COGNITION IN PRACTICE: CONCEPTUAL DEVELOPMENT AND DISAGREEMENT IN COGNITIVE SCIENCE

by

Mikio Shaun Mikuriya Akagi

B.A. with High Honors in Philosophy, Swarthmore College, 2008

MSc with Distinction in Philosophy, University of Edinburgh, 2009

Submitted to the Graduate Faculty of

the Kenneth P. Dietrich School of Arts and Sciences

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH

KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Mikio Shaun Mikuriya Akagi

It was defended on

March 28, 2016

and approved by

Edouard Machery, Professor of History and Philosophy of Science

Mark L. Wilson, Distinguished Professor of Philosophy

Mazviita Chirimuuta, Assistant Professor of History and Philosophy of Science

James F. Woodward, Distinguished Professor of History and Philosophy of Science

Dissertation Director: Robert B. Brandom, Distinguished Professor of Philosophy

COGNITION IN PRACTICE: CONCEPTUAL DEVELOPMENT AND DISAGREEMENT IN COGNITIVE SCIENCE

Mikio Akagi, PhD

University of Pittsburgh, 2016

Cognitive science has been beset for thirty years by foundational disputes about the nature and extension of cognition—e.g. whether cognition is necessarily representational, whether cognitive processes extend outside the brain or body, and whether plants or microbes have them. Whereas previous philosophical work aimed to settle these disputes, I aim to understand what conception of cognition scientists could share given that they disagree so fundamentally. To this end, I develop a number of variations on traditional conceptual explication, and defend a novel explication of cognition called the sensitive management hypothesis.

Since expert judgments about the extension of "cognition" vary so much, I argue that there is value in explication that accurately models the variance in judgments rather than taking sides or treating that variance as noise. I say of explications that accomplish this that they are *ecumenically extensionally adequate*. Thus, rather than adjudicating whether, say, plants can have cognitive processes like humans, an ecumenically adequate explication should classify these cases differently: human cognitive processes as paradigmatically cognitive, and plant processes as controversially cognitive.

I achieve ecumenical adequacy by articulating conceptual explications with *parameters*, or terms that can be assigned a number of distinct interpretations based on the background

commitments of participants in a discourse. For example, an explication might require that cognition cause "behavior," and imply that plant processes are cognitive or not depending on whether anything plants do can be considered "behavior." Parameterization provides a unified treatment of embattled concepts by isolating topics of disagreement in a small number of parameters.

I incorporate these innovations into an account on which cognition is the "sensitive management of organismal behavior." The sensitive management hypothesis is ecumenically extensionally adequate, accurately classifying a broad variety of cases as paradigmatically or controversially cognitive phenomena. I also describe an extremely permissive version of the sensitive management hypothesis, arguing that it has the potential to explain several features of cognitive scientific discourse, including various facts about the way cognitive scientists ascribe representations to cognitive systems.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EXPLICATIONS AND SCHEMATA

PREFACE

It is said that it takes a village to raise a child. I know myself that it was through the help of a great variety of people that I came from a cornfield in Ohio to the Cathedral of Learning.

I began thinking about the nature of cognition at Swarthmore College, where I frequently crossed the hall between the philosophy seminar room and Frank Durgin's visual perception laboratory. In 2008 I brought a headful of concerns to Edinburgh in order to take them up with Andy Clark and Pepa Toribio, hoping to put them to rest. When I arrived in Pittsburgh in 2009, I planned to focus my attention on other topics, but found that my thoughts returned to cognitive science. For my dissertation I undertook a project about science and functionalism, but succumbed once again to my fascination with cognition. The result, such as it is, is this document.

I was very fortunate to have the mentors I did. Robert Brandom, whose work I have admired since my college days, was unfailingly patient and supportive even as my project changed course. I am indebted to Edouard Machery for his sharp criticisms and sage advice. Mark Sprevak offered me much of his time and his characteristically measured judgment whenever I was in Edinburgh, and often when I was not. Mazviita Chirimuuta's support meant a lot to me, as did her willingness to join my committee late in the process. And I received a great deal of attention and criticism from Mark Wilson and Jim Woodward, whose sincere support has done much to shape my time in Pittsburgh. I admire each of these people tremendously; they have each left an indelible stamp on my thinking, which while perhaps not always apparent to them is unmistakable to me.

I must thank the University of Pittsburgh Department of Philosophy and the Office of the Provost for funding my work over the last seven years. I also received generous support from the

Wesley C. Salmon Fund, Robert Brandom, and the Andrew W. Mellon Foundation, which enabled me to spend an academic year in Edinburgh that was tremendously beneficial to me both intellectually and personally.

I owe much to my colleagues in Pittsburgh and elsewhere. I am particularly grateful to members of the Department of History and Philosophy of Science at Pitt and the Department of Philosophy at the University of Edinburgh for welcoming me so warmly when I wandered into their midst, as I often did. I benefited from valuable discussions with many people: Trey Boone, Julia Bursten, Zina Ward, Haixin Dang, Bihui Li, David Colaço, Josh Eisenthal, Kathryn Lindeman, Robert Steel, Josh Hancox, Jon Buttaci, Adam Marushak, Eric Hochstein, Emlen Metz, Lily Prudhomme, Evan Butts, Mo Salam, Stephen Ryan, Jamie Collin, Caitlin Tenison, Lucy Erickson, Amber North, Alistair Isaac, Andy Clark, Gualtiero Piccinini, Pepa Toribio, Carrie Figdor, Rob Rupert, Alex Doumas, Bill Bechtel, Zoe Drayson, and many others.

I must also thank my parents, my brother, my sister, and my stepfather for their many years of support. I don't think any of them were surprised when I finally decided that I wanted to be a philosopher when I grew up. I also owe much to the community at Lake Ridge Academy, and to my first philosophy teachers at Swarthmore College—Richard Schuldenfrei, Grace Ledbetter, Alan Baker, Hans Oberdiek, and Richard Eldridge—for starting me down the dark path. And special thanks are due to Eimear O'Carroll, my Pirate Queen, for her love and patience, and for helping me to remember the value in things besides philosophy.

Finally, I must acknowledge a very special debt of gratitude to Joe McCaffrey, whose conversation and criticism has made a tremendous difference to this document, and whose friendship has made a tremendous difference to me. I don't know what I'd have done without you, buddy.

1.0 INTRODUCTION

1.1 A PRACTICAL TURN IN PHILOSOPHY OF SCIENCE

We have minds (Figure 1.1 A). Cognitive scientists are people who study our minds and the minds of other critters, in order to determine how they work—how they enable us to perceive, learn, remember, decide, and behave (Figure 1.1 B). Philosophers of cognitive science (or rather, some of them) study cognitive scientists, in order to determine how they work—how they observe, hypothesize, explain, and revise their hypotheses (Figure 1.1 C). Some of the work I do here is of this latter sort. Some of it is pitched a level higher: I am also concerned with these philosophers of cognitive science and how they work (Figure 1.1 D).

I say only some philosophers of cognitive science, because this kind of project stands in contrast to other things that philosophy of the sciences can be and has been. Many philosophers of science contribute to science in a direct fashion, developing new hypotheses or articulating them clearly. In the context of cognitive science, one thinks immediately of Jerry Fodor's discussion of modularity (1985) or Ned Block's distinction between P-consciousness and A-consciousness (1995). Other philosophers of cognitive science aim to defend particular theoretical commitments (e.g. Chemero 2009, defending anti-representational, embodied cognitive science), to criticize currently accepted claims or practices (e.g. Machery 2009 on confusions regarding psychological accounts of concepts), or to articulate imaginative hypotheses (e.g. Clark 2008 on the notion of extended cognition). All of these projects have fairly straightforward prescriptive consequences for non-philosopher cognitive scientists: they provide new claims to consider and test, or they provide support for some existing projects rather than others, or they provide reasons for changing current

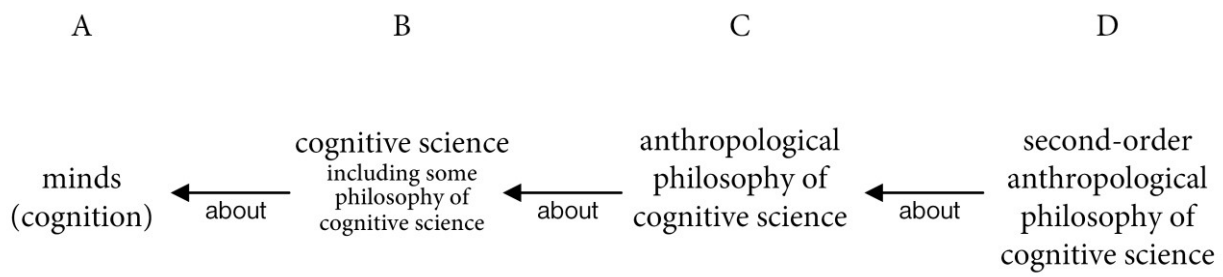

Figure 1.1: Philosophies of cognitive science and their objects of inquiry. This dissertation is primarily a contribution to the literature on (C) anthropological philosophy of cognitive science, which is about the practice and theory of (B) cognitive science, rather than (A) cognition itself. Parts of this dissertation concern (D) second-order philosophy of cognitive science, whose object of inquiry is (C) anthropological philosophy of cognitive science.

practices. I would categorize such concerns as a part of cognitive science proper, which, as an interdisciplinary enterprise, has always incorporated philosophical contributions (Figure 1.1 B).

