5.1 and 6.1 discuss how appraisals and syndromes are systems).

A Sketch of Complex-Systems Thinking

Since systems are defined as individuals whose constituents are situated in relation to
each other and to that system as a whole, from a broad scientific perspective, emotions are
systems. The rubric of “systems” provides the means for constructing frameworks to organize a
wide range of topics, where, importantly, the frameworks are construed as hypothetical
constructs. Emotions are viewed, more precisely, as complex-systems. The two requisite
categories for this meta-level approach are: 1) a (“static”) state space representation, and 2) a
dynamical representation (Auyang 1998). Each category unfolds the theme of “synthetic
microanalysis,” which in brief is the bundling of micro-relations required to form constituents
that, when situated in a system, have macro-relations to each other and the system (I expand upon
these claims in Chapter 9). The emergent dynamics between fine-grained and coarse-grained
levels bring into relief the various contours of the two categories, and also what it means to speak
of a system as an individual. To illustrate using the rainbow example, the “static” representation
would be given by the total solution space for the scaling law. The “dynamical” representation is
given by the interrelations between the wavelength dependent and independent solutions. The

\footnote{For example, in the generalized scaling law, the “system” (at this “level”) may be conceptualized as the sense of the equation (with relevant parameters/“constituents” of interest) as a “whole.” It can be conceptualized as an “individual” in that its state space is specifiable. And its “constituents” are things like state variables, control parameters, etc. The system and its constituents are interrelated to one another in providing particular solutions to the equation (say, if one component is fixed and the other constituents are allowed to vary). See also fn.17 below.}

\footnote{Strictly speaking, there is no time variable in the scaling law. For strictly speaking, Sunay Auyang’s (1998) contrast between static and dynamic representations concerns the distinction between, on the one hand, equilibrium models where the governing global constraints are set with respect to some variable, and on the other hand, modeling the evolution of the system over time, which may refer to micro/macroscale processes.}
“micro-relations” at the fine-grained level of the wave equation are just the ways in which waves destructively and constructively interfere. The “macro-relations” hold between the members of the equivalence class defined by the generalized fold caustic equation. And the way in which the wave “constituents” (packets of constructive and destructive interference) have macro-relations when situated in catastrophe equivalence classes is revealed emergently through the generalized scaling law—the asymptotic borderland that brings forth the sense of the “system” as a whole.17

**Emotions and Complexity**

The real question for my project is how to deploy these conceptual tools to construct a framework for understanding emotions. The key idea employed to this end is the use of informational stances. The functional approaches to emotions are captured by the functional stance; they represent the “static” state space of possible boxologized emotion-systems (remember that dynamic considerations are downplayed, not omitted). The developmental-social approaches to emotions are captured by the strategic and semiotic stances; they represent the dynamics of how emotion-systems “evolve” within and between multiple levels. The informational stances situate the thresholds between levels of investigation, and these levels in turn are framed by a complex-systems approach (as I argue in Chapter 9). Although I will make crucial use of complex-systems thinking, this project on emotions as complex-systems is not primarily about complexity; rather the project is an extended exploration of “evolutionary” considerations. Specifically, utilizing the general parallels between evolution and “psychology”—

---

17 Note that when we “move up a level,” the systems of raindrops form the macrolevel individuals that then delimit the “micro”-states of catastrophe equivalence classes. It should also be noted that while Batterman uses the rainbow example to explore inter-theory relations, I’ve appropriated the example to cohere with the complex-systems framework. That is, the rainbow example can be construed as examining inter-theory relations operating at the same “level” of investigation—i.e., the overall scaling law-raindrop-observer system (which also can be conceptualized as containing further “sub”-systems).
and the tangled relations to the human sciences—this meta-level project is attempting to provide conceptual tools for thinking about "emergence," and the implications that follow with respect to understanding what it means to speak of a "science" of emotion (as I discuss in Chapter 9).

As one last remark, recall that in Chapter 1 the argument was presented for why there can be no science of emotion—an argument which most notably drew from Paul E. Griffiths' superb work *What Emotions Really Are*. I have used his work through the course of my argument as a "foil," as a source of interpretation, and as a source of further exploration. Griffiths' orientation is very much along the lines of the heterogeneity/reduction debate, although he does not make explicit reference to these philosophical issues. As I read him, he has taken these philosophical themes, and has transfigured and then transferred them to the debates in philosophy of biology over the status of categories and systematics. And while it isn't old wine in new bottles, many of the intuitions about heterogeneity and projectibility guide the argument that there cannot be a science "proper" of the emotions (and thus no science of emotion). In chapters two through six I have attempted to realign such a claim from the standpoint of heterogeneity. I hope to further realign this claim by appealing to intuitions different from those grounded in heterogeneity. In particular, I appeal to the different intuitions coming from the new "science" of complexity, and, of course, how it is already making inroads to understanding evolution. Darwin had it right: evolution really is one long argument. Perhaps no less should be expected regarding the heterogeneity of emotions.

8.5: Evolution and Complexity: An Overview of Chapter 7 and Chapter 8

The Two Major Assumptions

---

18 Indeed, in many ways my work wouldn’t be possible were it not for what I hold to be one of the finest philosophical works on emotions.
19 E.g., Depew and Weber (1997).
To take stock, the first major assumption (Chapter 7) is that an evolutionary perspective situates the major emotions research programs. This assumption stems from the general parallel between biology and psychology; in particular, the link between the received view and the functional approaches to emotions, and the link between DST and the developmental-social approaches to emotions. The second major assumption, closely related to the first, is the examination of sciences from the standpoint of "heterogeneity." This assumption stems from 1) the diverse subject matters of biology—which parallel the manifest diversity of emotions investigated by the four research programs—and 2) the diverse physical implementers that support types of "function" (in its bare philosophical sense—see fn.4 above—and more importantly in its biologically robust senses).

"Function" and the Two Major Assumptions

In fact, the common theme running through both major assumptions is the conceptual placeholder of "function." With respect to the first major assumption, there are two uses of function pertaining to the received view and DST. Regarding the link between the functional approaches to emotions and the received view, function emphasizes types of boxologies, primarily at the computational level for Ekman's program, and primarily at the border between the computational level and the natural historical level for evolutionary psychology (see sections 7.1 through 7.3). These types of "function" (recall section 3.2 and section 4.2) have a more static flavor in comparison with the dynamic senses of "function" employed by the developmental approaches to emotions. Regarding the link between the developmental approaches to emotions and DST, "functional processes" emphasizes dynamic modes of activation, primarily at the border between the ecological-computational levels and the developmental level for the

20 Although the Social Constructivist Program has links to both DST and the received view.
Cognitivist Program, and primarily at the border between the ecological level and the population level for the Social Constructivist Program (see sections 7.4 and 7.5).

With respect to the second major assumption, "function" grants the special sciences—specifically psychology—(semi) autonomy from lower-level sciences. Heterogeneity concerns functional types that exhibit projectibility—types that are stable and yet multiply realizable. Heterogeneity also concerns the diversity of subject matters in sciences like biology. For the received view and DST, the robust senses of "function" at multiple interrelated levels of investigation indicate that diversity and consilience are organized around "functional explanation." In other words, diversity and consilience are what make biology a different kind of science, since their historical dimension grants biology its (semi) autonomy.

"Function" and Complexity

Thus the thread running through both major assumptions is "function"—in its multiple rich senses. The final issue to discuss is the relationship between evolution and complexity. The complexity framework extracted from the rainbow example illustrates that there is a contextualized sense of "function" (multiple realizability) pertaining to the universality of rainbow phenomena, whose emergence is in part expressed via the scaling law. One virtue of the complexity framework is that it moves beyond the reduction/anti-reduction debate, since the philosophical categories of homogeneity (Kim) and heterogeneity (Fodor) are too general to be usefully applied in various contextualized scientific accounts (e.g., the rainbow example; statistical mechanics; population genetics; the relation between classical dynamics and quantum mechanics; etc. (see fn.8 and fn.9)). In a similar vein, the issue is whether the complexity

---

21 Recall that DST de-emphasizes "static" senses of "function." Since developmental biology is conceptualized as "8(Functional Biology)/8t," and since DST's primary notion is the morphological field (which exhibits consilience through invariance, and diversity through developmental types), "function" for DST emphasizes "dynamically causal-functional processes."
framework can be linked to biologically informed senses of “function”—not merely the bare philosophical sense of the term (see fn.4).

Ron Amundson, in discussing the apparent incommensurability between the received view and DST, offers a suggestive analogy to illustrate a possible relation between the two. He writes:

I am still puzzling over the relationship between mainstream neo-Darwinism and EvoDevo. No EvoDevo-ite denies the importance of natural selection, but very few EvoDevo concepts pay the slightest attention to populations. Recognition of the phenomenalist origins of population genetics offers a possible solution to this conundrum. EvoDevo is a mechanistic theory about the internal dynamics of a complex system. Population genetics is a phenomenal theory about the statistical behavior of assemblages (populations) of such systems. EvoDevo stands to population genetics as the kinetic theory of gases stands to the phenomenal gas laws [emphases mine] (2005, p.11).

I propose to take this analogy seriously, for I think it provides a plausible link between the received view and DST—a link made possible by a complex-systems approach. The reason for taking this analogy seriously is simple: statistical mechanics is a paradigm for complex-systems thinking. While I have used the rainbow example to illustrate the complexity framework, the framework actually comes from Sunny Auyang (1998), who uses statistical mechanics to extract the conceptual tools for thinking about complexity. If population genetics is akin to the phenomenal gas laws (expressed by high-level laws of association), and if DST is akin to the kinetic theory of gases (expressed by differential equations stemming from classical mechanics), the natural question would be: what is the analogue of statistical mechanics? For statistical mechanics brings together the macrolevel phenomenal gas laws and the microlevel kinetic theory of gases.

The standard philosophical story on statistical mechanics is that it is a classic case of intertheoretic reduction (cf. section 8.1); the phenomenal gas laws are really just expressions of the kinetic theory of gases. For example, at the macrolevel, the ideal gas law expresses a phenomenal law, which maps correlations between parameters like temperature, pressure, and so
on. The "basal theory" is the kinetic theory, which is a part of classical dynamics. The "bridge laws" then link high-level parameters (temperature, etc.)—put in statistical form—to low-level parameters (point particles with velocity, etc.). But this story is wrong; the guts of the actual mathematics are hidden in the "logical form" of bridge laws, rendering the philosophic lesson of intertheoretic reduction inappropriate. When the mathematical structures are taken into account, statistical mechanics supports a complex-systems approach (see Auyang 1998, Batterman 2002). As with the previous discussion of Batterman (in particular, see section 8.3, "Beyond Kim's Reduction and Fodor's Anti-Reduction"), the point about statistical mechanics is that statistical aggregation involves a loss of information, so the system's macro-behavior is not linked to any precise details about microlevel state descriptions; rather each macrostate could have many micro-realizations—each macrostate is "multiply realizable."

So if we take seriously the analogy made by Amundson, the natural suggestion is to apply a complexity framework to the received view and DST. After all, the central lesson that biology exhibits consilience—the unexpected convergence of evidence from heterogeneous investigations—only reinforces that this science of exceptions should strive for the "resynthesis" of the Modern Synthesis with EvoDevo. However, it is currently unknown what form the resynthesis may take. As observed above, practitioners of DST do not generally concern themselves with statistical-population thinking; rather their emphasis is on the causal mechanics of epigenetic processes. In spite of this, the picture offered by Gilbert et al. (1996), as mentioned in Chapter 7, points to a future populational regulatory genetics. If DST is to fully explore organism-environment relations, the study of higher levels of interaction—within and among groups, and ecological relations more generally—requires that a populational stance eventually be
adopted.\textsuperscript{22} The \textit{conceptual} tools for thinking about "scaling relations" between microlevels and macrolevels are what a complex-systems approach offers. Thus taking Amundson's analogy one step further, why not seek a resynthesis by adopting a complex-systems perspective? The framework already applies to some of the most successful theories in physics, as well as to population genetics. And while \textit{particular models} from the population dynamic level may be "incommensurable" with models from DST, the point is that adopting the general \textit{conceptual} complexity framework would not exhibit incommensurability. For just as the generalized caustic equation is not helpfully viewed as incommensurable with the generalized wave equation, so likewise claiming that the received view and DST are incommensurable may be the wrong meta-scientific idea to apply. Pretheoretically, the complexity framework appears just the framework to adopt. In addition, given the links between evolution and the emotions research programs, it is natural to adopt the framework as it bears upon a possible science of emotion.

\textsuperscript{22} Concerning "emotions," section 7.5 discussed some possible relations between "population thinking" and the Social Constructivist Program.
Chapter 9: Toward a Science of Emotion

Recapitulation

Let me recap the connection between complex-systems thinking and the emotions research programs. First, I proposed in the previous chapter that a complex-systems perspective might organize the tension between the received view and DST. The suggestions by Amundson that the received view is analogous to the phenomenal gas laws, and DST is analogous to the kinetic theory of gases, are based upon, respectively, the association between the ("phenomenal") population dynamic level of the received view and the phenomenal gas laws, and the association between the ("mechanistic") developmental level of DST and the kinetic theory of gases. On the basis of Amundson's analogy, the proposal was made that just as statistical mechanics expresses a complex-systems perspective unifying the phenomenal gas laws with the kinetic theory of gases, so likewise perhaps a complex-systems approach can organize the tension between the received view and DST. Secondly, given the two links from Chapter 7—the link between the received view and the functional approaches to emotions, and the link between DST and the developmental-social approaches to emotions—the natural inference would be to adopt the complexity framework for the functional and developmental-social approaches to emotions.

However, the connection is not as straightforward as it may appear. One part of Amundson's analogy rests on the association between the received view's highest level—the "phenomenal" population dynamic level—and the phenomenal gas laws. By contrast, the link I've made between the received view and the functional approaches to emotions emphasizes the middle and lower levels of biology and psychology; in particular, the natural historical level for biology, and the computational and implementational levels for psychology. The other part of Amundson's analogy rests on the association between DST's "lower" developmental (and
anatomical) level(s) and the mechanistic kinetic theory of gases. (In terms of our four-dimensional space, the emphasis is on the causal-concrete level.) By contrast, the link I've made between DST and the developmental approaches to emotions emphasizes the middle and higher levels of biology and psychology; in particular, the developmental and population levels for biology, and the ecological level for psychology. Thus my links between biology and the emotion research programs apparently “invert” Amundson’s analogy.

However, this claim is only partially correct. For what requires clarification is the relation between the biological levels of Figure 11 from 7.1 and the biological levels of Figure 13 from 7.5. While the anatomical levels are roughly the same for both the received view and DST, the higher levels require more careful consideration. Firstly, the natural historical level is the middle level orienting the higher population levels (the general ecological level and the population dynamic level) and the lower anatomical level. And secondly, the developmental level is the middle level orienting the population level and the anatomical level. So from this broad perspective, there is a “similar” mapping between the levels of the received view and DST. But upon closer inspection, a number of differences arise. These differences, in conjunction with the “similar” middle-level orientation of both views, complicate Amundson’s emphases on the received view’s population dynamic level and DST’s (causal-concrete) developmental level.