The present work does not have such straightforward consequences for cognitive scientists. It is directed primarily at philosophers (though I hope parts of it are of broader interest). My main objective is to characterize an element of cognitive scientific practice as it is now, namely the way the concept COGNITION is used.[1] Accomplishing this objective requires motivating a kind of project that may seem unusual to those only familiar with the sort of philosophy of cognitive science described above, and it requires the development of some new logical resources. However, while my present aim is not to criticize first-order scientific practice, my conclusions do have revisionary consequences. First, I view them as a prolegomenon to sorting out conceptual and theoretical disputes in cognitive science more gracefully than we have for the last thirty years. While I will not be critical of current scientific practice in this work, I do not think the practice beyond criticism. However, understanding a state of entrenched disagreement is often helpful for resolving

[1] Throughout, I employ the convention of small capitalization to indicate reference to concepts. "Cognition" is a word, COGNITION is a concept, and cognition is a natural phenomenon.

such disagreements, and when disagreements are as complex as the present disagreements about the concept of COGNITION, some complex analytical machinery may be necessary. Second, my conclusions have revisionary consequences for common views and practices among philosophers of cognitive science.

Nevertheless, this project is not entirely idiosyncratic. I take it to be an example of a kind of anthropological or "practice-oriented" work that is currently gaining prominence in philosophy of science. The new mechanists and their critics, for example, try to account for the variety of forms that explanation takes in the special sciences (Machamer, Darden, and Craver 2000; Bechtel 2007; Craver 2007). They may take their claims to have revisionary consequences for explanatory practices in science, but in order to be helpful rather than intellectual busybodies (q.v. § 1.2) they must first understand the details and rationale of those practices. There is also recent work that describes the construction and use of scientific models (e.g. Weisberg 2007), the uses of scientific abstraction (e.g. Levy and Bechtel 2013), and the commensuration of experimental protocols that vary between laboratories (Sullivan 2009). This practice-oriented approach has begun to influence work on conceptual analysis and the adjudication of scientific disputes, for example about the concept of CONCEPT in psychology (Machery 2009). I extend this practical turn to the long-standing disputes in cognitive science over the nature of cognition.

I will take the opportunity afforded by introductions to clarify my practice-oriented methodology (I have found in the past that it saves trouble later on). I will begin by setting the stage with some historical remarks about the notion of COGNITION or THOUGHT in psychology and cognitive science (§ 1.2). With that background in place, I will say a bit about anthropological methods in philosophy (§ 1.3). Finally, I will clarify how I understand the term *concept* (§ 1.4) before outlining the remainder of the dissertation (§ 1.5). By necessity, these clarifications are more abstract than the material that follows. By courtesy, they are somewhat condensed. They will prove unnecessary for some readers, who might profitably skip to Chapter 2.

1.2 SOME POTTED HISTORY OF COGNITION

There is a whiggish history of psychology, the scent of which one often finds in the introductory chapters of undergraduate psychology textbooks. That history goes something like this: In the beginning (1879), academic psychology was a science of thought and sensation. The antediluvian psychologists were called introspectionists. They had a flawed method for discerning the ontology and dynamics of mental states, but they were guileless babes exploring a new and beautiful science. Their innocence came to an end in the 1920s, though, when psychology fell under the shadow of a tyrannical theoretical framework called behaviorism. The behaviorists said the introspectionists were unrigorous and unscientific, and harsh new methodological strictures were instituted. Under the behaviorist regime, the positing of "thoughts," "representations," and other intermediate states was outlawed. Instead, psychologists were forced to account for all psychological phenomena with an austere ontology of stimuli, responses, and associations between them. This went on for decades. But in 1959, a brave young linguist came and declared that capacities like language could not be governed by behaviorist laws ("R-expressions must be free!"). Psychologists and philosophers rallied to his side, and the cognitive revolution ensued. By the 1970s, the behaviorists were routed and psychologists were free once again to posit representations and other intermediate states. And now, having learned about rigor from their erstwhile behaviorist overlords, these new "cognitive" psychologists were much more scientifically productive than their introspectionist forebears. Some psychologists deepened their ties to their allies in linguistics, philosophy, and (a relative newcomer) computer science, as well as with neurologists and anthropologists. That alliance is called cognitive science.

Of course, this story is at best a strained representation of the history, and most textbooks are not this simplistic. Nevertheless, you can often still get a whiff of this narrative in most survey textbooks that open with a chapter on history (e.g. Myers 2004, just to take an example from my bookshelf) and many authors still regard representational commitments as the sine qua non of cognitive science, without which we risk reliving the unnecessary austerity of behaviorism (e.g.

Ramsey 2007; Buckner 2015). The central role of representation in scientific conceptions of cognition is one strand in this history. We can discern another if for the moment we understand the term "cognition" not as designating a determinate kind of phenomenon, but as functioning to demarcate the proper object of inquiry for the behavioral sciences as judged by scientists at a time (which is a use to which COGNITION is often put now; see § 2.2). If we understand the concept COGNITION this way, the three-act structure of the story traces a narrowing and then widening of scientific conceptions of cognition. The introspectionists took sensations, thoughts, and feelings to be their object of inquiry. The behaviorists narrowed their object of inquiry to the boundaries of these phenomena: externally observable stimulations and behavior. The cognitivists incorporate both of these conceptions, observing behavior and positing representational structures to explain their observations. But the widening trend in the story is made more apparent with a more careful recounting of some more recent history.

1.2.1 Cognition and computational functionalism

The early days of cognitive science coincide with the 20th century rise in popularity of computational functionalism, usually called simply "functionalism" (Putnam 1967a, 1967b; Fodor 1968, 1975; Newell and Simon 1976), and functionalism has since then been closely associated with cognitive science. There are many versions of functionalism, but roughly speaking functionalism is the claim that mental or cognitive states are functional states, which is to say states whose criteria of ascription and individuation are given in terms of their functional relations rather than by reference to their material components. Functional relations are characteristic causes and effects, or relations to inputs (i.e. sensations, perceptions), outputs (i.e. behavior, actions) and other functional states (see Appendix § A.1).

At least since the 1980s, functionalism has been a common stalking horse for criticism of the theoretical assumptions behind cognitive science (Searle 1980; Buechner 2011; Eliasmith 2002). However, the locus classicus for early functionalism is found in Hilary Putnam (1967b), where

functionalism is defended as a metaphysics not for cognitive science but for folk-psychological mental states such as pain, belief, and preference. Putnam's argument is essentially that functionalism is more extensionally adequate than a neural type-identity theory (cf. Place 1956; Smart 1959) because neural type-identities are subject to chauvinism objections (q.v. § 3.2) to which functionalism is immune. For example, we generally think of pain as a state that is experienced by both mammals and mollusks, although they have very different nervous systems and pain-states are realized by different neural structures in mammals and mollusks.

As cognitive science matured through the 1970s and '80s (or perhaps we might say, as it entered a rebellious adolescence) the ontology of cognitive kinds came to diverge dramatically from that of folk-psychological mental kinds like beliefs, desires, and intentions. A well-known example is the fracturing of the folk concept of MEMORY—psychologists now routinely make a host of distinctions between kinds of memory, e.g. between sensory, short-term, and long-term memory (Atkinson and Shiffrin 1968, 1971) or between episodic and semantic memory (Tulving 1972). These counterintuitive findings were so troubling that in the early 1980s several inquirers wondered whether folk psychological kinds would appear at all in our best psychological theories (Churchland 1981; Stich 1983). Less radically, serious worries were raised about early versions of the classical computational theory of cognition (CCTC) (cf. Turing 1950; Fodor 1975; Newell and Simon 1976; I take that particular label from Ramsey 2007). According to CCTC, cognition consists in computation over semantically-rich representations whose contents and interactions resemble those of conscious thought. The classical computational theory of cognition is famously exemplified by research projects in AI such as Roger Schank's restaurant script (Schank 1975) and the SOAR architecture (Laird, Newell, and Rosenbloom 1987).