As argued in chapter seven, the most important level for the received view is the natural historical level, since it organizes the search for function at and between other levels of investigation. While Amundson is correct in claiming that the phenomenal aspect of the population dynamical level is important for the received view — epistemically speaking — Mayr’s primary emphases are on narratives and the ontological objects of evolutionary study, namely individual organisms. And regarding DST, as a research program EvoDevo emphasizes the causal-concrete level to a greater extent than the received view. Amundson rightfully points out
that EvoDevo focuses on the more mechanistic aspects of evolutionary change. However, DST
also has a strong philosophical framework emphasizing the (differential) mutualism of
organism-environment interrelations. Thus the "inversion" I mentioned above is only partially
correct, since 1) epistemically speaking the population dynamic level is very important for the
received view (the level that Amundson highlights), and 2) as a research program, DST
emphasizes the causal-concrete level to a greater extent than the received view. But we must not
forget that Mayr highlights the natural historical level and the focus on individual organisms. It is
from this mid-level perspective that I've linked the received view to the middle and lower levels
of psychology, as the link pertains to Ekman's program and evolutionary psychology. For both
of these programs emphasize the mid-level aspects of EIIPS. We must also not forget that DST
has a strong causal-concrete orientation and a highly philosophical framework. It is from this
"crescent-like" perspective that I've linked DST to the ecological level of psychology, as the link
pertains to the Cognitivist Program and the Social Constructivist Program. For both of these
programs emphasize temporal aspects of organism-environment interrelations.

Prospectus

Given these qualifications, pursuing a complex-systems approach is still a promising
suggestion for the following reasons: 1) Amundson's analogy concerns the microlevel causal
mechanics of DST and the macrolevel phenomenal laws pertaining to the population dynamic
level; 2) statistical mechanics—a paradigm for complex-systems thinking—encompasses the
"causal" microlevel kinetic theory of gases and the macrolevel phenomenal gas laws; 3) the
distinction between microlevel and macrolevel can also be applied to the emotions research
programs, but the crucial caveat is that the four programs are not organized around the concern
with microlevel and macrolevel; rather they are organized via a conceptual "state space"
representation (capturing the functional approaches to emotions, which black-box developmental processes) and a conceptual "dynamical" representation (capturing the developmental approaches to emotions); 4) recall from the rainbow example that the distinction between microlevel and macrolevel is a distinction of degree, since one can shift perspective, whereby what was formerly a macrolevel then becomes a microlevel; and 5) likewise, what is crucial in complex-systems thinking is not a "reified" distinction between a fixed microlevel and a fixed macrolevel, but rather the qualitative moments of a state space representation and a dynamical representation—both of which may appeal to microlevels and macrolevels.

These five reasons underlie why I speak of understanding emotions (from a conceptual standpoint concerning scientific research programs) as modes. More accurately, there are two distinct but closely related claims: first, modes are systems (see section 8.4), and emotion-modes are complex-systems; and second, what frames a conceptual understanding of these modes is a complexity framework. In section 9.1 I expand upon these two closely related claims. In 9.2 I build upon 8.4 by further cashing out the tools for thinking about complexity. It turns out that complexity is closely related to the notion of emergence, which I discuss in 9.3. Sections 9.2 and 9.3 then lead to consideration of "levels" and "thresholds" as they relate to the three informational stances; in other words, section 9.4 emphasizes the qualitative nature of the conceptual complexity framework when applied to the emotions research programs. Lastly, 9.5 discusses the question whether there can be a science of emotion.

9.1: Emotions as Complex-Systems

My coverage of emotions-modes "as" complex-systems is speculative, since while researchers increasingly realize that neuro-physiological processes appear to be nonlinear self-organizing processes, the relevant studies are only beginning to get underway (e.g., Tononi et al.)
1998, Lewis and Granic 2000, Lewis 2005). Despite this shortcoming, looking ahead to section 9.4, I will expand on emotions “as” complex-systems from the conceptual standpoint of complex-systems thinking, in light of the tools to be provided in 9.2 and 9.3. In this section I begin by sketching three features of emotions relative to the functional and developmental-social research programs. The first feature is that emotions are conceived as systems, or in particular as modes. The second feature is that emotion-modes do not make hard distinctions between microlevels and macrolevels; these distinctions are employed as working levels of investigation. Thirdly, the claim that emotions are complex-systems trades on an ambiguity, namely that emotions refer to nonlinear self-organizing processes, and that the conceptual tools for analyzing complexity frame an understanding of emotions. I take up each of these points in turn.

**Emotions as Systems**

To reiterate from section 8.4, systems are characterized as individuals whose constituents are situated in relation to each other and to the system as a whole. For the functional approaches, emotions are conceptualized as FIIPS. FIIPS are systems in the sense that they are boxologies whose constituents are situated in relation to each other and to each boxology “as a whole” (i.e., the function that the box serves). In particular, for Ekman’s program, affect programs are FIIPS whose modal characteristics are cashed-out by the features of Fodorian modularity, and for evolutionary psychology, emotions are FIIPS whose modal characteristics are cashed-out by the features of superordinate programs. As for the developmental-social approaches to emotions, emotions are systems since their parameters are situated in relation to each other and to the behaviors manifested by the dynamical modes. In particular, for the Cognitivist Program, appraisals are modes of activation in a dimensional activation space, and for the Social Constructivist Program, syndromes are dynamic social modes oriented by the biological, psychological, and social systems.
Levels of Investigation

The second point is that microlevels and macrolevels are working levels of investigation. As it concerns the functional approaches to emotions, microlevels multiply realize functional macrolevels. In other words, the computational level taps into the implementational and anatomical levels of investigation, both of which multiply realize the computational state. Using affect programs a representative example, a few possibilities include: investigating Turing-like platforms which multiply realize “affect programs” at the computational level (Picard 1997); investigating “affect programs” ethologically, where multiple neuro-physiological processes realize similar functional states across species; or investigating affect programs in our own species, where these states are functionally similar in spite of (individually) unique neuro-physiological “fingerprints.” Furthermore, if we increase the “resolution” of the processes at the implementational or anatomical levels, what was viewed as a microlevel—the previously mentioned neuro-physiological processes—could become a macrolevel. For example, populations of neural networks could now be the macrolevel, and the microlevel may be neural networks in particular “modular” regions of the brain. These considerations concerning the relativity of “grain” also apply to the developmental approaches to emotions. Since emotions are viewed as dynamical modes of activation, the only shift in emphasis is on how microlevel changes relate to macrolevel changes; grain now highlights these dynamical factors. Thus microlevels and macrolevels are useful tools for understanding kinds of granularity, and are not reified levels of analysis.

Complex-Systems as Representations

Lastly, the third point is the most important, since it distinguishes between the conceptual scheme (or “representation”) for understanding complex dynamics, and what the scheme is intended to refer to. The distinction between a representation and what it is a representation of is
a distinction that should be carefully maintained when dealing with complex-systems, precisely because of the “manifest complexity” of the studied systems. For example, “actual” neural populations, “actual” ecological systems, and “actual” thermodynamic systems apparently involve multiple interrelations manifesting chaotic behaviors and functional stabilities. To all appearances, they are complex, and the categories that situate complex-systems thinking seem to reference (in some capacity) systems “in the world.” What matters for our purposes is that 1) the language of systems is crucial since it allows for thinking about different kinds of grain, i.e., the shifting borders between microlevels and macrolevels, and 2) the language of complexity frames the shifting dynamics between levels and between systems.

**Conceptualizing FIPPS as Complex-Systems**

With these caveats in mind, for the functional approaches to emotions, each FIPPS can be viewed as a complex-system in the following qualitative sense: 1) it makes sense to speak of macrolevels and microlevels; 2) the computational level discloses functional macrolevels; 3) the implementational and anatomical levels disclose relevant neuro-physiological microlevels (e.g., via chaotic neural net simulations); 4) the state space can be represented, for example, by what the parameters configuring such neural nets afford; and 5) the dynamical representation is given not only by the algorithms governing the chaotic changes in the net, but also by the targeted macrolevel “basins of attraction”—the functions at the computational level. In this way, developmental processes are black-boxed by functional approaches, since it is assumed that “developmental potentials” and “norms of reaction” will yield convergence to the macrolevel basins of attraction. To restate the received view’s position, development largely doesn’t matter with respect to function, since microlevel chaos is “cancelled out” by the tendencies expressed by functional basins. As for the latter, “norms of reaction” presuppose that function is robust—you can “feed” the function “nutritious food” or “junk food,” but just as long as the types of food fall
within "acceptable tolerances" as dictated by the macrolevel parameters, the system will mature into a proper FIIPS that is panversal. Think of it this way: a biological organism takes surrounding order and disorder, and transforms such energy into biological function, which is a particular type of order conducive to survival (see Chapter 7). FIIPS like affect programs are complex in the sense that they are robust. For in spite of different kinds of "foods" and various chaotic processes threatening the organism's growth and survival, what emerges is the very stuff of life—the emergence of order in the form of function that (locally) cuts against the drive of entropy.

**Conceptualizing Dynamical Modes as Complex-Systems**

By contrast, the developmental-social approaches to emotions focus not on FIIPS as basins of attraction, but rather on the dynamics that give rise to FIIPS and other developmental outcomes. Dynamical modes of activation are the focus of these approaches. Metaphorically, if FIIPS concentrate on adaptive peaks within an adaptive landscape, dynamical modes concentrate on the shifting relations that generate possible alternative landscapes; for even if there are developmental regularities like FIIPS, there may be other possible landscapes contextualizing why there is convergence to one landscape as opposed to another. Dynamical modes are also complex-systems in the following sense: 1) there are macrolevel and microlevel processes at work; 2) the ecological level discloses macrolevel tasks that developmental processes serve; 3) lower causal levels disclose relevant microlevel "morphological fields" (i.e., dynamical activation spaces with appropriate "differential equations"); 4) the state space is the total "solution space"; and 5) the dynamical representation tracks the unfolding of possible "landscapes." For developmental approaches, the very notion of a "norm of reaction" is challenged, since not only does it fail to take into account how changes at the microlevel may have cascading effects through development (which would lead to different developmental outcomes), but more
importantly, it presumes that FIIPS are basins of attraction which render negligible how a
constrained set of developmental resources can give rise to a diverse range of characters.¹ To put
the contrast informally, emergence for the functional approaches involves convergence to a single
basin of attraction, whatever that FIIPS may be, whereas emergence for the developmental
approaches involves a range of possible basins of attraction, whether they be new forms of
emotion-judgments (for a cognitivist), or new emotion-syndromes (for a social constructivist).

These different kinds of basins of attraction are, in general, different kinds of “wholes.”
Complex-systems thinking is useful precisely because it offers a means of organizing part-whole
relations across a wide range of sciences. Additionally, wholes are emergent patterns exhibiting
what Batterman calls “universality.” It turns out that wholes and emergence are deeply
embedded in considerations about information, although cashing out their interconnections will
take some work. In the next section I outline the general tools for complex-systems thinking, and
in section 9.3 I relate these tools to emergence. Section 9.4 subsequently applies the concepts
from 9.2 and 9.3 to situate emergence from an informational perspective; specifically, framing the
emergence of wholes requires the language of “levels,” “thresholds,” and “information.”

9.2: Tools for Complex-Systems Thinking

A good example involving emergence has already been given in chapter eight, namely
the rainbow example. I have also given a general overview of complex-systems thinking as it

¹ Employing Batterman’s language, for the functional approaches, “universality” concerns the panversal
nature of FIIPS; the details of development are largely irrelevant in understanding such universal
phenomena. For the developmental approaches, “universality” concerns a constrained set of
“morphological fields” (i.e., dimensional activation spaces with appropriate “differential equations”);
lower-level interactions are largely irrelevant in understanding such “epigenetic” phenomena.
relates to this example in section 8.4. The task here is to give a fuller picture of what conceptual tools frame an understanding of complexity and systems.

A Sketch of the Conceptual Complexity Framework

A major consideration not explicitly mentioned in 8.4 is how relations give rise to complexity in “many-body systems.” As the topic of relations is a philosophical “minefield,” it would be prudent to give an overview of relations that gives us what we need, but no more than that. Accordingly, Auyang (1998, p.49) distinguishes between monadic relations called “characters” (the main constituents of the state of an individual), and relations that are non-monadic (other explicit relational statements about a system). Both characters and (non-monadic) relations can be further distinguished into “types” and “values” (tokens). The importance of the type-value distinction concerns the opening of possibilities that occurs when we shift to type descriptions. A “state description” includes character types for each individual, plus relevant relation-types between individuals, which schematically capture a system’s salient, (usually) causal features. States are further distinguished into “microstates” and “macrostates,” where microstates are the product of lower-order constituents (as primarily determined by the relevant “microcharacters” and relations), and macrostates consist of higher-order “macrocharacter” types (with respect to the behaviors of interest). In order to bring such behaviors to the fore, three additional tools are required: a “distribution” of characters (which tend to be closer to the macrostate than the microstate (p.62)) across a population of constituents of a system, and “exogenous” and “endogenous parameters.” Normally, the exogenous parameters are control parameters that define the boundary conditions for the system in question, and the endogenous

---

2 For further examples see Batterman (2002), Auyang (1998), and Strevens (2003, 2005). The general complexity framework that I employ is extracted primarily from Auyang.

3 Types (monadic or relational) can be cashed out in terms of functions, which assign values from a range of (possible) characters with respect to an individual. Functions which utilize character and relation types are often used to delineate rulelike (and lawlike) behaviors (p.49).
parameters are the state variables of interest. Paradigmatically, the above concepts are expressed by a set of well-defined differential equations modeling a many-body physical system. For our purposes, what matters is that the conceptual framework extracted from using these tools is one where the distribution of characters forms a macroindividual (system) as a composite of microindividuals which are situated—by virtue of their exogenous and endogenous parameters—in relation to one another and to the system as a whole.

Individuating Complex-Systems on the Basis of Their (Dispositional) Behaviors

Systems are individuals on the requirement that they can be individuated (mathematically, the necessary condition is that their state space is specifiable). It is presumed that the external and internal couplings defined by the exogenous and endogenous parameters are weak enough to delineate behaviors of interest within that space. If it turns out that these couplings are too strong, thus making the delineation of behaviors problematic, several tricks can be used. Either define a new transformation that disentangles the couplings, or reshift the boundary conditions and use new relations to absorb these coupled “thickets.” The crucial point is that disclosing a system-as-an-individual is intertwined with the system’s manifestations (usually via a “phase portrait”); complex-systems are dispositional. This dispositional aspect organizes considerations about “parts and wholes.” For if many-body aggregations become “too noisy” then there is no intuitive sense that any particular aggregate has enough integrity (mathematically speaking) to be considered a legitimate candidate for being an individual. The

---

4 For example, see Auyang’s discussion of statistical mechanics (1998).
5 An example of the former involves the local use of a hyperplane to bring into relief certain qualitative dynamics that cannot be easily discerned by the governing equations. For the latter, an example would be using general fluid dynamical equations to model flow “above” the level of particle dynamics. In the rainbow example, the absorption of relations would be the way in which the generalized caustic equation only draws from certain catastrophes.
6 Traditional mereological concerns usually focus, in general, on the logic of partially ordered sets (e.g., Simons (1987) and Varzi (2004)). The problem, though, is that the focus on logic fails to reveal the structures that the language of state spaces, phase spaces, etc. affords.
case where any aggregate suffices for individuation ends up being a mere artifact of an observer’s choice, which cuts against the desire to discern regularities that make a consistent inter-subjective difference. (For example, if a system manifests universal behavior, like in the rainbow example, such stable behavior would make a consistent inter-subjective difference for investigators.)