The worries of the early 1980s were amplified by the success of connectionist models of cognitive processes (e.g. Rumelhart 1989), which failed to resemble folk-psychological and conscious processes in even more dramatic ways. For example, connectionist models do not trade in propositionally-structured representations, they invoke representational units that are distributed across parallel processing elements and which can overlap, and they have difficulty

replicating various features of natural language such as compositionality and recursiveness (Fodor and Pylyshyn 1988).

So Putnam's classic articulation of functionalism was meant for mental states, but the early development of cognitive science sees a widening gulf between folk-psychological mental kinds and scientifically useful cognitive kinds. As a result, we now commonly recognize a distinction between two kinds of functionalism: an analytic functionalism suited for philosophical psychology and metaphysics (Lewis 1972; Jackson 1998), and a psycho-functionalism suited for empirical psychology and cognitive science (Fodor 1968; Chalmers 2011). The major casualty of this divide is the naïve version of the computational theory of mind on which our ontology of cognitive states and processes would map cleanly onto our ontology of mental states and processes (though less naïve versions of CCTC still inform some cognitive science research). And this conceptual cleavage between cognitive theory and folk psychology soon grew more dramatic.

1.2.2 The cognition border wars, briefly

Since the connectionist challenge to classical cognitive science in the 1980s, cognitive scientists have advocated for a startling variety of (what I will call) "anti-classical" perspectives. Some claim that cognitive processes are realized by non-neural components in the body ("embodied cognition"), or by items and relationships outside of the body ("extended cognition"). Some deny that cognitive processes are realized by items outside of the body, but argue that cognitive processes crucially depend on extra-bodily items and interactions in ways that early cognitive science did not appreciate ("situated" or "embedded" cognition). Some deny that representations are crucial for cognition (anti-representationalism), or deny that cognition is literally a form of computation, or deny that the computation is illuminating even as a metaphor for cognition. Some claim that cognition can only be modeled as a dynamical process, drawing on the success of dynamical models in physics. Some argue that cognitive processes are to be found in plants, or in microbes, or in robots. Still others deny all these anti-classical claims. Most deny some but not all of them. I call

these disputes, considered together, the *cognition border wars*. I will provide an overview of the most prominent anti-classical perspectives in Chapter 2 (§ 2.2), though my focus throughout this document will be on the disputes concerning the extension of cognition rather than disputes about what methods should be used to study it. In the extant literature of the border wars, these topics are often considered together. In Chapter 5 I will describe several purported examples of anti-classical cognition (§ 5.3). The Appendix includes a discussion of some philosophical disagreements over whether there is extended cognition (§ A.2).

1.2.2.1 Common conceptions of cognition The credit for encouraging an explicit discussion of the nature of cognition belongs largely to Fred Adams and Ken Aizawa (2001). Before 2001, there were a number of suggestions regarding the nature of cognition but all were either quite vague or obviously liberal. In the early days of cognitive science Jerry Fodor (1975) adapted Putnam-style computational functionalism for cognitive scientific contexts (see § 1.2.1), but Fodor's functionalism is not meant to be a criterion of demarcation for cognition; it is an articulation of some methodological assumptions of classical cognitive science. Allen Newell and Herbert Simon (1976) articulated the "physical symbol system hypothesis," essentially that any intelligent system is an implementation of a Turing machine. Both of these suggestions are expressions of the classical computational theory of cognition, and have been much discussed as hypotheses about how to understand and represent cognitive architecture. However, they do not provide clear guidance on how to determine the extension of COGNITION, since there are many computational processes that are not cognitive processes.

Since the connectionist challenge to classical computationalism, cognitive scientists have tended to make do with a number of other awkward slogans that have similar drawbacks. For example, one familiar slogan is that "cognition is (just) information-processing." Such a slogan has value expressing the cognitive scientist's anti-dualist commitment that there is nothing magical or unnatural about cognition, but this is an unsatisfying criterion. It is not clear, for example, what sense is to be given to "information." It is sometimes alleged cognitive science and neuroscience

inherit an unproblematic, naturalized, and admirably formal notion of information from the work of Claude Shannon (1948). It is certainly true that the formal apparatus developed around Shannon information has been of use to cognitive scientists, but the claim that "a process is cognitive just in case it processes (manipulates, changes) states of Shannon information" is hopelessly liberal (q.v. § 3.2). Such a criterion would count phenomena like radioactive decay as cognitive processes, a verdict countenanced by no one as far as I know, and certainly not by the scientists who approve of the information-processing slogan. A more charitable interpretation of the slogan is that it gives voice to the commitment that cognition is *a kind of* information-processing, but then the slogan's value as a criterion depends on the sense of "kind of." Some cognitive scientists argue that Shannon information is an inappropriate notion of information altogether for cognitive science (e.g. Gibson 1979; Chemero 2009). Others argue that appeals to Shannon information in cognitive science are substantially enriched by local assumptions (Rowlands 2010, 110–119; Shea 2013, 498) (see also § 6.3.2).

Another familiar slogan is that "cognition is what the brain does." This slogan is also liberal, and also serves mostly to express naturalist commitments, and is also liberal. Not all brain activity is cognitive activity. Neurons decay with age, metabolize sugars, sometimes rupture, sometimes become cancerous. These are things "the brain does," but are poor candidates for cognitive processes. They are things the brain does incidentally, rather than things it does in fulfilling its function. This slogan is arguably chauvinistic (q.v. § 3.2), as well. If there is such a thing as extended cognition or plant cognition, then there are cognitive processes that are not brain processes. Other slogans associate cognition with "adaptive behavior" (or "flexible behavior," or "flexibly adaptive behavior," &c.). However, it is not clear what criteria distinguish flexibility, adaptiveness, or even behavior. I have never heard slogans of this kind defended as precise criteria of demarcation for cognition, often because they are acknowledged to be imprecise.

Given this poor showing for cognitive demarcation criteria, it became a commonplace in some circles that it is impossible to articulate such a criterion. This view is still held by some (Chemero 2009, 212n8; Clark 2010b, 62).

1.2.2.2 The mark of the cognitive In the present century, however, philosophers have attempted to improve the situation by defending more careful criteria for ascribing the concept COGNITION, and there has been more explicit discussion about the nature of cognition in the last few years than there was in the 20th century. A particular proposal about these criteria is called a "mark of the cognitive" (a self-conscious nod to Brentano). There are four extant "marks" that have been described and defended as such. I will describe them briefly in order to illustrate some of their varieties and commonalities. I do not intend to argue against any of these views directly in this document, as their aims are rather distinct from mine (see § 2.4). Nevertheless, a brief summary of extant views will make the contrast with my own approach clearer.

Adams and Aizawa coined the expression "mark of the cognitive," and defend a view about some aspects of the "mark": cognition must be a particular kind of causal process (though we do not yet know precisely how to characterize this process), and this causal process must somehow bear "non-derived content." These are supposed to be necessary, but not sufficient, conditions of cognition. Mark Rowlands (2010, esp. 110ff) defends a set of jointly sufficient (but not necessary) conditions:

> A process *p* is a cognitive process if
>
> 1. *p* involves Shannon information-processing,
> 2. this information-processing has the proper function (à la Millikan 1984) of making information available to a subject *s* (or to subsequent processing operations) that was not previously available,
> 3. this information is made available through the production in *s* of a representational state (with non-derived content), and
> 4. *p* belongs to *s*.

Cameron Buckner (2015) defends a view on which cognition is a homeostatic property cluster (q.v. Boyd 1991, 1999), essentially a cluster concept used provisionally to identify a natural kind. Buckner offers eight properties for the cluster: cognitive processes tend to be

1. context-sensitive,
2. rapid (i.e. they do not rely crucially on chance or brute force methods),
3. categorical (i.e. they respond discontinuously to stimuli),
4. high-order (i.e. they track abstract relationships),
5. multi-modal (i.e. they integrate information from several sensory channels),
6. inhibitable,
7. capable of monotonic integration of cues,
8. capable of generating and monitoring expectations.