Two things need clarification. The first is that the specifiability of a state space is normally associated with a system where dynamic changes in the system are not emphasized; rather attention is drawn to all the possible states that the system can attain, and what counts as a state of the system under consideration. A phase space, by contrast, is concerned with how regularities within the state space arise when dynamical considerations are centrally taken into account. Secondly, although “state space” and “phase space” can be used interchangeably, since ultimately a (“total”) phase space is the same as a (“total”) state space, the difference is one of perspective—the total solution space for a set of differential equations, for example, can be given conceptually by all possible solutions (state space), or by a complete family of phase portraits. (Another way of saying this is that the shift from state space to phase space is a shift from analysis to topology, where one is concerned to discern the topology of various regions that correspond to families of solutions to equations governing the system.) The latter is a qualitative space, whereas the former is usually quantitative. This difference in perspective is important with respect to seeing how systems are dispositional. For in the case where a system is seen from the standpoint of a state space (when a quantitative solution space is given), it might be difficult to discern how certain regularities manifest themselves; but a qualitative portrait gives a picture of what, say, attractors (assuming there are any) lie where. Phase spaces make perspicuous the dispositional nature of those systems, for which they are appropriate.7

7 *A fortiori*, the general situation with respect to differential equations is that most do not lend themselves to neat analytical solution, so it would seem that qualitative representations are a “practical” necessity. There are two deeper points: the first is that Hamiltonian systems are separated into integrable and non-integrable
Phase Spaces and Dispositions

If our systems-based theories and models are dispositional in nature, two philosophical questions arise: what if a phase space isn’t specifiable, or there is no phase space? and what classifies systems as complex? With respect to the first question, there are two cases. For the first case, if there is a phase space but it isn’t specifiable, this could mean several things. It could mean that a phase portrait doesn’t reveal any discernable patterns (pure noise?), and so while strictly speaking there is a phase space, there aren’t any behaviors of interest. Or in a stronger sense it could mean that there truly is a phase space “in the world,” as it were, but for some reason or other we cannot find the specifiable pattern—e.g., it might be embedded in some higher-dimensional mathematics that our minds (for some reason or other) don’t have access to.

Either way, the problem remains that we cannot discern any dispositions. I set this case on the side, since it is not relevant to the project of understanding systems that are complex.

In the other case, if there is no phase space then certainly it isn’t specifiable, so it would seem to follow that there are no dispositions. But this presupposes that phase space representations are the only tool required for understanding dispositions, which in turn assumes that some mathematical representation can be given. So assuming that a (consistent) mathematical representation is given of, say, a range of phenomena, the claim would be that there simply is no phase space for that representation. It is not clear what sense this makes, since apparently there are problems which cannot be “solved,” and yet solutions “exist.” A not uncontroversial example might appeal to Gödel’s first incompleteness theorem, where a mathematical system satisfying the conditions of the theorem would have true statements (the “solutions that exist”) which cannot be proved (cannot be “solved”), at least relative to that system. A less exotic example would be the n-body problem, which was once widely presumed to be unsolvable (analytically). Suppose this were actually the case. Still, relative to phase spaces, there are partial qualitative “solutions”; additionally, in some sense a solution exists as a total Platonic phase space.
another way to read the claim that there is no phase space. Auyang, while focusing on tractable cases involving many-body systems, also mentions two other kinds of systems, namely "organic systems" and "cybernetic systems." She means by the former a system with many different kinds of components that are "highly specialized and tightly integrated. Organic systems are conducive to functional descriptions" (p.10). The latter is a blend of many-body systems and organic systems, where the types of relations between constituents are far greater than in many-body systems. Now perhaps with cybernetic systems, which seem to be the most complex of systems (the kinds of systems with the most interesting mix of functional order and chaotic disorder), there might be no phase space due to a high degree of complexity. It follows that there would be systems which have no (total) phase space since they cannot be given a complete mathematical representation. The issue that remains is whether the concept of disposition is also dismissed. My brief answer is "no," since the general category of disposition is wider than its particular exemplification in phase portraits. A fuller justification of this claim is given in the remainder of this section and in section 9.3.

Information and Classifying Systems as Complex

The second question above, what classifies systems as complex, has already been partially answered in this section and in section 8.4 where the basic framework was laid out. There is a further dimension to complexity that Auyang mentions which helps to classify a system as complex, namely information. Although she notes that there is no precise comprehensive definition of "complexity"—there is only the intuitive sense that applies to "self-organized systems that have many components and many characteristic aspects, exhibit many structures in various scales, undergo many processes in various rates, and have the capabilities to

---

9 And why not? The general situation seems to be that there are thus far no complete mathematical representations of whatever such representations are taken to (metaphysically) represent. Or more weakly, no uncontestable cases of purported completeness.
change abruptly and adapt to external environments” (p.13)—two precise notions of complexity coming from the informational/computational sciences are given, but their import is restricted by their very precision. The two notions are “information-content complexity,” measured by the smallest program capable of specifying a well-defined binary sequence to a computer, and “computation-time complexity,” measured by the time it takes to solve a problem with that problem’s most efficient algorithm (pp.13-4). The latter measure of complexity distinguishes between tractable problems (polynomial-time algorithms) and intractable problems (exponential-time algorithms). Many problems fall into this latter class, the importance of which is that a “brute” approach is infeasible, and so to tackle such problems researchers “direct their attention to reformulating the problem, relax[ing] some conditions, or find[ing] alternative tractable problems” (p.14). Accordingly, the relevance of informational complexity is that while problems and solutions in natural science are less rigorously definable than in computation science, natural scientists have adopted similar strategies [emphasis mine]. If the lowest-energy configurations of an \( n \)-atom system involve specifying the state of each atom, then the problem is intractable. With \( n \) a trillion or more, physicists do not even try a brute-force approach. They wisely spend their effort in formulating more tractable and interesting problems, for which they develop many-body theories. With insight and versatile strategies, problems regarding systems with the highest information-content complexity [think: total information given by the state space] need not be the most difficult to solve. Statistical methods deal with random systems [—systems which have a high degree of information since random interactions between constituents result in a greater number of permutations—] more adequately than with systems with some regularities\(^{11}\) (p.14).

A paradigm of an intractable problem in physics is the \( n \)-body problem. There are (at least) two senses of “intractable” involving, on the one hand, the presumed lack of a solution to the equations that set-up the \( n \)-body problem, and on the other hand, the computational difficulty

\(^{10}\)Note the resonance with Cartwright’s informal characterization of a capacity, as discussed in section 8.2.

\(^{11}\)For example, Bernoulli trials (with appropriate chance set-ups) are easier to deal with than systems that have, say, causal regularities, since in the latter case much more work is required to screen off “false” (or non-causal) correlations. See Strevens (2003, 2005) for a probabilistic interpretation of the philosophical foundations of complexity, where he argues that, as with the above random systems, “chaos” actually enables and supports complexity (informally, the idea is that chaos “erases” the details of the system, allowing higher-order “functionalities” to show themselves).
involved in finding solutions that precisely track systems with large $n$ (mainly because these systems exhibit sensitive dependencies on initial conditions). The (presumed) lack of a solution stems from the claim that the regions of chaotic dynamics implicitly contained in the n-body equations are non-integrable. This is the sense in which an n-atom system is “intractable” in the passage above. Closely related to this point is the second claim that a brute-force approach is not feasible for large $n$ due to the concomitant computational (run-time) intractability, from which the emphasis on strategies gains importance.

However, these two senses are separable from a theoretical standpoint for the reason that, in fact, the n-body problem has been solved.\(^{12}\) The catch is that the proof’s solution converges at a rate so slow as to be practically unusable with respect to finding actual solutions, which are the objects of interest for physicists. So while the n-body problem is not intractable in the sense of being unsolvable (intractability\(_1\)), effectively the problem is still intractable when trying to find actual precise solutions to the problem (intractability\(_2\))—various sophisticated correction techniques are used which redefine the problem into approximative tractable form. The reason for drawing attention to these two senses of “intractability” as they relate to information and complexity is that Auyang’s discussion of emergence relies on an epistemic “elision” that stems in part, as I argue below, from a failure to distinguish between these senses. It is important to discuss this elision, since it reveals the need for the general category of disposition. This category frames emergence, complexity, and the realization that most systems “contain” far more information than we can fully and precisely account for. In other words, the sciences of complexity are made (epistemically) possible by the realization that many systems can be put into tractable form without taking into account all the microlevel details of the system; as with Batterman’s notion of universality, for complex-systems the details are largely irrelevant with

respect to understanding emergence—i.e., the system's details turn out not to matter concerning
the emergence of a macrocharacter which "floats free" of some of the underlying micro-detail;
and since it reflects a loss of information, there is no backtracking from it to a unique microstate
description of the system.

9.3: Emergence from an Epistemic Point of View

According to Auyang (p.175), emergence occurs relative to certain "transformations of
things." The contrast class for emergence is resultant characters, which are bounded by the
individuals contained in a system. The distinction between emergence and resultant characters,
as a first approximation, is the distinction between a structural pattern exhibiting behaviors in
accordance with Batterman's notion of universality, and a pattern accounted for solely by an
aggregation (of some kind) of the characteristics of a system's constituent individuals. The
manner in which emergence occurs relative to certain "transformations" requires further
explication.

Elemental Descriptions and Constituent Descriptions

From the standpoint of how systems are conceptually framed, a system's state has two
kinds of description for constituents (and thus two descriptions for the system). One involves the
character and (non-monadic) relation types pertaining to its micro-constituents, capturing the
features of interest when thinking about small systems. Auyang calls this the "elemental
description." The other "constituent description" also has character and relation types, but here
the character types absorb (into dispositional characters) many of the relation types used in the
elemental description.13 These types of constituents are then used in describing a system's

13 It would be tempting to claim that these higher-order "functions" take the lower-level functions as their
values. While this may hold for certain character and relation types, it is unsafe to assume that higher-order
functions are decomposable into lower-level functions. Using the rainbow example, recall that the lower-
microdescription and its macrodescription. The "microdescription" of a system is derived by aggregation of the constituent characters; by contrast, the "macrodescription" of a system introduces character types with "simpler values"—at this level, most character and relation types are qualitatively different from the character and relation types in the microdescription. ("For example, a value of a microcharacter type for a gas contains trillions of trillions of numbers for the momentum of every molecule; a value of a macrocharacter type is a single number for the pressure of the gas" (p.176).) For systems described in terms of resultant characteristics, their macrodescriptions ultimately amount to microdescriptions (the aggregation and averaging of elemental descriptions of constituents). By contrast, for other systems, examining the borderland between microdescriptions and macrodescriptions reveals emergent characters; the emergence of qualitatively different characters concerns these borderland "transformations" as they relate to the macrolevel properties of the system.

Auyang's "Left-Over" View of Emergence

Auyang has what I call a "left-over" view of emergence. An overview of her argument proceeds as follows: first, she examines models that are classified as "individual-independent"; then focusing on the residual interactions among individuals, such residues enable us to define superordinate types that bring into relief an expanded range of characteristics. "To call the balance of characters emergent is reasonable. Thus I suggest we call emergent only those macrocharacter types with no satisfactory microexplanations in individual-independent models and their modifications" (p.178). As an example of an individual-independent model, take a set of equations modeling classical particle motion. The particles are individuals (as are systems of particles), with "weak enough" couplings between them (and so are independent in the sense that they can be individuated). The residual interactions among individuals (particles, or systems of level theory "reduces" to the coarser-grained theory, but at the asymptotic limit the theories do not smoothly interrelate.
particles) disclose "new" degrees of freedom, which in turn enable a range of higher-order types (e.g., temperature, which delineates higher-order "residual" behavior among particles, or systems of particles). But from the microlevel perspective of particle mechanics, no amount of modification of the initial conditions, boundary conditions, correction parameters, etc. will yield a satisfactory account of such higher-order behavior (recall Auyang's remarks on n-atom systems; see also Auyang 1998, p.53, on temperature). The balance of characters is then, I submit, the balance between two well-accepted theories dealing with particle mechanics on the one hand (the kinetic theory of gases), and a higher-order account of gaseous behavior on the other (the phenomenal gas laws). It is a “balance” since a borderland theory (statistical mechanics) examines these “autonomous enough” theories, revealing where the residual interrelations occur. And such residual relations allow for emergent characters insofar as these characters (characters like temperature in statistical mechanics) are non-resultant—they are "left-over."

**Non-Resultant Characters as Emergent Characters**

But why exactly are these characters non-resultant? Auyang claims that a character which is not resultant is emergent (p.179), where by a resultant character she means: (1) "it is qualitatively similar to the properties of its constituents, belonging to the same character type [or superordinate type] as the constituent properties"; and (2) "its microexplanation can be given by approximately analyzing the system into independent parts with distinctive characters such that it is the sum or average of the characters of the parts, where the micro-analysis includes independent-individual models, the superposition principle, and other available means" (p.178-9). However, the claim that an emergent character is a "non"-resultant character has an ambiguous scope—is the scope of the denial a denial of (1) and a denial of (2), or more weakly is it the denial of both conditions (one condition holds but not the other)? I shall explore both interpretations in order to further articulate Auyang's intuitions about emergence.
The Second Interpretation: The Denial of One Condition but not the Other

The latter scope is considered first. There are two cases. Case (A): condition (1) holds but not (2). Case (B): condition (2) holds but not (1). For case (A), emergence would be tied to the denial of (2), and for case (B) emergence would be tied to the denial of (1).

Case (A)

In case (A), assume that for some system, (1) holds. Assume further that the denial of (2) is premised on the "unsolvability" of the n-body problem, for example. From this latter assumption, it would follow that no satisfactory microexplanation is possible (the denial of (2))—it would be impossible to get from "particle" physics to a higher-order account of aggregative behavior. Since emergence for case (A) is based on the denial of (2), the problem is that the above denial is based on the false premise of unsolvability (i.e., the n-body problem is not intractable). However, note that (2) is hedged, so perhaps the denial of (2) is that a microexplanation cannot be given by some approximate means using, say, averages. This is unclear, since there are "approximate" techniques for the n-body problem—a paradigmatic many-body problem—which appear to be employed at the "micro"-level. These techniques transform an intractable problem into tractable (approximative) form. Either way, even if (2) is denied, however that is interpreted, acceptance of (1) raises the more troublesome problem of how qualitatively similar characters illuminate what is meant by emergence. Much work would need to be done squaring the denial of (2) with the acceptance of (1).\(^\text{13}\)

\(^\text{13}\) For example, KAM theory gives solutions to the Hamiltonian under certain conditions. It is "approximative" relative to such conditions. See Diacu and Holmes (1996).

A putative counterexample to this claim might be the concept of rigidity, which can be multiply realized at different scales. For example, the rigidity of a bridge makes it seem as if we can have "qualitative similarity" across levels (the bridge is "rigid" as are its relevant "rigid" parts) and yet still have emergence. That is, we can satisfy (1) and still have emergence since the overall rigidity of the bridge is not reducible to the rigidity of its parts, as the bridge's rigidity concerns the overall structural relations between constituents. However, this putative counterexample actually supports the last point of this section; namely, even if characters like rigidity are "qualitatively similar" across levels—since the bridge's rigidity...
**Case (B)**

In case (B), we accept (2) but deny (1). First problem: acceptance of (2) does not assist in understanding emergence. For example, the solution to the n-body problem may be interpreted as supporting the "in-principle" reducibility of a system’s behavior to the behavioral sum of its parts (thus satisfying (2)). However, the solution doesn’t make much of a difference in understanding chaotic and non-chaotic dynamics (insofar as emergence relates to these dynamics), whence this lack of epistemic difference means it is *possible* for condition (2) to be compatible with emergence. To repeat, though, such mere compatibility does not assist in understanding emergence, as the n-body problem is still intractable in the second sense.

Second problem: regarding the denial of (1), which amounts to qualitatively dissimilar characters, it could be objected that qualitative dissimilarity does not suffice for emergence, since things that are “qualitatively dissimilar” might not be appropriately interrelated to produce emergent patterns of interest. And while qualitative dissimilarity may be a necessary condition for emergence, this depends on what is meant by the terms “qualitative” and “dissimilar”—an objection that also holds for claims to sufficiency.¹⁶

**The First Interpretation: Denying Each Condition**

is “similar” to the rigidity of its relevant parts—such mere “qualitatively similarity” does not help to illuminate the structural relations among parts that allow for the emergence of a bridge’s rigidity.

¹⁶ Note that cases (A) and (B) are not intended as criticisms of Auyang, but rather as further articulations of some of her (I think largely correct) intuitions about emergence. By exploring objections to the two cases, I am attempting to indicate where conceptual misunderstandings and misapplications of emergence may occur. Accordingly, concerning the denial of (1), for Auyang “qualitative dissimilarity” pertains, more precisely, to properties among dissimilar character types. For example, the macrocharacter temperature would be qualitatively dissimilar to microcharacters at the level of the kinetic theory of gases. Note also that temperature in general does not satisfy (2)—temperature generally is not to be identified with the “sum of its parts” (average kinetic energy). As Auyang writes:

Consider the most often cited definition "Temperature is the mean kinetic energy of particles." The definition is generally false. Particles streaming in a common direction have a definite mean kinetic energy, but this mean kinetic energy is not the temperature of the stream of particles; temperature is not well defined for such systems. Temperature is identifiable with mean kinetic energy only if the system is in thermal equilibrium or describable by a certain velocity distribution, both of which are system concepts (p.53).
In sum, from cases (A) and (B) together, what we have are objections to the denial and to the acceptance of (1) and (2). This seems to be an unhappy situation, but what I want to argue is that Auyang has the right intuitions, and what the two conditions need are further articulation. As a step toward this goal, now consider the former scope of the denial, which denies each of the conditions. While the above objections in cases (A) and (B) (pertaining to the denials of (2) and (1), respectively) also apply here to an extent, a significant difference can be found when each denial is read in relation to the other. Keeping in mind that each of the denials needs to be read in the context of the tools provided in section 9.2, the strategy I will employ argues that the denial of (1) is an epistemic claim about the disclosure of dispositions, and the denial of (2) is a “transcendental” claim for the category of disposition. These two denials mutually inform the problem of emergence.

The Denial of (1): Epistemically Disclosing Dispositions

The denial of (1) holds that there are (macro)characters of a composite system that are qualitatively dissimilar to the properties of its constituents (since they do not belong to the same character types as the constituent properties). To use the rainbow example again (see section 8.4), the generalized scaling law relates characters expressed by the generalized wave equation, and characters expressed by the generalized caustic equation. But the borderland character types expressed by the scaling law (in relation to the appearance of rainbows) are qualitatively dissimilar to the above characters, because the two equations do not smoothly interrelate. Recall from Chapter 2 that dispositions are epistemically parasitic on midlevel generalizations (specifically, see Ch.2, fn.20); for the case here, the “midlevel” generalizations are the well-worked-out caustic and wave equations, and the dispositional scaling law (it was argued in
section 8.2 that rainbow phenomena are expressive of a capacity) then builds upon the resources that these two equations afford.¹⁷

The tools given thus far yield a more general language for framing these ideas. Accordingly, the midlevel generalizations are a part of a microdescription, and these generalizations interrelate to rainbow phenomena and the generalized scaling law, both of which are a part of a macrodescription. Hence Auyang’s notion of “qualitatively dissimilar” characters is situated by systems with constituents of the right sort and number, the proper kinds of character and relation types that express generalizations, and appropriate micro and macro descriptions. Relative to these tools, qualitatively dissimilar characters are disclosed epistemically at the border between micro and macro descriptions—specifically, between elemental or constituent character types and macrocharacter types. (Recall that for Auyang, to “call the balance of characters emergent is reasonable. Thus I suggest we call emergent only those macrocharacter types with no satisfactory microexplanations in individual-independent models and their modifications” [emphases mine] (p. 178).) The rainbow example illustrates (scaled) qualitative dissimilarity between the midlevel wave and caustic generalizations (particular types of capacities—since they express causal “powers”), and the higher-order (dispositional) scaling law (which is epistemically parasitic on these midlevel generalizations). Most importantly, since the scaling law interrelates to raindrop-sunlight systems and perceptual systems, the overall scaling law-raindrop-observer system is expressive of both capacities (“causal powers”) and tendencies (“phenomenal laws of association”); for Batterman, these interrelations between types of dispositions situating emergence should be viewed as explanatory narratives (2002, p. 127).

¹⁷ They are “midlevel” in that the general equations express superordinate character types with respect to the microlevel relations involving particular waves and caustics; but such characters then become part of the elemental description—whence they no longer are superordinate—in relation to the superordinate types expressed by the generalized scaling law. See section 8.4, fn. 14, fn. 15, and fn. 17; see also fn. 23 below.
As dispositions are unavoidably messy, it ought not to count against Auyang that the terms "qualitative" and "dissimilar" are vague; rather it is a strength of her account that it has the right sort of flexibility which captures the relevant conceptual features from an array of examples in the sciences. What I've attempted to do is to delineate the general constraints on how such terms are employed, by virtue of the tools used to frame complex-systems thinking. (Indeed, for Auyang there will be epistemic and pragmatic elements involved in finding the independent constituents—along with mid-level dispositional character types—by which to conduct synthetic microanalysis.)

The Denial of (2): The Category of Disposition as a Condition for Understanding Emergence

The denial of (2) claims that an emergent character cannot be explained by analyzing the system into independent parts with distinctive characters such that the system is the "sum" of the parts. This may appear like an ontological claim, but the emphasis is on explanation. In particular, even if a problem is classified as intractable, this does not guarantee a strong claim to what cannot be explained (cf. the second objection in case (A) above). The conditions underlying this "explanatory lack" stem primarily from intractability, which pertains to dynamical considerations for systems with multiple constituents, whose couplings are described by the system's characters and relations. The problem of emergence in this case is not primarily about what is emergent, but rather the conditions framing the general process of emerging. In other words, even if a system "can" be analyzed into independent parts with distinctive characters (as in the existence of a solution to the n-body problem), the emerging of superordinate characters is not merely the "sum" of the system's parts; this claim is a claim about the "transcendental" conditions—the necessary presuppositions—for understanding the process of emerging. (Keep in mind that the denial of (2) is to be read in relation to the denial of (1)—they mutually inform one another as they bear upon understanding emergence. Accordingly, explanation can also "go
in the other direction"; i.e., we may already have macrocharacters that we want to understand/explain, such as the midlevel and higher-level dispositional properties alluded to in the previous section concerning the rainbow example and the denial of (I).)

Firstly, to give an explanation of a subject matter requires that something about that subject matter becomes understood. Secondly, in order to claim an understanding of X, "informativeness" is presupposed, in that without the relevant information, one cannot claim to have understood X to the same extent as one would with such information in place. Most generally speaking, something is informative to the extent that it marks a difference that makes a difference. But what conditions are required to mark a difference, and to recognize what makes a difference? Given that a difference is relational, a mark of a difference must have a prior relatum which can be compared to a subsequent relatum, and such a mark makes a difference when that difference can be noted in some respect. As this relates to the problem of emergence, what is emergent is the mark of a difference, which involves a comparison between character types from the constituent and elemental descriptions. And what makes a difference is the process of emerging that gets tracked (usually) by a phase space representation. The sense in which an emergent character isn't merely the sum of its parts has to do with a claim to informativeness. For even in the n-body problem (as noted in the first discussion for case (B)), while the solution doesn't reject emergence, it doesn't help either. That is, while the "emergent" characters involved in the n-body problem are specifiable "in-principle" by the sum of its parts (thus metaphysically eliminating emergence), from an explanatory perspective the solution makes no relevant difference.

The remaining question is, what grounds the possibility for understanding difference-makers (e.g., phase portraits) which track the process of emerging? Given a mark of a difference,

---

18 Bateson (1972).
19 Remember the importance of perspective in the discussion of the rainbow example.
such a mark must be able to be noted in some respect. This “ability” requires a space of possibilities that relates (in the minimal case) the actualization of the subsequent relatum to the prior relatum. The point is that the notions of actualization and possibility interrelate via the category of behavior, since we are focusing on dynamical considerations; i.e., the placeholders of actualization and possibility are interrelated at the conceptual level via the general category of behavior. Behavior is the category required in order to make sense of how differences can be tracked, whether the particular behaviors involve a range of what can be actualized, or what possibilities tend to be actualized. But the category of behavior in conjunction with marks of differences is just the general notion of a disposition—what I have been calling the “general category of disposition.” This general category is the “transcendental condition” for understanding the process of emerging. Thus the particular dispositions tracked in the n-body problem are “informative” relative to what gets disclosed by the qualitative phase portrait: the behavioral patterns are emergent, and the behaviors dynamically emerge as certain regions of “possibility” are “actualized.”

9.4: Levels, Thresholds, and Stances

The two previous sections may appear as a significant detour from the topic of emotions, but it turns out that they are crucial in situating the three informational stances and their relation to complex-systems thinking. As an initial characterization, “levels” are delineated by a system’s micro and macro descriptions; “thresholds” concern the emergent relations between these levels; and the informational stances capture the different granularity of emotion-systems. In fuller

\[20 \text{Depending on how one looks at it, one could assign weight to the actuality of the prior relatum that then can be related to some subsequent/possible relatum; or one could assign weight to the actualization of a subsequent relatum that is actualizable relative to some prior (possible) relatum.}
\[21 \text{This is the response to the second objection in case (A)—the use of a phase space is no longer properly viewed as a microlevel tool given its role in disclosing emergent dynamics.} \]
detail, recall that a microdescription of a system is derived by aggregation of the constituent characters, and that a macrodescription of a system introduces character types with "simpler values"—a macrostate has character and relation types, where most macrocharacters are qualitatively different from the character and relation types in the microdescription. Thresholds focus on these qualitatively dissimilar characters—that is, non-resultant characters—which emerge when comparing the system's macrodescription and microdescription. It turns out that levels and thresholds are deeply connected to information; disclosing their relation to information requires further discussion of levels, thresholds, and types of "qualitative dissimilarity."

Levels, Perspectives, and Causal Thickets

A threefold distinction made by William Wimsatt (1994) assists in identifying three types of qualitative dissimilarity. Wimsatt distinguishes between what he calls "levels," "perspectives," and "causal thickets." He characterizes "levels" as "local maxima of regularity and predictability in the phase space of alternative modes of organization of matter" (ibid). In terms of our framework, the individuation of a system is determined by its dispositional behaviors ("the phase space of alternative modes of organization"), where the behaviors focused on are the salient stabilities ("local maxima") of well-behaved dispositions ("of regularity and predictability").

Regarding our more general use of "levels," Wimsatt's particular use of the term would identify states through the characters that determine local stabilities. Consequently, the most basic sense of "qualitative dissimilarity" is just the identification of "dissimilar" states either within or between systems, not taking into account any explicit interactive relations between levels.

"Perspectives," by contrast, explicitly take into account interactions between levels, where the dispositions characterizing each level behave "differently enough" (as qualified by the local maximality condition) to individuate them as levels. This is the second sense of "qualitative dissimilarity." Lastly, "causal thickets" are the most "complex", since there are disputes over
differing perspectives and their purported boundaries. So while perspectives have organizational traction, with thickets the apparent multiple boundaries create problems in characterizing (putative) characters and states. For this third case "qualitative dissimilarity" is subject to dispute, since there are many ways to "cut the (causal) cake," and thus no unequivocal identification of emergent phenomena.

Mapping Levels-Perspectives-Causal Thickets to the Three Informational Stances

The importance of the threefold distinction is that it maps to the conceptual structure of the three informational stances. That is, levels are first-order stabilities ("functional stance"), perspectives are second-order comparisons between levels ("strategic stance"), and causal thickets are third-order comparisons between (disputed) perspectives ("semiotic stance"). However, one major difference is that Wimsatt's three terms refer to causal networks. It is important to observe that Wimsatt's causal account differs from the broader epistemic account that has been advocated thus far. To concretely illustrate the contrast, again consider the rainbow example. It would be inappropriate to classify Batterman's borderland account of emergence as a causal account, since the borderland concerns the interrelations between systems and perspectives. Indeed, as noted previously, his account of emergence, while involving causal mechanisms, is really best viewed as an explanatory narrative (2002, p.127)—a different kind of explanation of emergence.

Thus placing the causal element of the threefold distinction on the side, the upshot is that our systems-language concerning "levels" and "thresholds" is applicable to the three informational stances. In order to further expound the parallel between levels-perspectives-thickets and the three stances, the following issues will be addressed in sequence: what is

---

22 This is also partly why I claimed in section 8.2 that the rainbow example may not be an example of a capacity—it would be a capacity insofar as the wave and caustic equations express causal relations, but the standpoint from which emergence is disclosed through the interrelations between rainbow phenomena and the "asymptotic" scaling law is not appropriately classified as a causal account.
information? why "stances"? and why three informational stances? (Throughout the discussion of these issues, it should to be kept in mind that the three stances frame emotions "as" complex-systems at the meta-level.)