Cognitive processes need not exhibit all of these properties according to Buckner, but tend to exhibit many of them. Buckner anticipates that someday we will discover a form of representational system (perhaps like the one Adams and Aizawa describe?) that explains why these properties sometimes cluster together. Finally, Adams and Rebecca Garrison (2013) defend a view on which cognition is the production of behavior for reasons, where these reasons are representational states bearing non-derived content, and are the reasons of the behaving system (not those of a designer, or those imparted by natural selection; see § 6.2.1).

1.2.3 Conceptual disagreement

This is, anyway, a chaotic state of affairs for cognitive science. It is to this tangle of conflicting views that I intend to bring some order. However, unlike the extant contributions to the literature on the "mark of the cognitive," I intend to bring order in the form of clear and ecumenical description, rather than sectarian prescription (see §§ 2.4–2.5). That is, I do not intend to settle precisely what

the "mark of the cognitive" should be, nor to adjudicate which controversial cases of cognition are genuine cases. Rather, I intend to introduce a model—the sensitive management hypothesis—that expresses the norms for using the concept COGNITION, as well as which of the norms are in dispute among cognitive scientists and which may be considered common ground between them.

A contribution like this is possible because the disagreements at issue are what might be called *conceptual disagreements*. By contrast, a factual disagreement is a straightforward disagreement about what is the case: whether the Eiffel Tower is taller than the Tōkyō Tower, whether Bond is in Zürich, or whether any cephalopod axons are myelinated. Purely factual disagreements are simply about what is empirically the case, and can generally be resolved by empirical methods. A verbal disagreement, on the other hand, is a disagreement arising because disputants unknowingly use words in different ways. For example, if Susanne says that Mary was thrown out of the bar (a place of business serving intoxicating beverages) and Joe insists that Mary was never thrown out of the bar (a legal association), their disagreement is merely verbal. Verbal disagreements are often resolved when disputants come to recognize that their disagreement is verbal. Conceptual disagreement is a third form of disagreement, in which disputants endorse different norms for the use of a word or concept, but where recognition of this fact is not sufficient to resolve the disagreement because they maintain that the other party uses the concept incorrectly.

Chapters 2–4 will be concerned with articulating novel resources for describing states of conceptual disagreement. In the remainder of this introduction, I will take a moment to say why I see this project as a philosophical one.

1.3 LOGICAL ANTHROPOLOGY

In trying to describe the nature of cognition, one commits oneself to doing a kind of metaphysics. However, in taking a practice-oriented approach to this problem, I mean to practice a kind of *critical metaphysics* in something like the Kantian sense. My aim is not to describe cognition as it

really is, independently of anything about us as inquirers or our modes of understanding. Rather, I aim to describe cognition as we think about. Better, I aim to describe the concept COGNITION as cognitive scientists use it (Ch. 5). Or better still, I aim to describe how cognitive scientists think of something when they think of it as an object of cognitive scientific inquiry (Ch. 6).

This is a kind of inquiry I call *logical anthropology*. Anthropology is concerned with human practices, perhaps especially with an eye to such things as rituals, forms of social organization, power dynamics, comparisons across communities, and so on. There is also a strong tradition in analytic philosophy of describing human practices, but with an eye toward their rationality. Epistemologists, for example, are interested inter alia in characterizing and evaluating our practices of forming beliefs and our use of evidence. Philosophers of language are interested in our practices of evaluating utterances and other meaningful performances. Moral philosophers and philosophers of action are interested in our deliberative practices and our practices of evaluating actions and holding people morally responsible. Philosophers of science are interested in scientific practices such as explanation and theory-revision. The description and criticism of these practices are examples of logical anthropology.[2]

My concern is with cognitive scientists' use of one of their own central theoretical terms. That is to say, I aim to characterize the practices of judging something (a system, a process, a behavior, a state) to be cognitive, of disagreeing about such judgments, and of reasoning about cognition in general. As a bit of anthropology, this project is of a piece with other venerable philosophical topics. After I describe some logical resources for understanding states of entrenched disagreement like one finds among cognitive scientists about COGNITION (§ 2.5, Ch. 3, Ch. 4), I will be in a position to offer a rational reconstruction of the scientific concept of COGNITION (Ch. 5). In doing so, I will draw from the extant literature on cognition, and strive to be faithful to the variety of views expressed therein. Nevertheless, I will say a word about why I adopt the

[2] This conception of philosophy bears some resemblance to what George Graham and Terry Horgan call "ideological inquiry" (Graham and Horgan 1994), what Katrin Flikschuh calls "philosophical fieldwork" (Flikschuh 2014), and (in the context of conceptual analysis) what Frank Jackson calls "modest conceptual analysis" (Jackson 1998, 42–44).

particular method that I do. In doing so, I will address the relation between my question a more nakedly philosophical question in the vicinity: "Forget what cognitive scientists think; what is the nature of cognition, really?"

1.3.1 Logical anthropology and speculative metaphysics

Criticisms of the anthropological method can come from either of two sides. The first side is what might be called the side of analytic or "speculative" metaphysics. The speculative metaphysician might ask, "Why bother figuring out what scientists think cognition is? Why not just figure out what it really is?" More generally, one might suppose that it is a better use of time to figure out how things really are, rather than what experts who are not trained in philosophy seem to think, but do not say out loud. After all, discerning what is implicit in scientific practice and making it explicit seems to be a roundabout way of figuring out how things really are, and the scientists might not be right. There is a weaker and a stronger reply to this sort of worry. The weaker reply is that science is an expensive and complicated enterprise, involving tremendous investments of funds, person-hours, and intellectual effort on the part of extremely qualified and clever people. (For non-scientific practices like those studied by epistemologists and philosophers of action, one might say that our deliberative practices are shaped through a significant multi-generational investment.) Surely, there is some value in understanding what return these investments have borne. This reply suggests that there is some value in logical anthropology, but does not vindicate it as a way of learning what the world is like rather than how we go about trying to learn what the world is like. The stronger reply is that the scientific enterprise is our best and most rigorous effort to learn what the world is like, and it is hubris to try to beat science at its own game from the armchair. Logical anthropology can clarify what the fruits are of our best scientific efforts to date. If it is followed by careful and informed criticism, then that is our best method for learning what the world is like. And given the existing disagreements between scientists about the nature of cognition, it is a major undertaking to complete that first step of characterizing the unripe fruits of science.

One of the benefits of logical anthropology in this context is that, done well, it is a methodological safeguard against philosophical busybodyism. There is a worry that the questions that naturally interest philosophers are not the questions that actually interest scientists. It would not do to hold scientists responsible for failing to do well something that they do not see themselves as doing. For example, consciousness science is not at bottom a scientific attempt to resolve the hard problem of consciousness, and consciousness scientists should not be faulted for addressing so-called "easy problems of consciousness" like the relationship between executive control and attention or the neural correlates of conscious contents (Chalmers 1995).[3] It would also not do to draw premises from scientific research without a proper understanding of the aims of those who write it. For example, a philosopher interested in the hard problem should not draw claims from just any paper in consciousness science as if they represented empirical contributions to resolving the hard problem of consciousness.

1.3.2 Logical anthropology and experimental methods

The second side that criticism of the anthropological method can come from might be called the side of experimental philosophy. An experimental philosopher might object, "If you're so interested in what scientists think, why not ask them and collect a lot of data? Why go through the rigamarole of rational reconstruction?" Actually, there are two different activities one could engage in here: asking a lot of scientists what they think, and collecting data that test hypotheses about how they think about cognition (cf. Machery 2004; Devitt 2012); I will address them each in turn. To clarify, however, the goal of an anthropological investigation is to make explicit what is implicit in practice; in many cases, including the case of explicating COGNITION, participants in the practice lack reflective knowledge of the rules of the practice. Just as competent knowers, speakers, and agents

[3] Regrettably, the discourse in consciousness science leaves something to be desired in terms of philosophical sophistication (see Irvine 2012 for review). I choose the example because most philosophers will recognize the stark difference between trying to answer "easy problems" and resolving the hard problem, not because it is a sterling example of clarity among empirical scientists and confusion among philosophers.

judge, speak, and act without being able to describe their own practices with anthropological clarity, scientists often engage in their projects without being able to report various features of their practice. Indeed, I have found that it is easy to elicit a variety of responses from cognitive scientists about what they think cognition is, but these responses tend to be slogans that lack philosophical rigor or sophistication (e.g. "cognition is information processing," or "cognition is what the brain does," or "cognition is the production of flexible, adaptive behavior"; see § 1.2.2.1). Given that my interest is in characterizing the implicit criteria that give shape to a messy practice, then, I can say why I have not collected original data in my treatment of my topic.