**What is "Information"?**

Earlier it was claimed that X is informative to the extent that X marks a difference that makes a difference. There is no universally agreed upon definition of what (semantic) information is, in part because it "is notoriously a polymorphic phenomenon and a polysemantic concept so, as an explicandum, it can be associated with several explanations, depending on the level of abstraction adopted and the cluster of requirements and desiderata orienting a theory" [emphases mine] (Floridi, 2005). In spite of this, a widely accepted working characterization of (semantic) information involves the following three conditions (ibid): p is an instance of information iff (1) p consists of one or more data, (2) the data are well-formed, and (3) the well-formed data in p are meaningful. More fully, a datum is "a putative fact regarding some difference [emphasis mine] within some context" (ibid); data are well-formed relative to a "syntax," broadly conceived; and the constraints placed on p make well-formed data in p meaningful, relative to some system. The major lesson I want to draw from these three conditions is that the characterization of information as a difference that makes a difference accords with the above definition, if suitably interpreted. Regarding condition (1), Floridi notes that a datum is neutral in four ways—taxonomically, typologically, ontologically, and genetically. (The first way is really the only one that matters; the remaining ways are also compatible with our account of information.) Taxonomically, a datum is a relational entity, but the general characterization is neutral concerning what specific relata data are. In 9.3 I noted that dispositions are required to be informative, meaning that dispositions yield (marks of) differences that make a difference. I also noted that a difference requires at least two relata, generally speaking; these relata are likewise
taxonomically neutral. Turning attention to conditions (2) and (3), *marks* of a difference—to be construed as marks—presuppose some “syntax,” and such marks *make* a difference insofar as some system of interpretation is employed. Thus our most general characterization of information (as a difference that makes a difference) accords with the features of (semantic) information sketched above.

**What is a “Stance”?**

In addition, our informational approach to framing the category of disposition accords with the above characterization of (semantic) information. For remember that the general category of disposition—the meta-level category covering the range of particular dispositions—requires marks of explanatory difference and the category of behavior. Furthermore, given that there are different levels of informational analysis pertaining to particular demands, it is only reasonable to expect that analyses of dispositions may likewise be partitioned into different levels. For remember that dispositions include a variety of “subspecies” (capacities, tendencies, and the more traditional sense of “dispositions” as “deterministic-like processes,” as with the solubility of sugar). So assuming that particular dispositions exhibit marks of difference, the issue remains how one ought to understand a system behaving in certain ways under certain conditions. This is where Dennett’s notion of a “stance” gains traction, since a stance is a predictive strategy treating selective phenomena as “objective,” and where, in turn, the adoption of the stance brings into relief such phenomena. Even stronger, (a) such phenomena can only be brought into view by that stance, and (b) the phenomena can be assessed as objective only by assessing the success which the stance affords (Dennett, 1987, p.15). For (b), it is important to notice that a stance is an *as-if* predictive strategy, i.e., it is a hypothetical construct adopting the attitude that any assertions issuing from the strategy are warranted relative to the strategy’s consequences. The justification for (a) is that stances simply work in the sciences, and are
regularly employed (p.16); more importantly, they are the only practical methods giving us any reasonable amount of predictive power for select phenomena (e.g., see p.23). These characteristics of a stance are 1) compatible with the idea that dispositions are informational (i.e., a stance, as a predictive strategy, treats dispositions as systems, and examines their consequences); 2) compatible with the project of understanding emotions as complex-systems (i.e., recall from 8.4 that viewing an emotion as a complex-system is a hypothetical, as-if construct); and 3) compatible with the notion that there are multiple relations between complex-systems and the (meta-level) category of disposition (i.e. there are multiple types of stances, or informational levels of analysis).

Why Three Informational Stances?

So why three informational stances? The answer is that the threefold distinction maps to the earlier distinctions between levels, perspectives, and thickets. A general way to express the threefold distinction is that it distinguishes between single-level analyses, between-level analyses, and multiple-level analyses. However, the notion of “levels” must be qualified. First, there is a sense in which higher levels “build upon” the lower levels, to the extent that each informational stance discloses certain phenomena. The operative sense I have in mind that distinguishes the three stances from one another goes as follows. The functional stance black-boxes developmental processes, and concerns itself primarily with the ways in which different behaviors manifest themselves (reliably) under appropriate conditions; if such behavior is judged to be “reliable enough,” then the functional stance has done some work toward further understanding those processes. The strategic stance looks at relations between these boxes to delineate patterns of interaction that the functional stance might not so readily discern. In a similar fashion, if this stance reveals behaviors that look “reliable enough,” then it too has gained some claim to “projectibility.” The same general lesson cautiously applies to the semiotic
stance; i.e., since there are multiple-level interactions between systems, the behaviors are more heterogeneous. Thus judging what constitutes reliable behavior becomes problematic when adopting the semiotic stance. What this stance primarily affords is a way to keep track of crucial presuppositions in representing putative patterns of behavior.

It is important to keep in mind that the levels are not constitutive, "ontological" things where, for example, the levels gain the reliable traction they do because they really refer to processes in the world. Such a picture would raise a host of problems that should be avoided.23 Furthermore, while higher levels build upon the lower levels, the two lower stances actually presuppose the "highest" stance (recall section 6.2). To make this clear, it suffices to note that the functional stance already presupposes a semiotic distinction between sign, signified, and interpretant (the capacity in which the sign and signified are related), since a boxology is a sign/representation of a signified process, where the aptness (capacity) of such a representation is determined in part by what reliable consequences that representation affords. Clearly the same sort of lesson holds for how a semiotic distinction is presupposed in the strategic stance as well. So, in brief, my answer to the question "why three informational stances?" is two-fold, namely that single-level, between-level, and multiple-level analyses appear to capture the general types of analyses that can be undertaken from a systems standpoint; and secondly, these three stances keep track of the different reliable consequences which representational schemes afford.24 (In this two-fold sense the stances are conceptual moments needed to understand system-processes.)

Complexity and the Three Informational Stances

The final task in this section is to give an overview of the link between complexity and the three informational stances. Since most of the labor has already been done above, the

23 For critiques of (reified) conceptions of "levelism" see Heil (2003) and Schaffer (2003).
24 Although the semiotic stance tracks the assumptions regarding representations which problematize "reliable consequences."
discussion will be brief. Thus far complex-systems thinking has granted us the rubric of levels, thresholds, emergence, etc. Wimsatt's particular senses of "levels," "perspectives," and "thickets" are expressible within our broader language of levels—i.e., single-level analyses, between-level analyses, and multiple-level analyses. Given the parallel between these three levels and the three informational stances, the tools of complex-systems thinking also are applicable to organizing the four emotions research programs. Additionally, since emphasis is placed on the conceptual complexity framework—a set of "root-metaphors" for thinking about and framing putatively complex-systems—the three informational stances are likewise tools for thinking about and framing the emotions research programs (and their investigations of particular emotions).

In conclusion, to summarize section 9.1 through to the above, there are several senses in which emotions may be viewed as complex-systems. First, emotions may actually be (activations of) complex-systems, as recent neuro-physiological studies appear to suggest. (More generally, biological systems appear to exhibit non-linear self-organizing behavior; witness the informal phrase, "chaos is the very stuff of life.") Secondly, FIPS and dynamical modes appear to exhibit the features of complexity; that is, the major concepts of the functional and developmental-social approaches to emotions implicitly employ the tools of complex-systems thinking. Both of these first two points were covered in section 9.1. The third sense in which emotions "are" complex-systems highlights the meta-level. First, the three informational stances capture the conceptual features of the four emotions research programs at the meta-level—i.e., the concepts of the functional approaches are captured by the (meta-level) functional stance, and the concepts of the developmental-social approaches are captured by the (meta-level) strategic stance and the (meta-level) semiotic stance. Secondly, since the informational stances link with complex-systems thinking, the meta-level tools afforded by the complexity framework can also be applied to the
emotions research programs. (In particular, the functional approaches to emotions are conceptually framed as a “state space” representation, and the developmental approaches are conceptually framed as a “dynamical” representation.) And thirdly, the tools of complex-systems thinking—levels, thresholds, emergence, etc.—can be applied at this meta-level to represent changing relations between research programs, where these shifting relations concern possible borderland accounts by new, “emergent” emotions research programs.

9.5: Toward a Science of Emotion

Overview of Chapters Two Through Nine

Is a science of emotion possible? Thus far I have broken this question down into several parts: first, what is meant by “science”? The liberal conception of science employed through the previous chapters is one adopting a “heterogeneous perspective.” I argued that our four-dimensional space organizes different kinds of “sciences,” whether paradigms, research programs, or philosophical frameworks, and as a consequence, it is plausible to speak of various sciences of emotions which go beyond the narrow concern with projectibility. The second move I’ve made is to examine actual emotions research programs, using the four-dimensional space to align their differing structures. If these four programs represent the current major approaches to emotions, the four-dimensional space organizes what it generally means to speak of (contemporary) sciences of emotions—i.e., each of the contemporary programs can be called a “science” in some sense.

The third move I’ve made links the four major emotions research programs with the concepts and tools coming from “biology” and “psychology.” Looking ahead to the final section,

---

More precisely, recall in Chapter 2 that “causal homeostasis” was replaced by the widened notion of “resiliency,” in addition to other desiderata of heterogeneous sciences indicated by the 4-D framework. It should also be noted that Griffiths’s later works acknowledge the importance of more “ecological” theories of emotions. See Griffiths (2004a) and Griffiths and Scarantino (in press).
if it turns out to be the case that future research programs are extensions of biology, then a science of emotion would be possible, since biology’s consilient framework—neglecting for the moment the tension between the received view and DST—would also encompass the heterogeneous category of emotion. In other words, just as biology makes possible a science of life (even though the heterogeneous details of life-processes are not fully known), likewise biology may provide a research agenda for investigating the category of emotion (again, I discuss this speculative point more fully in the final section).

The fourth move I’ve made is the use of complex-systems thinking, and its potential application to the tension in biology between the received view and DST. The complexity framework also offers a potential way of framing the differing sciences of emotions covered in chapters three through six, given the general parallels between biology and psychology. If this framework provides “elegant” (and not merely ad hoc) organization of the four emotions research programs, then we would have a means of establishing a possible science of emotions, above and beyond just a conceptual organization of these sciences of emotions (i.e., the “conceptual shapes” provided by the three informational stances).

Overview of the Remaining Argument

However, I have still not addressed whether a science of emotion is possible. All the hard work up until now was just to suggest a theoretical map for why there may be reason to hope that a science of emotions may be possible. The transition to the general, “all-encompassing” category of emotion—if this even exists—is much more difficult, and of necessity, highly speculative. Keeping the above in mind, the remainder of this chapter will do just that—speculate about what bearing complex-systems thinking has on a science of emotion. To do this, the discussion will run in three parts. First, I will engage in some semantic “spadework” on the use of “emotion” and its relation to the emerging field of affective science. Secondly, at the
meta-level, the complexity framework may suggest future borderland approaches defining novel research programs. In this case, not only would the conceptual complexity framework suggest new, qualitatively "emergent" disciplines, but also new, emergent types of emotions (especially at the higher social and population levels). And lastly, I will revisit the question, what sense does it make to speak of a "science" concerning a category as general and variegated as the category of emotion?

9.5.1: Emotion and Affect

Two Reasons for the Shift from "Emotion" to "Affect"

As mentioned in the Preface and Chapter 8, "emotion" is a term of art employed by various research programs. However, the widening range of scientific investigations about emotive phenomena has led to a recent shift in terminology reflecting this expanding demand. There are two general reasons why the shift from "emotion" to "affect" makes sense. First, because "emotion" carries certain pre-theoretic associations pertaining to felt subjective states—in particular, "emotion" usually connotes an affective state, of short duration, whose felt quality has an intensity and "texture"—these phenomenological aspects of emotion exhibit tangled relations to scientific senses of "emotion." Here is a brief recap of some of these tangles. (1) For conceptual analyses based on folk intuitions, recall from chapter three that Griffiths' objections to \textit{a priori} analytic accounts of emotions concern the sustained anomalies these accounts encounter relative to scientific work on emotions (more subtly, Griffiths advocates a kind of conceptual analysis, but the analysis should take seriously contextualized scientific accounts). (2) Ekman's program: affect programs are viable scientific concepts, since they are causally homeostatic, although their modular characteristics do not fully mesh with folk intuitions about emotions. (3) Evolutionary psychology: folk categories (like folk biology, for example) work well for certain
purposes, but they also get "unhinged"—i.e., research reveals certain systematic mismatches between folk testimony and operative mechanisms—meaning, in the context of FIIPS, that FIIPS will have both systematic matches with aspects of folk categories, as well as systematic mismatches with these categories. (4) Cognitivist Program: folk intuitions cohere with cognitivist interpretations of emotions (e.g., valence as a pleasure-pain continuum), where degrees of, say, pain, in juxtaposition with the formal objects of appraisals, yield modes of activation that are classified as emotion-judgments; yet in spite of this, cognitivist modes of activation are far more abstract (e.g., potentially infinite modes of activation) than folk intuitions allow. (5) Social Constructivist Program: folk intuitions partially mesh with social classifications of emotions, but when emotions are syndromes (covert constructs), clearly experience is disjointed from a theoretical understanding of how these emotions are socially constructed.

The second reason for the shift from "emotion" to "affect" is due to the internal demands of scientific practices, which require an increased range of distinctions concerning affective phenomena. In other words, given the modern acknowledgement that "emotion" is not strongly distinct from "cognition," scientific intuitions have shifted about what emotions are; for while "emotion" is still a term of art, as research has evolved, so has the need for better working characterizations of various "affective phenomena." Since "affect" connotes a wider category compared to pre-theoretic intuitions about "emotion"—intuitions that nevertheless are still tacitly employed, in some capacity, by researchers—the new field of affective sciences includes a diverse array of research programs. Accordingly, a road map of the major working characterizations is provided in what follows.²⁶

"Emotion"

²⁶ As alluded to in the Preface, thus far my use of "emotion," in the context of the major research programs, can be read as shorthand for "affective phenomena."
In the *Handbook of Affective Sciences* (2003), six types of affective phenomena are distinguished (in terms of a working scientific, third-person language). The first is *emotion*, which "refers to a relatively brief episode of coordinated brain, autonomic, and behavioral changes that facilitate a response to an internal or external event of significance for the organism" (p.xiii). Regarding our four emotions research programs, these coordinated changes may be represented, *in third-person terms*, as affect programs, FLIPS, or appraisals, depending on the crucial parameter of (brief) temporal duration. In particular, since affect programs are of brief duration, and have adaptive significance for a wide range of species, affect programs fall under this general characterization of "emotion." Likewise, appraisals and other FLIPS, assuming they remain within "acceptable" temporal constraints, may count as emotions.

*"Feelings"*

Secondly, *feelings* refer to the "subjective representation of emotions" (*ibid*), meaning that they refer to the subjective representation of certain relatively brief and coordinated neurophysiological changes. Affect programs, FLIPS, and appraisals are scientific (third-person) representations for understanding emotion-systems; "feelings" (as a *third-person* term) concerns the first-person experiences of the states represented. Research programs implicitly acknowledge feelings. For example, the *intentional stance*, as used to understand FLIPS, is a third-person perspective (referencing first-person systems) which implicitly relies on first-person experiences. More generally, third-person representations of emotions tend to rely on testimony—third-person reports of first-person experiences.

*"Attitudes"*

Thirdly, *attitudes* are "relatively enduring, affectively colored beliefs, preferences, and predispositions toward objects or persons" (*ibid*). Attitudes are frameable by the intentional stance. First, note that the *functional stance* applies primarily to functional approaches to
emotions, where the emphasis is on reverse engineering (particularly the computational level). However, the computational level, in producing “boxologies,” deemphasizes attitudes (see sections 3.2 and 4.2); with respect to functional approaches, it is the intentional stance that is used to understand attitudes. Additionally, the intentional stance coheres with the Cognitivist Program’s use of appraisals. For recall that the Cognitivist Program is generally capturable by the strategic stance, which applies the intentional stance to second-order intentional systems. So whether we are talking about individual attitudes (first-order intentional systems) as they relate to appraisals, or more ecological “attitudes” (second-order intentional systems) as they relate to higher-grained appraisal-systems, it remains that the intentional stance squares with cognitivist approaches.

“Mood”

Fourthly, mood “typically refers to a diffuse affective state that is often [though not always] of lower intensity than emotion, but considerably longer in duration. Moods are not usually associated with patterned expressive signs [e.g., precise facial expressions for affect programs] that typically accompany emotion [in the sense mentioned above], and sometimes occur without apparent cause” (ibid). Clearly affect programs are distinct from moods. “Wider” types of FIIPS may be compatible with moods, in spite of the fact that FIIPS tend not to focus on temporal factors. However, the diffuse nature of moods fits better with the developmental-social approaches to emotions. Appraisals as modes of activation explicitly take into account temporal factors concerning parameters like valence, intensity, etc.—tools which enable far greater traction regarding moods. (Note that these tools are applicable even when there apparently is no “formal object” which may “appraise”/“trigger” a mood—an objectless mood.) Additionally, developmental approaches allow for “moods” at various group/population levels (e.g. “a nation’s
fearful mood,” classified as a social constructivist mode of activation) that move beyond the focus on individual dynamics.

“Affective Style”

Fifthly, building upon emotions, feelings, attitudes, and moods, the “macrocharacter” that absorbs these lower-level parameters is the notion of affective style, which refers to “relatively stable dispositions that bias an individual toward perceiving and responding to people and objects with a particular emotional quality, emotional dimension, or mood” (ibid). First, moving beyond temporally constrained states (like emotions), moods are “macrolevel” modes of activation that differentially tap into “microlevel” emotions, attitudes, and feelings; but looking at the larger dispositional landscape, moods are midlevel, temporally “stretched” states. Moods, along with emotions, feelings, and attitudes, may jointly form a microlevel for speaking about macrolevel affective styles. These styles are complexity landscapes that allow researchers to see the trajectories of types of personalities (Lewis and Granic 2000). People, quite starkly, are “affective creatures” through and through—the broad notion of “dispositions,” stable or otherwise, gives us a means of situating stable basins of attraction (multiply realizable “functions”) as well as possible shifts in landscapes (dynamical modes of activation).

“Temperament”

Lastly, the sixth macrocharacter, which goes hand-in-hand with affective styles, is the notion of temperament. This refers to “particular affective lifestyles that are apparent early in life, and thus [may have a significant “genetic” influence]” (ibid). At the developmental level, temperament is akin to the morphological field concept; and at the ecological level, the roles that this key “unit” plays would be important for developmental approaches to emotions. If affective styles are matured stable basins of attraction within an “adaptive landscape,” temperament is the field that enables a constrained range of developmental outcomes. In other words, temperament is
the "developmental middle level" between population level affective styles and individual developmental potentials.

9.5.2: Affect and Complexity: Borderland Research Programs

Thus far the complexity framework provided offers tools for organizing a possible science of emotions, or more appropriately, a science of affects (recall the remarks in the Preface). With respect to the discussion in section 9.1, the sense in which affects "are" complex-systems applies primarily at the "object level" (i.e., affects are actually (activations of) complex-systems; or complex-systems representations assist in modeling affective phenomena). This section picks up on the discussion at the end of section 9.4, and speculates on the third, meta-level sense in which affects "are" complex-systems. At the meta-level, the issue concerns the informational stances and the borderland relations between their corresponding 4-D shapes. First let us briefly review the informational stances individually.

The Functional and Developmental-Social Approaches to Affects and the Three Informational Stances

For functional approaches to affects, recall that the central concept is that affects are FIIPS. Several limits to this general approach were discussed in chapter seven. Moving to the meta-level, the functional stance delimits the concepts of Ekman's program and evolutionary psychology. For developmental-social approaches to affects, recall that the central concept is that affects are dynamical modes. Likewise, in chapter seven several limits were discussed. Moving to the meta-level, the strategic stance delimits the general features of the Cognitivist Program; however, there is "overlap" since the more individualistic aspects of this program are in part capturable by the functional stance (i.e., the intentional stance frames the psychological categories that are used in dimensional representations). As for the Social Constructivist
Program, there is also overlap, since the strategic stance captures the program's population modeling assumptions, while the semiotic stance captures the qualitative features of (covert) affective social constructs.

Thus the three 4-D shapes carved out by the informational stances have regions of overlap, indicating that while the stances delineate qualitatively different perspectives—that is, the functional stance is "first-order," the strategic stance is "second-order," and the semiotic stance is "third-order" (see section 9.4)—the stances are also intertwined with one another. These two joint characteristics of qualitatively different perspectives with regions of overlap lend credence to the claim that the meta-level stances may also be viewed as (meta) systems embedded within (meta) systems. So, for example, with respect to a borderland affect research program, the functional stance may delineate a microlevel, and the strategic stance may delineate a macrolevel. "Emergence"—at this conceptual meta-level using the qualitative concepts afforded by the complexity framework—would be twofold: there would be the "emergence" of a new research program, as well as the "emergence" of novel affective modes investigated by this borderland account. However, it is important to reemphasize that these levels are not fixed levels of analysis. For example: the microlevel could come from the semiotic stance (e.g., viewing covert affective syndromes as "memes" operating on individuals) and the borderland account could use the strategic stance to view a population of (competing) memes as the macrolevel; the macrolevel could come from the functional stance (e.g., some FIIIS) and the microlevel could use the strategic stance to understand neural-level populations (e.g., neural "strategies" concerning "neuronal group selection"); and so on.

Speculations about Future Borderland Accounts

Thus the first speculative point is that borderland accounts draw upon the 4-D shapes delineated by the informational stances, and new research programs may be conceptually
envisioned as emergent 4-D shapes “connecting” regions from these three shapes. Thus a science of affects includes potentially new research programs building on the resources of extant research programs.

The second speculative point is that since affects have different kinds of grain (i.e., the working distinctions between emotions, feeling, moods, etc.), a complex-systems approach provides a means for dealing with grain. Additionally, it is plausible to link the study of affective grain at the “object level” with emergent 4-D shapes at the meta-level. Thus, while particular studies of affective phenomena may suggest new research programs (a “bottom-up” approach), by contrast, starting at the meta-level, speculative 4-D shapes may direct future particular studies of affective phenomena neglected by existing programs (a “top-down” approach).

The third speculative point is that a science of affects will most likely be highly heterogeneous. Yet in spite of the heterogeneity of affective phenomena, at the meta-level, if our four-dimensional space generally organizes heterogeneous sciences, then various 4-D shapes would still fall within our conceptual framework. Furthermore, emergent 4-D shapes would be framed by complex-systems thinking applied to our four-dimensional space. Thus a science of affects would be “unified” by the tools of complexity. Such “unity” would differ from the senses of “unity” stemming from traditional philosophy of science; for given that biology is a model science concerning heterogeneity, “consilience,” rather than “unity,” would be the operative concept.

9.5.3: The Category of Emotion: Complexity as a Framework for Heterogeneity?

If the complexity framework captures affect research programs in consilient fashion, there would be a possible science of affects. Additionally, if it turns out that the general category of emotion is just the range of possible affects examined by research programs, then a science of
emotion would be possible, since a science of affects would be coextensive with a science of (the general category of) emotion. But is the general category of emotion/affect exhaustible by a science of affects? In other words, do the theoretical boundaries of science include all possible affects, and thus the general category of emotion?

The Qualia Objection

One objection to establishing a science of emotion is that there may be affects which are out of the reach of scientific inquiry. For example, if first-person experiences of affects are centrally about the “ineffable” quality of what-it-feels-like, akin to the qualia of consciousness, then perhaps no scientific account can capture such affective qualia. Indeed, debates in philosophy of mind concerning consciousness often turn on the qualia problem—briefly, how could materialistic processes give rise to the qualia, the “what-it-feels-like,” of conscious states? The same form of the problem is applicable to the qualia of affective experience, since consciousness is one particular “brand” of affective experience. Interpreting “materialistic processes” as the domain of scientific investigation, and interpreting “qualia” as the domain of felt, first-person experiences of affects, an objection to a possible science of emotion is that the general category of emotion includes such first-person experiences, but science by its very “nature” (as providing third-person accounts) cannot capture these experiences. Thus while a science of affects may be possible, a science of “emotion”—a science concerning the general “all-encompassing” category, if it even exists—is not.

Structure of the Remaining Argument

I will place on the side objections akin to the above. By the category of emotion, I still primarily have in mind the types of third-person accounts that investigate heterogeneous affects. Granting that the general category of emotion might exceed the boundaries of science, especially

---

27 Indeed, the types of arguments concerning the “hard problem” of consciousness are applicable here. See, for example, Chalmers (1997).
when it comes to first-person accounts, I want to concentrate on the manner in which a "science" of affects may exhibit consilience, thus potentially giving "unity" to the category of emotion from a "scientific" perspective. In other words, this final section brings us back to the first and second chapters, and the focus on "science." There are three issues in particular that I will discuss. The first issue concerns the four affect research programs and their tangled relations to the human sciences. The second revisits the meta-level debate between DIT and DST, and the Social Constructivist Program's two competing philosophical frameworks over what values a science ought to adopt. The third issue revisits the general parallel between biology and psychology; specifically, I focus on evolution as the consilient framework by which all of biology "hangs together."

The Bigger Picture: The Four Affect Research Programs and Their Tangled Relations to the Human Sciences

Recall that Ekman's program distinguishes between affect programs and display rules, where most higher-order, socially informed senses of "emotions" pertain to display rules and the unspecified relations that affect programs have to these rules. Display rules implicitly appeal to an indefinite range of human-science concepts. It appears that while affect programs are "properly" scientific, these higher-order senses of "emotions" are comparatively less "scientific" (as their concepts exhibit relatively less projectibility). (See Chapter 3 for more.)

Recall that evolutionary psychology investigates more centralized processes (compared to Ekman's program). Recall also that there are two general distinctions indicating where human-science concepts may become more prominent: 1) since evolutionary psychology looks for cognitive functional adaptations (FIIPS), to the extent that various human sciences bear upon this search, these fields would be constrained (and mutually informed) by the primary search for FIIPS; or 2) a stronger distinction may be made between the search for various functional
(adaptive) processes, and "other" human science concepts which are concerned with non-functional (non-adaptationist) processes. (See Chapter 4 for more.)

The developmental-social approaches to affects, by contrast to the above functional approaches, are more explicitly situated within the human sciences. Recall that the Cognitivist Program, while emphasizing the causal nature of its abstract dimensional activation spaces ("causal" in that the mathematical activation spaces model, in some capacity, the neuro-physiological processes underpinning affective experiences), explicitly starts with a mid-level orientation respecting the "lived" aspect of affects. That is, recall that dimensional representations attempt to map fields of affective experience, and that the Cognitivist Program appropriates these phenomenological representations to yield causal-abstract-phenomenological representations of appraisals. These phenomenological considerations draw from an array of human-science concepts concerning folk senses of "emotion." (See Chapter 5 for more.)

Finally, the Social Constructivist Program also begins with a mid-level orientation respecting the phenomenological dimensions of affective experience; recall that what this program emphasizes are various social structures and their dynamic relations with appraisals. Recall also that various types of constructivisms are situated within the human sciences; most importantly, the major difference concerns how these approaches conceive of investigating affects—that is, the philosophical frameworks concern "dualistic" modeling of affect-structures on the one hand (DIT), and "dialectical" conceptions on the other (DST). (See Chapter 6 for more.)

**DIT and DST: Differing Conceptions of "Science"**

To begin, first note that both DIT and DST offer third-person accounts. DIT's model are "dualistic," separating the representation from the represented; DST's models are "dialectical," tracking the interrelations between representations at the "object level," and the potential changes.
that a representation has on the represented at the “meta-level.” Still, the models are *models*, not first-person experiences. Keeping this in mind, a problem for a prospective science of emotion stems from the aspect of the Social Constructivist Program not capturable by DIT, namely the qualitative dimension of the Social Constructivist Program. The qualitative dimension emphasizes that affect syndromes are *dynamically* nominalistic (i.e., they are covert constructs). Recall from chapter six that philosophical disputes between DIT and DST amount to disputes at the meta-level over the core assumptions of each philosophical framework. At this level, DIT and DST offer competing views of science. DIT generally presupposes that models of dynamic processes “dualistically” match sign with signified, where any change to what is signified because of the sign is negligible. The existence of affect syndromes problematizes such an assumption, which is why this aspect of the Social Constructivist Program cannot be captured by DIT. By contrast, DST assumes that sign and signified “dialectically” inform each other. The problem here, though, is that if DST isn’t to be merely a philosophical framework, the *differential* changes in relations between sign and signified require “empirical” documentation. For example, with respect to biology, first, at the *object-level* (focusing on interrelations between signs), *EvoDevo*’s experimental documentation of differential changes still significantly underdetermines the extent to which most evolutionary changes are due to regulatory shifts; and secondly, at the *meta-level* (focusing on interrelations between sign and signified), DST is on shakier ground (compared with DIT, which is also theoretically rich and “empirically” poor—in terms of the relative lack of studies carried out—though comparatively less so), since neither population developmental biology nor its cultural extension yet exist. Thus, at the level of philosophical frameworks, disputes between DST and DIT really do amount to philosophical disputes about conceptions of science and how science ought to proceed on the basis of these conceptions.
At this level, it would appear that there is a problem in establishing a science of emotion. On the one hand, the dualistic picture of OIT would hold that modeling affects does \textit{not} significantly change the phenomena; our modeling practices do not render impossible the project of making sense of the category of emotion. For even if new affect-characters arise in the process of cultural evolution, the population dynamic models of OIT can still (retrospectively) track the general evolutionary process. Just as evolutionary biology makes sense of the category of life, OIT (or some variant thereof) can make sense of the category of emotion. But on the other hand, DST's dialectical picture would hold that modeling affects might significantly change the phenomena; our modeling practices may make the category of emotion intractable. First, we might be modeling a "moving target" that shifts as we model it. And second, affects may significantly change in fundamentally unforeseeable, intractable ways \textit{because} of our modeling practices (recall the nature of affect syndromes)—our very modeling practices may be responsible for "novel" affective phenomena. Indeed, the general problem with the human sciences is that they investigate subject matters which can be modified by the very theories used to try to understand them.