Concerning the former, exploratory sort of data, I prefer literature review to collecting new data in this case. I expect that the opinions of individual scientists might include a lot more variation than the views expressed in prominent publications, and I am interested in characterizing the state of scientific discourse rather than the psychological states of individuals per se. A review of prominent literature seems to be a more direct way to measure the state of the discourse than the collection of data on individual intuitions. Collecting exploratory data would be a reasonable method if I were not so fortunate as to have a corpus of literature to draw from, but I am in the fortunate position of having a literature to review.

Concerning the latter, hypothesis-testing sort of data, I concede that that is a good method for going about some anthropological investigations, but resist it in this particular case. Just as I prefer literature to exploratory surveys because my object of inquiry is a social practice rather than a psychological process, my hypothesis should be tested against the views expressed in prominent publications, at least in the first place. It would be interesting, having made the arguments I do here, to test my hypothesis against a body of data on scientist intuitions concerning specific cases, and on the patterns of covariation in such intuitions. That is a matter for further work.

1.4 CONCEPTS AND CONCEPTUAL TRUTH

I am loosely guided here by the Sellarsian dictum that "grasp of a concept is mastery of the use of a word." *Concepts*, as I mean them here, are theoretical entities that are like words or expressions except that they are invariant across successful translations between languages. Thus, the English "red," the German "rot," and the Japanese "赤い" correspond to the same concept, RED. (A German writer might render it as ROT but we are referring to the same concept.) Concept-talk, understood this way, is rather like discourse with Quinean semantic ascent (1960, 270–276), but in which one explicitly masks off phonological and syntactic features of a word to focus on the semantic (and perhaps pragmatic) features. Thus, concepts are entities that like linguistic expressions refer, have meanings, can be used or mentioned, &c. Since concepts have these properties, they are associated with various norms of proper use: criteria of ascription, criteria of individuation, inferential proprieties, theoretical and practical consequences of application, and so on. These norms derive in part from social conventions about the use of concepts, and in part from facts about the world that make those social conventions apt or inappropriate.

My primary interest in this dissertation is how scientists use the concept COGNITION, e.g. what they take its extension to be, what consequences they infer from judging something to be cognitive, and which considerations they appeal to when challenged on their judgments that a phenomenon is cognitive. However, I do not take myself to be committed to any strong or controversial view about the nature of reference or meaning (e.g. to the view of Brandom 1994). Even those who reject use-theories of meaning or inferentialist semantics acknowledge that words or concepts have uses, that there are proprieties of inference regarding concepts, and that these uses and proprieties may be fruitfully investigated. I also intend to sidestep a number of concerns about the nature and individuation of concepts in natural language, and the unity of concepts or concept usage as an explanandum (e.g. Machery 2009). My concern is as I said with the use of the concept of COGNITION, understood as the language-indifferent use of the word "cognition" and its cognates in the rarefied discourse of academic science literature.

1.4.1 Conceptual explication without analyticity

I regard these investigations into the norms of proper use for concepts as a species of conceptual analysis, or, as I will prefer to say in what follows, *conceptual explication*. I prefer the latter term in part because I do not understand this activity to consist in breaking a concept into parts or describing its "true nature." Rather, it is simply making explicit various norms of proper use, and as I said above I am officially agnostic as to what, if anything, such use-norms have to do with "meanings." So an explication of a concept is judged not according to whether the explication corresponds faithfully to something happily called a "meaning," but according to whether it can elucidate patterns of ascription and inference that feature the concept. Indeed, when I do resort to *means*-talk in what follows, it is generally some form of "speaker meaning" that I invoke, rather than "semantic" (i.e. "word" or "sentence") meaning. An explication is a product of philosophical labor, judged by its potential to contribute effectively to matters of theoretical or practical interest and not by whether it accurately limns the ultimate structure of reality. For example, a successful explication of RED might be expected to describe the norms for ascribing redness in a variety of contexts, e.g. by unaided sight and by spectral analysis, in diverse conditions of illumination or irradiation. Or it may describe norms for only a restricted set of such contexts, making explicit various implicit facts about the use of "red" and cognate terms in e.g. folk but not laboratory contexts.

And since explication in my sense does not trace something called a "meaning," I would resist the accusation that this activity violates W. V. O. Quine's prohibition on distinguishing between analytic and synthetic statements. Whereas Kant understands analytic truths to be those where the predicate concepts are contained within the subject concept (Kant 1787, a6–7/b10–11), I do not claim that conceptual explications—in the sense used here—make explicit the components of concepts. Quine understands analytic sentences as those that are true "by virtue of meanings and independently of fact" (Quine 1951, 21), but I do not insist that conceptual explications make explicit the meanings of expressions, nor that they exclude other facts. Rather, conceptual explications in

my sense are true just when they truly describe (some of) the norms governing the use of a concept, where those norms describe linguistic conventions as well as other facts about the world. I am happy enough to call these norms "conceptual truths" only so long as they are not understood to be analytic in the sense of Quine or Kant. Indeed, explications might make reference both to the sorts of considerations normally considered analytic (e.g. "bachelors are unmarried") and to the sorts of considerations normally considered synthetic (e.g. "bachelors do not tend to live tame lives"). I therefore do not need to enforce a distinction between conceptual truths and obvious or unquestionable beliefs, much less an immutable distinction (cf. Machery MS, Chap. 1 § 3.2). And like Jackson (1998, 42–44), I accept that conceptual explication does not furnish us with apodictic claims about the nature of the world apart from the way we think of it.

The term "explication" is notably used by Rudolf Carnap and by Quine, and I take their conception of explication to be more amenable to my method than a notion of "analysis" with undertones of breaking into parts. In particular, Carnap and Quine see explications as products of philosophical labor to be judged on the basis not of faithfulness to meaning, but on the basis of their potential to contribute to various practical or theoretical endeavors. Thus, the spirit (if not the letter) of Carnapian or Quinean explication is amenable to the ecumenical project I describe in Chapter 2. However, I do not mean to undertake all of Quine's and Carnap's commitments, either. I will take a moment to distinguish my conception from theirs before concluding this chapter.

1.4.2 Resisting ontological gentrification

Carnap clarifies that "By the explication of a familiar but vague concept we mean its replacement by a new exact concept" (1947, 7). He approvingly cites Russell's explication of numbers. Quine describes a similar activity, lingering over the history of the ordered pair in mathematics (1960, 257–262). Quine characterizes the vague concepts for which explication is a remedy as "defective nouns" (258), clarifying that

> We do not claim to make clear and explicit what the users of the unclear expression had unconsciously in mind all along. We do not expose hidden meanings, as the words 'analysis' and 'explication' would suggest; we supply lacks. We fix on the particular functions of the unclear expression that make it worth troubling about, and then devise a substitute, clear and couched in terms to our liking, that fills those functions. Beyond the conditions of partial agreement, dictated by our interests and purposes, any traits of the [unclear expression] come under the head of "don't cares." (258f)

Quine is enthusiastic about revisionary consequences. He is eager to take language that he regards as vague or ambiguous and to replace it with "crisp" language that is better suited to his purposes. That is a fine thing to do, I think, but it is not an activity that suits my anthropological interests. The activity of explication need not have an austere Quinean character. My divergence from Quine can be traced to two themes: Quine's resistance to psychologizing, which I share but from which I draw different conclusions, and Quine's famous aesthetic preference for ontologies like "desert landscapes," which I do not share.

I do agree with Quine that explication does not reveal hidden "meanings." And I agree that there need not be unique true explications for any concept; Quine insists that there are several distinct but perfectly adequate explications of ORDERED PAIR. I admitted above that there might be several distinct but adequate explications of RED, even explications that imply different facts or that are valid for RED-ascriptions in different circumstances (e.g. everyday color ascriptions, chemical analysis) for the sake of different projects (e.g. epistemology, spectroscopy). Regarding the first theme, Quine is concerned that he not be understood as making explicit "what the users of the unclear expression had unconsciously in mind all along." He is concerned in part because he wishes to avoid the "paradox of analysis," which is roughly the puzzle that arises when one recognizes that devising an explication of a concept is supposed to be an intellectual achievement, and yet what it does is simply make explicit what users of the concept all already must have known.

Quine illustrates the paradox with some disparaging remarks about Oxford ordinary language philosophy (259). I would resolve the paradox by attending carefully to what is meant by "known."

The appearance of paradox tends to vanish when one recognizes the complexity of linguistic practices generally, and the extent to which their details are opaque even to their competent practitioners. The bread and butter of 20th century generative syntax and semantics live in this gap; few English-speakers understand the norms of anaphora, do-support, or the that/which distinction, and yet all fluent English speakers follow these norms and recognize their violation. There may well be a clever philosophical puzzle about this, but it is no methodological objection to conceptual explication as a way of making explicit what is already implicit in a socially-instituted practice. I happily assimilate the project of conceptual explication to the practice articulating descriptive grammars in linguistics, noting only that what we call conceptual explication caters more to the interests of the logical anthropologist.

Regarding Quine's preference for desert landscapes, I must express some qualified resistance. Quine says of his conceptual housekeeping that it

> is the task of making explicit what had been tacit, and precise what had been vague; of exposing and resolving paradoxes, smoothing kinks, lopping off vestigial growths, clearing ontological slums. (1960, 275)

I admire Quine's talent for metaphor, but the last two phrases in this list should not be assimilated to the first four. The former elements are an apt enough description of logical anthropology and an acknowledgement of the sorts of idealizations it might involve. The latter two elements introduce methodological prescriptions that make me uncomfortable. After all, it is not that vestigial growths like the tonsils and appendix are worthless organs; it is just that one can if necessary live without them, and there are circumstances where they do more harm than good. More to the point is Quine's last metaphor, of clearing ontological slums. Slums may contribute in an important way to the life of a city, and after all there are reasons that people live there. A practice-oriented

philosopher of science should not declare features of a conceptual practice—Quine's ontological slums—worthless or pernicious just because she does not understand them, or out of a misguided preconception about what respectable science looks like. My point will be clearer with an example.

Mark Wilson is associated with the view that concepts have lives that appear, from the standpoint of someone accustomed to working with formal languages, rather disorganized (see e.g. Wilson 1982, 2006). For example, the notion of HARDNESS in engineering is assessed in many different ways, several of which do not converge—resistance to indentation, resistance to impression, result of scratching, result of traumatic impact, &c., where the appropriate assessments depend on the type of material and the use to which it might be put (2006, 337ff). The Quinean impulse is to make distinctions, to replace a "defective expression" like "hardness" with univocal and precise ones. However, there is a reason to imagine more precise conceptions of hardness as part of a common concept with a complex structure rather than as entirely distinct notions. Some of these assessments give convergent results in some circumstances. Some of them are not clearly distinguished for certain purposes (because they do not need to be). And even the more precise characterizations do not correspond uniformly to more ^{S}fundamental^{S} physical properties. In short, the concept of HARDNESS—the messy, unsanitized version used by actual engineers—has a role in the practice of engineering that could not be served effectively by Quinean replacements. Wilson calls complex conceptual structures like this, where the norms governing the use of a concept vary according to the circumstances of the user, "facades." He remarks:

> In real life… facades sometimes perform these fine offices in such a discrete and imperceptible manner that, as an undesirable side effect, they create *ur*-philosophical perplexities when their structuring is misunderstood and utopian projects are plotted upon an erroneous diagnosis. (Wilson 2006, 203)

Quine's taste for desert landscapes, or his misguided understanding of scientific respectability, leads him to advocate what Wilson archly calls "wholesale ontological destruction" (2006, 274).

Clearing Quine's "ontological slums" would hobble scientific practice. Philosophy at best deprives itself when it does not explore these neighborhoods, and at worst contributes to a harmful sort of conceptual gentrification when it succeeds in evacuating them.

In short, while I do take many Quinean lessons to heart, I reject his meta-theoretical austerity. Rather, I take my methodological cues from the David Lewis who said

> I am not one of those philosophers who seek to rest fixed distinctions upon a foundation quite incapable of supporting them. I rather seek to rest an unfixed distinction upon a swaying foundation, claiming that the two sway together rather than independently. (Lewis 1973, 92)

In what follows, I will provide a Lewisian foundation for understanding the scientific concept of COGNITION. The foundation I offer does not provide a surface on which to build—for the concept I seek to explicate is a concept that is already in wide use—but rather an anchor through which an existing structure can, hopefully, better withstand the stresses now placed on it. And while I will not map every dark alley in the ontological slums of cognitive science I will endeavor to be no agent of gentrification.

1.5 COMING ATTRACTIONS

In the next chapter I will describe and motivate the central problem of this dissertation: the problem of cognition. This is the problem hinted at in my discussion of the cognition border wars (§ 1.2.2 above). In motivating that problem, I will describe four functions that its solution might serve, and clarify which of those functions I prioritize. I will also introduce the first logical resource that will allow me to describe my solution: ecumenical extensional adequacy. In Chapter 3, I will introduce a second logical resource: construal-based explication. This chapter will draw on the

anthropological commitments I described above in Section 1.3, making them more concrete in the context of conceptual explication. Chapter 4 describes a third logical resource, parameterization, which completes the logical toolkit with which I can construct a Lewisian foundation for COGNITION. In parameterized explication, some terms can be assigned different interpretations so that the explication determines different extensions under different conditions. With these tools in place, I will describe my favored explication for the concept of COGNITION, the sensitive management hypothesis. The sensitive management hypothesis is a parameterized explication that successfully models the pattern of disagreement exhibited in cognitive science (or so I will argue; Ch. 5). I will also describe a particularly permissive version of the sensitive management hypothesis, on which the interpretations of the parameters remain construal-sensitive. I will defend the permissive version of the sensitive management hypothesis against some objections and sketch out how it can shed light on other features of cognitive scientific discourse, such as the ascription of representations (Ch. 6). Finally, in my concluding chapter (Ch. 7) I will revisit the four functions described in Chapter 2. I will evaluate my approach in general, and the sensitive management hypothesis in particular, against the standards for success I outlined in the beginning. Let us turn to that beginning, then, by reconsidering the problem of cognition.

2.0 RETHINKING THE PROBLEM OF COGNITION

2.1 THE PROBLEM OF COGNITION

Cognitive science is the interdisciplinary study of something called "cognition." Cognitive scientists and other researchers traffic in talk about cognition and its cognates (cognitive states, processes, &c.), and take these things, whatever, they are, to be their object of study. There is general agreement about the component disciplines of cognitive science—parts of psychology, computer science, neuroscience, artificial intelligence, philosophy, linguistics—and there is also agreement on which topics belong to cognitive science—e.g. perception, learning, memory, decision-making, language, motor control, &c. However, there is no agreement about what cognition is (Adams and Aizawa 2001; Godfrey-Smith 2002; Prinz 2004; Lyon 2006; van Duijn, Keijzer, and Franken 2006; Adams and Garrison 2013; Buckner 2015). Many feel a need for more clarity than this. Fred Adams and Rebecca Garrison claim that it is "embarrassing to say the least for there to be a science of cognition… that is unable to say what constitutes cognition" (2013, 340). Jesse Prinz says that "It is scandalous that cognitive science has not settled on a definition of cognition" (2004, 41). This state of affairs, that there is no satisfying account of the nature or extension of cognition despite a felt need for such an account, is the *problem of cognition*.

Some philosophers deny that there is any satisfying, unique solution to the problem of cognition (Chemero 2009, 212n8; Clark 2010b, 62). Many share a common attitude that "there really isn't a lot at stake, scientifically, in our efforts to delineate the conceptual boundaries of cognition" (Ramsey forthcoming, MS 11). Nevertheless, philosophical interest in the problem has renewed in the present century with several attempts to describe "the mark of the cognitive" (beginning with

Adams and Aizawa 2001). These "marks" usually take the form of partial definitions of COGNITION—necessary and/or sufficient conditions for ascribing cognitive properties—that serve to justify some set of theoretical perspectives and undermine others. However, none of these proposals has proven popular. More recently, some philosophers have turned their attention from "direct assault" on the problem (Aizawa forthcoming, MS 3) toward more modest, substantive questions about the scientific concept COGNITION, such as whether cognition is identical to a kind of behavior (Aizawa forthcoming; Shapiro 2013), or whether cognition is representational as a matter of conceptual fact rather than as a matter of empirical fact (Ramsey forthcoming).

I have two aims in this chapter. The first is to articulate the problem of cognition more explicitly than has been done before, inter alia describing what is at stake, in order to combat the common attitude that the problem of cognition is not worth addressing. Specifically, I will describe four ends that would be served by a characterization of cognition. In doing so, I assess whether the dominant strategy for answering the problem of cognition—that of seeking a traditional definition aiming to justify certain theoretical perspective over rival perspectives—is likely to accomplish those ends. I am not sanguine about its prospects. My second aim is to describe an alternative criterion of success for understanding the scientific concept of cognition—what I call ecumenical extensional adequacy. I argue that an ecumenical approach is more likely than the dominant strategy to serve more of the ends that motivate the problem of cognition. I do not offer a solution to the problem of cognition or a mark of the cognitive in this chapter; the approach I advocate is so thoroughly absent from the contemporary discussion that it merits a defense independently of any specific ecumenically adequate solution. What I do offer is a clarification of what is at stake in the problem of cognition, and the articulation of an approach to resolving it that has been overlooked so far.