\textit{Evolution as an Indefinitely "Evolving" Science}

However, from the standpoint of evolution, the above problem is only apparent. Emergent phenomena are not an "architectural" problem for the life sciences; rather the heterogeneity of emergence is the starting place for \textit{understanding} the complicated nature of biotic-abiotic interactions.\footnote{In addition, recall that complex-systems thinking applies primarily to many-body systems which have a certain degree of tractability; Auyang's analysis does not focus on highly complex (and potentially "fundamentally" intractable) systems like cybernetic systems (see section 9.2). These latter systems, while \textit{qualitatively} representable using the conceptual tools of complex-systems thinking, are intractable in terms of representing the full complexity of the system's "nested" relations. In other words, supposing that the full complexity of such systems matters (unlike many-body systems whose details are largely irrelevant with respect to "universality"), these systems would actually be \textit{transformed} into qualitative many-body}
DIT and DST, even if our modeling practices change the phenomena being studied (in particular, even if our practices give rise to emergent affective patterns of behavior), this is not appropriately viewed as a problem for a science of emotion. Informally, evolution and the tools for studying evolutionary processes themselves "evolve." Clarifying this idea brings us to the third issue, namely the focus on consilience and the parallel between biology and psychology.

Recall that complex-systems thinking implicitly uses the general category of disposition when qualitatively organizing the application of the informational stances to heterogeneous affect research programs and their affect-systems. While precise physico-mathematical models illustrate complex-systems thinking (e.g., statistical mechanics), when it comes to the special sciences and their heavy use of (non-mathematical) concepts, mathematical tools like phase spaces are replaced by qualitative notions such as the general category of disposition. As it pertains to evolutionary considerations, there is a deep sense in which the general category of disposition frames an understanding of evolutionary processes. This claim accounts for the manner in which evolution and the tools for studying evolutionary processes themselves "evolve," as I explain below.

In "The Evolutionary Contingency Thesis" (1995), John Beatty argues that biological generalizations are historically contingent. Briefly, the idea in strong form is that if the "tape of life" were played again—Stephen Jay Gould's notion that if we "rewound" the history of life and let things unfold from there—the results might be radically different from our history (p.57-8). More weakly, various "rule-making" capabilities of evolutionary agents (agents like mutation, natural selection, epigenetic rules, etc.) not only have the capacity to make further

representations. The point is that such transformations—qualitative or quantitative—are the norm in science; science, by its very nature, "reduces" complexity through abstraction.
"rules/generalizations," they also have the ability to break these rules. From the viewpoint of natural history, while principles like the principle of natural selection (as well as other potentially complementary principles coming from, say, developmental biology) might help to understand past evolutionary events—as well as to frame certain local predictions—the implication is that particular formulations of dispositions will not exhaust the category of disposition in general. Why? (Besides the obvious and unhelpful point that that is what distinguishes particular from general.) There is a straightforward “empirical” reason that Beatty appeals to. Evolutionarily well-informed accounts of biological phenomena give no reason (inductively) to presume that the historical outcomes of evolution are precisely and narrowly constrained. (If we did have inductive reason to assume such constraints, it would allow for explanation via particular dispositions.) Hence we ought to be “on the lookout for multiple accounts of each [biological] domain [of inquiry]” (p.75). In a more metaphysical vein, the general picture of evolutionary change suggests that not only do the rules (particular empirical generalizations) evolve, the evolutionary agents themselves evolve as new material conditions arise (p.75).

The “rules” of evolution are just “low-level” dispositions, whether they are basins of attraction in general ecological models, say, or potential adaptive peaks within adaptive landscape models (at the population dynamic level). The evolutionary agents are the “mid-level” dispositions driving change. The metaphysical picture offered by Beatty is that not only can the rules change (e.g., basins of attraction change with different ecological parameters), but the

---

29 The idea is that “what the agents of evolution render general, they may later render rare” (p.53). For example, at one time evolution favored “the rule” of hairiness in our lineage, but later opted for hairlessness (p.53). Another example discussed is the Krebs cycle (p.54). Here I advocate only the weak contingency thesis, which is agnostic as to whether a radically different history would result if the tape of life were played again.

30 Mitchell (2002) puts the former point as follows: “The domain of alternative evolutionary solutions [particular “rules”] to adaptive problems defines a form of complexity. This consists in the wide diversity of forms of life that have evolved despite facing similar adaptive challenges” (p.336). An arguable example of the latter point where evolutionary agents “evolve” is the current group-selection debate as it relates to cultural evolution and differing models of what mechanisms underpin such evolution.
evolutionary agents themselves can “evolve” (e.g., group selective mechanisms; shifts in fields
giving rise to novel regulatory networks). The latter point is actually more radical: it creates a
space for new evolutionary principles at the scientific meta-level, where there may be extensions
to accepted principles like natural selection. An example of this speculative possibility, I think, is
exhibited in the novel tools of DIT, which extend the received view to the realm of cultural
evolution (especially since DIT involves Lamarckian modes of “inheritance”; see Chapter 6).
And as I mentioned in chapter seven, another potential extension would be a cultural analogue of
population regulatory genetics. In any event, at the meta-level, the general category of
disposition creates an epistemic space framing the manner in which evolutionary theory may
“evolve.”

Thus the category of disposition accommodates the differing scientific values of DIT and
DST, since both evolutionary frameworks fall within the epistemic space afforded by this
category. More generally, Beatty’s argument suggests that the category of disposition frames the
“metaphysics” of evolution. That is, the consilience of evolution (probably) applies to future
ways in which evolutionary theory evolves; while we don’t know the shape of evolutionary
theory to come—especially concerning its extensions into the cultural realm—the sprawling
structure of evolutionary theory will still (probably) be consilient. Hence, 1) given that the
general category of disposition creates an epistemic space for the metaphysics of evolution; 2)
given that this category is presupposed by the complexity framework (sections 9.2 and 9.3); and
3) given that the complexity framework may be applicable to a science of affects, it follows that
the category of disposition appears to frame both evolution and a science of affects. Furthermore,
this category would also frame the general category of emotion, since it is being assumed that
from a scientific perspective, the category of emotion is coextensive with the indefinite
“evolving” range of affects examined by research programs. The final topic to discuss, then, is
consilience. For as the category of disposition links the "evolving" architecture of evolutionary theory with the category of emotion, if the former is consilient, then consilience would also apply to the general category of emotion.

*Consilience of the Category of Emotion? Emotion as a Regulative Ideal*

The argument for the consilience of an extended evolutionary theory, while speculative, is based on an appeal to inductive evidence. First, evolution, generally speaking, continues to garner increasing support from its numerous domains of investigation. That is, biology generally exhibits consilience: its growing, coherent web consists of independent, yet convergent, lines of investigation (from semi-autonomous fields of investigation). Furthermore, belief in evolution stems from its internal history, which exhibits consilience between established fields, as well as between new and established fields. Thus there is inductive reason to think that evolution, as a scientific framework, will work in the future, judging from its past success. Secondly, evolutionary theorizing has just begun to be applied to the cultural realm. Evolutionary biology itself is a relatively young scientific field, at least with respect to the founding of the Modern Synthesis. Given the significant inductive support for evolutionary theorizing from the standpoint of its internal history, an analogous argument may be made for extensions "external" to the traditionally conceived boundaries of biology (e.g., the typical distinction made between biology and culture). As with evolution's internal consilience—the consilience of evolutionary theory itself, and biology's past success—there may be inductive hope for expanding the evolutionary domain so that its "external" extensions also exhibit consilience. It is in this second sense that an extended evolutionary theory will "probably" exhibit consilience. In particular,
consilience would hold within and between established and extended domains, since the extended domains "evolve" by bootstrapping off of already "projectible" domains.\footnote{A mathematical analogy might help to illustrate the structure of the argument here. Actually establishing that a science of emotion is possible is not really possible, I think; it would be akin to assuming a "Platonic" view of mathematics, and showing "existence" either by example, or by indirect proof. Rather I opt for an "intuitionistic" argument, where there is reason to hope that a science of emotion may be possible, on the basis of an indefinitely evolving "science." Thus it is not clear if the all-encompassing category of emotion even exists; the most that can be said is that emotion serves as a regulatory ideal guiding investigations of various particular emotive-cognitive processes.}

This speculative inductive appeal differs markedly from Griffith's use of inductive concerns. Recall from chapter one that causal homeostasis is the mechanism by which scientific categories apparently do their work; i.e., categories are scientific insofar as they refer to "natural kinds," where natural kinds are picked out by those categories exhibiting causal homeostasis. Causal homeostasis is one way to find workable inductive inferences that are worthy of being labeled "scientific." However, my whole project has attempted to reconfigure these intuitions about science by adopting a heterogeneous viewpoint. In accordance with the different intuitions provided by the four-dimensional space and the new intuitions coming from complex-systems thinking, it should be no surprise that the speculative inductive argument offered above does not appeal to causal homeostasis. Rather, the broader intuitions underlying this meta-level argument stem from 1) the focus on biology's consilience as organized around the notion of evolutionary narratives (chapters two and seven); 2) the parallel between biology and psychology, and the links to the major affect research programs (chapters two through seven); 3) the expectation that the affect research programs will also exhibit types of consilience (relative to the search for FIIPS, and the search for dynamical modes of activation); 4) the use of the complexity framework as a potential means to "unify" the differing kinds of consilience exhibited by the received view and EvoDevo (chapter eight); and 5) the use of the complexity framework as a potential means to frame a consilient science of affects (chapters eight and nine).
The point is that evolution is still "one long argument," which conveys the idea that different kinds of consilience still demand some sort of integration. My speculative hypothesis is that complex-systems thinking may provide the tools for establishing a possible science of affects—for integrating different affect research programs (see fn.31). But it is not clear if the general category of emotion even exists, as it would include all possible affective processes. Still, given the contemporary acknowledgement that we cannot divide "emotion" from "cognition"—even stronger, it seems that all centralized systems are emotive-cognitive (or simply "affective") systems—and given that contemporary research programs are actively investigating variegated affective systems, it appears that it would be unsafe to simply discard the "general category of emotion." For even if this category doesn't exist, the hope to better understand centralized systems expresses the hope to understand "the mind." Is the mind nothing but affective systems? Is the mind "equivalent" to the general category of emotion? Nobody knows, and perhaps nobody will ever know. Still, the hope to better understand the dimensions of centralized systems persists. Perhaps the best way to acknowledge that 1) discarding the general category of emotion would be unsafe, and 2) researchers hope to establish a "science" of the mind, would be to posit the general category of emotion as a regulative ideal—the general category preserves this hope as an ideal; and it regulates practices in the search for various consilient relations between studies.

More broadly, from a "metaphysical" perspective, the link between evolution and emotion makes sense because evolution concerns an indefinite science of "life"; while the general category of emotion is heterogeneous, it is less so than the general category of life, whence a possible science of emotion would apparently be included within an "indefinitely evolving" science of life. In other words, the general (all-encompassing) category of "life" is likewise a regulative ideal, since various studies seek to understand its dimensions, even though nobody fully knows what exactly "it" is, or if it even exists. And finally, at this metaphysical level, as
both evolution and complexity are framed by the general category of disposition, exploring emotion from an extended evolutionary perspective—specifically, the various extended, tangled relations between affect research programs and the human sciences—utilizing complex-systems thinking also makes eminent sense. For evolution is not just one long argument, it is also crucially an indefinitely evolving argument.
Glossary


ARE: Adaptively Relevant Environment. An alternative, as provided by behavioral ecologists, to the EEA. This alternative holds that (behaviorally) adaptive “pieces” of our minds stabilized over time, which makes it appear as if our cognitive architecture has been the same since the Pleistocene, but more probably mosaic changes occurred within the Pleistocene and after that as well.

Affect Program: An affect program is a coordinated set of changes constituting the emotional response; the changes are complex, coordinated, and automatic. They are complex since the elements of an affect program include (a) expressive facial changes, (b) musculoskeletal responses such as flinching and orienting, (c) expressive vocal changes, (d) endocrine system changes and consequent changes in the level of hormones, and (e) autonomic nervous system changes; they are coordinated because the above elements occur together in regular, patterned ways that are distinguishable (e.g., the pattern of fear is distinguishable from the pattern of anger); and they are automatic because they unfold without conscious direction.

Algorithmic Level of Analysis: The details of each “box” delineated by the computational level are cashed-out at this “architectural” level—i.e., the “architectural instructions” are given for how to build each box, stopping short of actually building the boxes.

Appraisals: Modes of activation within a dimensional (activation) space.

Capacities: Capacities are about causal structures; they are the “ontological ground” for causal
laws. In other words, capacities are required presuppositions for revealing, under appropriate conditions, certain causal abilities of models. See also tendencies and dispositions.


Causal-Phenomenological Axis: Causal generalizations are captured by a homogeneous partitioning of the statistically relevant factors for an attribute (the ideal being maximal resiliency). Phenomenological generalizations, by contrast, track correlations between events (e.g., the correlation between weather and a barometer). The difference between causal and phenomenological generalizations is that one can have the latter without the former—the familiar claim that correlation is not necessarily cause.

Causal Homeostasis (see Projectible): A robust version of projectibility (I often use “projectibility” as shorthand for “causal homeostasis”). A category/concept exhibits causal homeostasis when it provides successful inductive inferences to novel instances of the category. More fully, causal homeostasis looks for scientifically projectible inductions relative to an established scientific context, whereupon the causal aspect of causal homeostasis is the postulation of an underlying system of causes explaining why such inductions are projectible in the first place. The purported referents of causally homeostatic categories/concepts are natural kinds.

Central Systems (see also Peripheral Systems): Central systems are global in that they are not localized, and may in principle access any part of the “web of information” coming from peripheral or other central processes. This “in-principle” global access is what distinguishes peripheral systems from central systems.

Complex-Systems: Understanding a system as complex utilizes two conceptual moments.
The two requisite categories for this meta-level approach are: 1) a ("static") state space representation, and 2) a dynamical representation. Each category unfolds the theme of "synthetic microanalysis." which in brief is the bundling of micro-relations required to form constituents that, when situated in a system, have macro-relations to each other and the system. (The rainbow example is used to illustrate these notions in section 8.4).

Computational Level of Analysis: From this "middle-level" perspective, a system is understood via a "flowchart" of boxes (each indicating what is processed and why), where the system's resulting operation is "functionally defined" by these "boxologized" constraints. While this is called a "computational" level, it is more appropriately viewed as the level from which the overall "architectural role" of the system is grasped (e.g. the overall "flowchart" of how a car generally works, without any of the details of how to build it, nor the actual building of it). This level is also one aspect of the ecological level.