I will begin by rehearsing some historical considerations that have recently made the problem of cognition more urgent. Extant treatments of the problem of cognition are usually motivated by a desire to resolve open questions about the nature of cognition that have become pressing since the 1980s. The fact of widespread expert disagreement about the nature and

extension of cognition produces practical demands for policing the boundaries of the cognitive scientific enterprise. I will also argue, however, that the scientific concept of cognition is in the midst of unresolved conceptual development. An explication of COGNITION might aim to characterize this development-in-progress, rather than to make bets about how it will turn out. In the second half of the chapter, I suggest that whereas the extant "marks of the cognitive" aim primarily to serve the boundary-policing function, the dominant approach to the problem of cognition serves that function imperfectly. I then outline an alternative approach to the problem of cognition that aims primarily to make explicit the implicit commitments of cognitive scientists regarding their object of study. This is to be done *ecumenically*, not by taking sides in intramural disputes but by articulating what is supposed in common and what is at issue in those disputes.

2.2 THE COGNITION BORDER WARS

The problem of cognition is not a matter of merely academic curiosity. It is sometimes observed that biologists are not particularly impeded for want of a definition of LIFE (Cleland 2012; Machery 2012), and it is therefore unclear why one should want an explication of COGNITION. I will say at the outset that I agree that biologists can do biology without a definition of LIFE, and that much of cognitive science can likewise proceed unimpeded without an explication of the scientific concept of cognition. On this matter I am in the company of Andy Clark (2008, 239n3), Robert Rupert (2013, 42–43), Ken Aizawa (forthcoming, MS 1–3) and many others. Nevertheless the problem of cognition is pressing, and more pressing than corresponding questions in biology, for several reasons. First, unlike definitions of LIFE for biology, explications of COGNITION are taken to settle criteria of demarcation for cognitive science. Biologists will study viruses even if on many proposed definitions they are not considered living, and they would study some non-homeostatic processes even if biological systems turned out to be necessarily homeostatic. But the claims that "plants have no cognitive processes" or that "models that posit no representations are not cognitive models" are

taken to have the consequence that cognitive scientists have no business studying plants or considering non-representational hypotheses. A second reason, which compounds the urgency of the first, is that contemporary cognitive scientists disagree strongly about the boundaries of cognition, and therefore about the boundaries of cognitive science.

Recent interest in the problem of cognition was inspired in large part by what I call the *cognition border wars* (§ 1.2.2), and in particular by controversy over the hypothesis of extended cognition (Clark and Chalmers 1998) which holds that cognitive processes are constituted in part by processes that extend outside of the brain and the body. However, the hypothesis of extended cognition is merely the straw that broke the camel's back. It is a latecomer in a host of (what I will call) anti-classical perspectives in cognitive science that began gathering support in the 1980s, following the connectionist challenge to classical cognitive science (§ 1.2.1). Some of these perspectives, like connectionism and dynamicism (Rumelhart 1989; Thelen and Smith 1994; van Gelder 1998), promote non-classical strategies for modeling cognitive processes. Other anti-classical perspectives have unintuitive consequences about where cognition can be found (so they cover all the ground that Chemero and Silberstein 2008 identify as the "new philosophy of mind"). Arguments that machines can exhibit genuine cognition are as old as classical cognitive science (Turing 1950; Putnam 1967a), but anti-classical partisans are apt to recognize cognition in a host of other contexts. Proponents of embodied cognition argue that cognitive processes extend out of the brain and into the non-neural tissues of the body (Gibbs 2005; Chemero 2009; Clark 1997; Varela, Thompson, and Rosch 1991). Proponents of group and social cognition argue that the coordinated activity of multiple corporeal agents sometimes counts as a distinctive kind of cognitive phenomenon (Hutchins 1995). Some researchers now hold that plants have cognitive processes (Trewavas 2003; Calvo Garzón 2007). Enactivists hold that cognition consists in activity on the part of a whole organism, and not in the manipulation of representations or information (Varela, Thompson, and Rosch 1991; Hurley 1998; Noë 2006), and sometimes that all living organisms—including microbes—engage in cognitive activity (Stewart 1996; van Duijn, Keijzer, and Franken 2006; Thompson 2010). I will describe some cases in greater detail below in Section 5.3. These anti-

classical perspectives are all controversial—none is generally accepted, but each is defended by cognitive scientists from a range of disciplinary backgrounds. Importantly, though philosophers have often been the most systematic exponents and critics of the anti-classical perspectives, the debates of the border wars turn on bodies of empirical and formal research in the other component disciplines of cognitive science (see e.g. Clark 1997, 2008; Rowlands 2010 for this style of argumentation). These perspectives are each embraced by groups of researchers of many backgrounds, and inform cognitive scientific research in several disciplines. For example, enactivism was spearheaded by biologists and found a firm proponent in the psychologist Eleanor Rosch. The ecological approach was first articulated by psychologist J. J. Gibson. Proponents of embodied cognition and its value to empirical work are found throughout psychology (see e.g. Gibbs 2005 for discussion) and computer science, especially in robotics (Brooks 1991; Webb 1994). So the border wars should not be dismissed as fuss about the overly bold claims of iconoclastic philosophers.

The result is that there are open questions in cognitive science now where no one (or very few) saw open questions up until the 1980s:[4] Might cognition involve non-representational processes? Are some properties of muscles, or the structure of the body, parts of cognitive mechanisms? Can artifacts or the structure of the environment be parts of cognitive mechanisms? Do plants have cognitive processes? Microbes? The state of affairs created by these border wars might be called *the fact of widespread expert disagreement* about the nature and extension of cognition. Shortly after Andy Clark and David Chalmers argued that cognitive processes are realized by structures outside of the body, Fred Adams and Ken Aizawa argued for a partial definition of COGNITION meant to exclude the possibility of several of these anti-classical

[4] The history is of course complicated. Some border controversies have predecessors. The characterization by scientists of "unconscious" processes by comparison to highfalutin cognitive processes like inference goes back at least as far as the 19th century, e.g. Helmholtz' (1867) "unconscious inference." Scientific consideration of microbe cognition goes back at least as far as early enactivism among Chilean biologists in the 1970s (Maturana and Varela 1980, originally published in 1970). However, the mainstreaming (or re-adoption) of these perspectives has accelerated since the 1980s, when the border wars began (§ 1.2.2).

perspectives. They have been joined by other philosophers (notably Rupert 2009; Rowlands 2010; Buckner 2015) describing rival accounts (§ 1.2.2.2).

Thus, the first reason to care about characterizing cognition correctly is to settle the open questions inspired by the border wars. Sven Walter (2010) for example argues that questions about extended cognition cannot be answered until we know the "mark of the cognitive." Normally, these questions are taken to have normative consequences for the way cognitive science is done. If cognitive science just is the study of cognition, then clarity about the extension of COGNITION delimits the proper scope of the cognitive scientific enterprise (Rupert 2004). That is to say, cognitive science is the study of cognition, and the extension of COGNITION determines demarcation criteria for cognitive science.[5] Without a definition of COGNITION, or something like a definition, it is feared that the practice of cognitive science might be wildly misguided. We might be studying the wrong object, or studying it the wrong way. Addressing these concerns appears to be a primary dialectic goal of most of the present literature on the "mark of the cognitive": Adams and Aizawa (2001, 2008) conclude based on their proposed "mark of the cognitive" that many anti-classical research programs, including those into cases of putatively extended and social cognition, are misguided. Rowlands (2009, 2010), on the other hand, concludes based on his alternative "mark" that certain anti-classical research programs are legitimate, including those concerning extended cognition.