Concrete (Pole): See Concrete-Abstract Axis.

Concrete-Abstract Axis: The concrete dimension focuses on the actual domain—i.e., what is empirically found—of a scientific generalization. "Active invariant counterfactuals" cash out how good a generalization is, causally speaking, relative to the actual domain. On the other hand, the "less active invariant counterfactuals"—"less active" since they aren't really doing any "ontological work," although they are also viewed as expressing empirical regularities—are the phenomenological generalizations. Both causal and phenomenological generalizations are concrete in that they express invariant counterfactuals with respect to some actual domain. By contrast, the abstract dimension focuses on dispositions (in particular, capacities and tendencies). Dispositions express nomically strong generalizations with vague ceteris paribus clauses, where ceteris paribus clauses function to make adjustments between the actual domain of the
generalization (what is empirically found) and the intended domain (the hoped for range of the generalization).

Consilience: A field of investigation (theory, model, etc.) exhibiting a growing, coherent web of independent, yet convergent, lines of investigation is said to be "consilient."

Constituent Description: A description using "specialized predicates to absorb most relations and characterize individual constituents within the many-body system more concisely" (Auyang 1998, p.176).

DST: Developmental Systems Theory. DST includes the particular research program of EvoDevo as well as the more general philosophical framework covered in Chapter 5.

Dimensional Approaches: Approaches employing "Cartesian coordinate" representations—i.e., representations with axes and other parameters for classifying emotive phenomena (e.g., a two-dimensional framework with a positive-negative valence axis, and an intensity axis).

Dispositions: "Dispositions" is used to cover talk of (abstract) capacities and tendencies. Dispositions are conceptualized as referring to nomically strong generalizations with vague ceteris paribus clauses, where ceteris paribus clauses function to make adjustments between the actual domain of generalization and the intended domain (see section 2.2, especially fn.20).

Domain Specificity (or Domain-specific Inputs): These inputs concern the constrained range of information "coming into" the processing system. ("Parameterizing" domain-specificity results in the broader distinction between access specificity/generality mentioned in 4.1, which generates certain possible FIIPS.)

DIT: Dual Inheritance Theory. Extensions of the received view of evolution to the cultural realm; the dual systems concern (population) genetic systems, and (population) cultural systems.
Ecological Level: The computational level plus the intentional stance. The computational level outlines the general architectural roles of systems; the intentional stance further constrains the application of the computational level of analysis. Basically, what is called the “Ecological Level” in Chapter 2 and Chapter 7 is what I call the “functional stance.”

Elemental Description: A description of a system’s “constituents in terms of the character types and relation types familiar in small or minimal systems” (Auyang 1998, p.176).

Emergence (or Emergent Characters): “Characters” (one-place predicates delineating the main constituents of a state) which are non-resultant. Emergence involves the comparison between a microdescription and a macrodescription. See section 9.2.

Encapsulation: Encapsulation holds that the internal operations in a processing system cannot draw information from outside. (“Parameterizing” encapsulation results in the broader distinction between process specificity/generality mentioned in 4.1, which generates certain possible FIIPS.) In slightly different terms, a system is encapsulated to the extent that it isn’t a centralized system. (Note that encapsulation is different from the concept of peripheral systems, since encapsulation is one potential feature of such systems.)

EEA: Environment of Evolutionary Adaptedness. The “crucible” in which our species evolved—in particular, the environments in the Pleistocene that gave rise to Homo sapiens (sapiens), and in which (according to the Santa Barbara School) we acquired our species-typical cognitive adaptations.

EvoDevo (see DST): Evolutionary Developmental Biology. See section 7.4.

Fidelity Axis (low-high): This axis recognizes that theoretical models come in many types and operate in different ways depending on the variegated demands of practicing scientists. What this third axis acknowledges is that some representations “lie” more than others, in that the initial conditions and assumptions of some representations are more artificial than
others—i.e., fidelity tracks the capacity in which models intend to explicate general theoretical ideas and intuitions.

Four-Dimensional Conceptual Space: The four axes situating heterogeneous sciences discussed in Chapter 2, and applied in subsequent chapters. The four axes are the causal-phenomenological axis, the concrete-abstract axis, the fidelity axis, and the paradigm/research program/philosophical framework axis.

Functional Stance: The 4-D subspace capturing the functional approaches to emotions. This stance gains traction by employing reverse engineering, in addition to Dennett’s intentional stance.

FIIPS: Functionally Individuated Information Processing System(s). Fodorian modules are the narrowest kinds of FIIPS. More generally, the three parameters configuring possible FIIPS are domain specificity, encapsulation, and inaccessibility. Tweaking these three parameters yields putative, “widened” FIIPS.

Implementational Level of Analysis: The “lowest” level of reverse engineering, where actual implementation occurs—the schematics given by the computational level and the algorithmic level are physically instantiated at this level.

Intentional Stance: An as-if predictive strategy to understand “intentional systems” that behave in “life-like” ways. This stance is used cautiously; the intentional stance is used only when it is judged to yield predictive power that is hard to come by either via reverse engineering, or via other types of explanatory strategies (e.g., cause-effect accounts). First you treat the object whose behavior is to be predicted as a rational agent—that is, the agent has more-or-less consistent beliefs relative to dealing with aspects of the world, and is rational enough to draw certain implications from such beliefs. Then you figure out what beliefs that agent ought to have, given its place in the world and its purpose—
that is, figure out what “true” (i.e., warrantedly assertible) beliefs the agent ought to have to cope in its environment. Figuring out its beliefs is intertwined with figuring out the desires it ought to have (desires satisfying the most basic needs like survival, food, etc.). And finally you assume that this rational agent will act to further its goals in light of its beliefs. The intentional stance and the computational level operate at similar levels of analysis in that both concern the ecological roles of systems; the intentional stance, though, offers a more explicit understanding of these roles for intentional systems.

Macrodescription: A macrodescription “introduces character types with simpler values and seals the unity of the composite system” (Auyang 1998, p.176).

Microdescription: A microdescription “of a composite system is given in terms of character types, each value of which specifies the behaviors and relations of every constituent” (Auyang 1998, p.176).

Modules (see also FIIPS): Modules take certain kinds of information as inputs, then compile and transduce the inputs into “standardized” cognitive outputs. Briefly, they are mandatory, opaque (inaccessible), and encapsulated (see section 3.1).

Monomorphic Trait (see Polymorphic Trait): Monomorphic traits are exhibited by (nearly) all (conspecific) individuals within a “normal” environment. Monomorphic traits are traits which are not polymorphic.

Monomorphic View of Mind: The view that our species-typical cognitive “traits” are monomorphic “traits.” This thesis argues that apparent cognitive “adaptations” acquired by people across cultures are not a result of genetic differences, but rather a result of differences in environments. These apparent “adaptations” are really expressions of underlying, species-typical adaptations. Monomorphic psychic “traits”
concern the same set of genetically "determined" developmental potentials—the same "developmental program." (See Chapter 4, fn.13.)

Paradigm: "Paradigm" includes exemplars—which are recognized achievements providing model problems and solutions for a community of practitioners—and disciplinary matrix—which includes the shared global commitments of a community of practitioners embedded in a "form of life" (a form including ordered elements such as symbolic generalizations, models, values, heuristics, etc.).

**Paradigm/Research Program/Philosophical Framework Axis:** Philosophical frameworks are orientations or stances that bring into relief the different kinds of integrity holding between methods, theories, and aims. At one end of the spectrum where integrity is high, a paradigm offers a worldview that yields a well-substantiated philosophical framework. With a weaker degree of integrity, research programs disclose a vaguer picture (or pictures) if only because the various consequences of the methods, theories, and aims are not as well worked out in comparison to paradigms, perhaps due to the heterogeneity of the subject matters being investigated. Lastly, when integrity is weakest, philosophical frameworks might offer mere worldviews without any mention of methods and theories, ideals that tap into some existing senses of methods and theories, or perhaps a new concept that will help to reorganize existing methods/theories or to create new methods/theories. For example, when integrity is called into question concerning whether certain social sciences are really sciences, philosophical frameworks open a space for genuine dispute over differing conceptions of "science"—philosophical frameworks assist in disclosing where the very category of "science" is problematized.

Peripheral Systems (see also Central Systems): Peripheral systems are usually modular (e.g.,
the visual system), and are peripheral in that they operate on narrowly constrained inputs and produce narrowly constrained outputs, which may in turn be passed to centralized systems. Furthermore, they are "peripheral" as a contrast class for central systems.


Philosophical Frameworks: Philosophical frameworks are orientations or stances that bring into relief the different kinds of integrity holding between methods, theories, and aims. Importantly, when integrity is called into question concerning whether, for example, certain social sciences are really sciences, philosophical frameworks open a space for genuine dispute over differing conceptions of "science"—philosophical frameworks assist in disclosing where the very category of "science" is problematized.

Polymorphic Trait: Polymorphic traits maintain significant ("discrete") variation of the relevant types within a population (for example, eye color—green, brown, etc.). More generally, "discontinuous"/discrete genetic variation resulting in discrete phenotypes within a population—and the maintenance of those types within that population—characterizes polymorphic traits.

Projectible (see Causal Homeostasis): Projectible predicates take (past) correlated properties that are representative of what such predicates are applied to; what makes them projectible is that they hold up in new instances—they are "good" inductions in that they reliably work in these new instances.

Proximate Causation (see also Ultimate Causation): Proximate investigations, usually subsumed under the field of "Functional Biology," search for answers to "how X operates."

Research Program: Research programs embody a pluralistic approach to understanding
"scientific" practices. This is a pluralism where theories progress by a process in which 1) theory $T^*$ has excess empirical content over $T$ (i.e., $T^*$ can predict novel facts not easily accessible by $T$), 2) $T^*$ captures and explains $T$, and 3) some of the new excess is corroborated. Refutation of a theory is retrospective, since there is no crucial experiment that rejects a theory except when judged by hindsight after a better theory has been put into place. As an analogue to Kuhn's normal science, Lakatos proposes a methodology of research programs that appraises a succession of theories as scientific or not through "negative" and "positive heuristics" (i.e., those paths not to pursue and those to pursue). The metaphor used is that of a hard core and a protective belt. Negative heuristics tell us not to upset the core tenets of a theory within a program, whereas the protective belt can be toyed with using auxiliary hypotheses, should apparently anomalous results make their appearance. The positive heuristics (within a program) suggest to us how to refine the protective belt by providing more apt models.

Resiliency: Given a sentence $q$, a set of sentences $S=\{p_i\}$, and Pr($q$)=n, the resiliency function is defined as $\text{Res}(\text{Pr}(q)=n) = 1 - \max_{p_i} \text{ln-Pr}(q/p_i)$ for $p_i$ ranging over $S$. The best-case scenario is when $q$ is independent from $S$; the more that $q$ depends on elements of $S$, the lower the resiliency. The degree of resiliency then determines the degree of "nomic generality." Note that resiliency takes the ("widened") place of what (narrower) work projectibility does for Griffiths and Fodor.

Resultant Characters (see also Emergence): A character 1) whose properties are qualitatively similar to its constituents, belonging to the same character type as the constituent properties; and 2) whose microexplanation can be given by approximately analyzing the system into independent parts with distinctive characters such that it is the sum or
average of the characters of the parts, where the micro-analysis includes independent-individual models, the superposition principle, and other available means.

Reverse Engineering (see also Functional Stance): The predominant method for understanding systems at three interrelated levels of analysis. First, the computational level of analysis adopts a middle-level orientation to understand the overall architecture of a system’s function(s). The system then gets broken down to see how its parts contribute to these function(s); breaking down the system’s parts invokes the algorithmic and implementational levels of analysis. These three levels jointly delineate reverse engineering.

Semiotic Stance: This stance expresses the threefold distinction between sign, signified, and "interpretant." Signs are things that stand to some agent for something in some capacity; signification concerns what the signs stand for; and interpretants relate objects to signs and give those relations significance—it specifies how a sign is to be taken, relative to what it is a sign of, in its capacity as a sign. This stance is presupposed by the functional stance and the strategic stance.

Strategic Stance: This stance builds upon the functional stance covered in section 4.2; it looks at relations between "boxes"/agents to delineate patterns of interaction that the functional stance might not so readily discern. The way in which the strategic stance reveals patterns is through two uses of "strategy"—a static use and a dynamic use, both of which find technical application in game theory. In static game theory, the search for equilibrium points occurs in non-repeated games where the space of possible strategies depends upon the information available to the agents ("perfect" or "imperfect" information); the upshot is that in static games, strategies look at possible dispositions (expected value payoffs) without considering repeated encounters with agents. In short,
the picture is a "time-slice," where the aim is to find and classify solutions relative to the information available. For dynamic games, the intentional stance accommodates this dynamical shift through a change of perspective, where strategies are viewed as intentional systems themselves. That is, the units of action in static game theory are agents, and now the units are strategies (systems, in turn, are sets of strategies); give the strategies intentional system status, then look at the ways in which strategies ("second-order agents") interact with other strategies.

Syndromes: Covert emotion systems are "syndromes" since they are interpreted as "passions" (in the sense of being disclaimed and passively received). What makes covert social constructs powerful and strange (from a meta-level view) is that individuals are appraising an emotional syndrome as a passion, and thus are making "active" judgments using individual scripts, yet for the practice to make a difference for that form of life it must be tacitly understood as disclaimed.

System: An individual whose constituents are situated in relation to each other and to that system as a whole.

Tendencies: Tendencies are about how phenomena behave; they are the "ontological ground" for (phenomenological) laws of association. In other words, tendencies are required presuppositions for revealing, under appropriate conditions, certain phenomenological abilities of correlational/associative models. See also capacities and dispositions.

Ultimate Causation (see also Proximate Causation): The search for ultimate causations seeks answers to "why X evolved."

Universality: The two features of universality are: 1) the details of the system (those details that
would feature in a complete causal-mechanical explanation of the system's behavior) are largely irrelevant for describing the behavior of interest; and 2) many different systems with completely different "micro" details exhibit the same behavior.
Bibliography


Barkow, Jerome, Leda Cosmides, and John Tooby (eds.). 1992. The Adapted Mind:


----- . 1999. "The Units of Selection Revisited: The Modules of Selection." Biology and


Cunningham, Donald and Gary Shank. "Semiotics, An Introduction."

http://www.indiana.edu/~educp550/shtcrs.html


Gilbert, Scott, John Opitz, and Rudolf Raff. 1996. “Resynthesizing Evolutionary and


of Evolutionary Psychology edited by David Buss. New Jersey: John Wiley and Sons, Inc.


James, William. 1884. "What is an Emotion?" Mind 19:188-204.


Mills, Susan and John Beatty. 1979. "The Propensity Interpretation of Fitness." Philosophy of
Science 46:263-86.
Ross, Don and Paul Dumouchel. 2004. "Emotions As Strategic Signals." Rationality and