If the correct explication of COGNITION can contribute to disputes about the legitimacy of various research programs then the problem of cognition is not only a matter of scientific but also of social significance. After all, cognitive science research is often funded by national funding bodies with limited resources. The funding of misguided research programs unjustifiably draws

[5] Ramsey (forthcoming) also claims that cognitive science should be understood as the study of cognition, whatever cognition is, but denies that any speculative "mark of the cognitive" should limit our inquiry. I am inclined to agree that it should not, as will become clearer, but disagree with Ramsey that there is therefore no important end served by trying to resolve the problem of cognition (I will outline four such ends in this chapter). However, cf. Rupert (2013) for a dissenting view, that cognitive science is not aptly characterized as the study of cognition; his dissent is based on the premises that in order for that description to be a happy one cognition must be a well-behaved natural kind, and that cognition is not a natural kind.

funding away from legitimate research. Consider, for example, the recent trend of dedicating substantial resources toward brain-centered research projects (such as the BRAIN Initiative, the Human Connectome Project, and Henry Markram's Human Brain Project) and away from more traditionally "cognitive" behavioral or formal work. It behooves us to be as clear as we can about the nature of cognition, and therefore the value of various strains in cognitive science, in order to reflect on what, if anything, we are missing out on due to these trends. Thus, although William Ramsey asks for a rationale "beyond turf wars and funding issues" that justifies interest in the problem of cognition (forthcoming, MS 11), I am inclined to think that that alone would be enough. After all, the products of philosophical effort—arguments, explanations, accounts, conceptual analyses, distinctions, and so on—are not merely attempts to limn the ultimate structure of reality. They are devices for thinking clearly and justifying one's claims, not just by their original expositors but by other inquirers. Philosophers of cognitive science are part of the scientific community, and if our work has edifying practical consequences for the study of cognition then so much the better for all of us. Nevertheless, I contend that there indeed are reasons beyond turf wars and funding issues to care about the problem of cognition.

2.3 THREE MORE REASONS TO CARE ABOUT THE PROBLEM OF COGNITION

To reiterate, I am not claiming that we need a "mark of the cognitive" in order to do cognitive science at all. However, there are reasons we might want to have a characterization of cognition ready to hand. There are at least four functions that a solution to the problem of cognition could serve for inquirers. Above I described the boundary-policing function that has been the focus of contributors to the literature on the "mark of the cognitive" in the present century. However, there are also epistemological, public relations, and metaphilosophical functions that a solution to the problem of cognition could serve (Table 2.1). I will motivate these remaining three functions with

Table 2.1: Four functions for an explication of COGNITION.

Four functions for an explication of COGNITION	
I. boundary-policing (§ 2.2)	II. epistemological (§ 2.3.2)
III. public relations (§ 2.3.3)	IV. metaphilosophical (§ 2.3.3)

arguments that the scientific concept COGNITION is changing, and that this change is well-described as progress.

2.3.1 The concept COGNITION is changing

Perhaps the clearest example of change is that cognitive scientists now routinely distinguish between two fairly distinct senses of "cognition." A *highfalutin* kind of cognition, sometimes called "higher cognition," is roughly synonymous with "rational thought" and figures in expressions like "cognitive therapy" and "cognitive control." Cognition in this sense is normally understood in contrast to phenomena such as perception and affect. Prinz (2004), for example, suggests a working definition of COGNITION meant explicitly to contrast with perception. More recently, Nico Orlandi (2014, 6f) discusses the perception-exclusive notion of COGNITION. The highfalutin sense of "cognition" is closely related to the traditional sense of "cognition" inherited from Latin. In the late 20th century, though, scientists began using the word "cognition" to refer to a more *inclusive* category of phenomena that contains all the proper objects of study for cognitive science. This category uncontroversially includes un-thoughtlike phenomena such as perception, affect, motivation, and motor control. (The distinction between two senses of "cognition" is drawn explicitly in Greene et al. 2004, 389; and Rowlands 2009, 7.) "Cognition" in this inclusive sense is a

scientific neologism of recent decades. Since highfalutin cognition is a more restrictive phenomenon, understood explicitly in contrast to phenomena like perception and affect that are generally agreed to be proper objects of cognitive scientific study, the problem of cognition cannot be about the extension of highfalutin cognition. The explicit restriction of "highfalutin cognition" as a descriptor for e.g. non-perceptual contexts implies that it cannot serve the boundary-policing function. However, its sincere use in scientific contexts should incline us to think that it is a scientific concept. That is to say, the contrast is not between a pretheoretic (highfalutin) and a scientific (inclusive) notion of cognition, but rather between two notions used simultaneously by scientists. These senses are nevertheless distinguishable in terms of their consequences of application—to claim that a process is cognitive in the highfalutin sense implies inter alia that it is not perceptual or not affective (or is otherwise to be contrasted with "lower" cognition). To claim that a process is cognitive in the inclusive sense is implies inter alia that it is in the proper domain of cognitive scientific inquiry.

2.3.2 The conceptual change concerning COGNITION is progress

The distinction between highfalutin and inclusive cognition is evidence of the conceptual change already accomplished in cognitive science, but the border wars are inter alia about whether inclusive cognition encompasses even more natural phenomena than perception, affect, and motor control. The open questions of the border wars were undreamt of or safely ignored before the 1980s, but are matters of legitimate controversy now. The reason that the questions of the border wars have become pressing to cognitive scientists is that their unintuitive conclusions have been motivated by argument and evidence. Advocacy for embodied and situated cognition, in particular, has been conducted with arguments that draw on intuition-twisting empirical studies (see e.g. Clark 1997; Gallagher 2005; Gibbs 2005; Rowlands 2010 for some book-length reviews of such arguments). As a result, though anti-classical perspectives like enactivism and dynamicism remain controversial to cognitive scientists, it has become increasingly common for empirical work to embrace other

perspectives like situated cognition and Gibson's ecological psychology (Robbins and Aydede 2008).

Briefly, then, here is an argument for the claim that COGNITION is undergoing conceptual progress. First, the concept is certainly undergoing a process of change. This is evidenced, first of all, in the relatively recent distinction between what I called highfalutin cognition and inclusive cognition. Second of all, the border wars raise open questions about the nature of cognition that were mostly considered either absurd or radical before the 1980s. That these questions have meanwhile become legitimate implies that cognitive scientists think of their object of inquiry in a new way. Hence, the concept COGNITION as used by cognitive scientists has changed and is changing (recall that by *concept* I mean the language-independent use of a word, as guided by linguistic and empirical norms). Furthermore, since the disputes of the border wars have been motivated by appeal to empirical and formal evidence, the conceptual change driven by the border wars is well-characterized as progress, rather than as arbitrary, or as a reflection of merely notational trends, or as conceptual drift. The sense of "progress" here is not that of closer approximation to the truth, whatever that might be, but that of responsiveness to evidence. I contend, therefore, that we should understand the border wars as evidence of ongoing conceptual progress among cognitive scientists concerning their understanding of the object of their inquiry. That is, the open questions of the border wars are, in part, questions about how scientists should understand and ascribe the concept COGNITION, and contributions to the border wars are inter alia attempts to predict and influence the course of conceptual change.

Conceptual progression of this sort is a cornerstone of science. Much of the most interesting scientific innovation is conceptual innovation. Our science is more advanced than Aristotle's not because he simply didn't examine the world carefully enough, but among other things because we have learned better vocabulary in which to couch our questions and theories. Contemporary scientific concepts like GENE and even TEMPERATURE are hard-won fruits of scientific inquiry (Griffiths and Stotz 2006; Chang 2008), and their careers of development pushed against pretheoretic intuitions (a familiar reference regarding temperature is found in Francis Bacon's

Novum Organum, II.XI). If the border wars are evidence of conceptual progression concerning COGNITION, then we are presently living through a moment of major conceptual change in the scientific understanding of the mind. New research questions and theoretical trends in cognitive science reflect what we as a scientific community have learned about cognition from decades of doing cognitive science. For example, the notion of inclusive cognition has proved useful because whatever important dissimilarities there may be between rational judgment (highfalutin cognition) and perception or emotion, there are some significant continuities captured by cognitive scientific methods. That these continuities are important is generally agreed upon notwithstanding disagreements concerning the nature of these continuities—whether they are information-processing interactions, for example, or common recruitment by linguistic capacities.

The countenancing of counterintuitive claims by proponents of anti-classical research traditions is a symptom (though not a criterion) of conceptual maturation. However, the lack of any agreement about how to resolve the problem of cognition implies that whatever we have learned has not been made very explicit, and that the border wars remain controversial implies that cognitive scientists disagree on how to conceive of their object of study. We may hope that a successful explication of COGNITION would make some of that knowledge more explicit, in the way contemporary treatments of TEMPERATURE or GENE make explicit what we have learned about heat and genetics by doing physics and biology. And since cognitive scientists disagree about how the concept COGNITION is to be applied, we might hope furthermore that an explication would make explicit which if any commitments are shared among cognitive scientists, and which commitments are not shared. Thus, in addition to the boundary policing function described in the previous section, there is a second, epistemological, reason to address the problem of cognition. An explication of COGNITION should make explicit what cognitive scientists have learned, but not yet clearly articulated, about their object of study.

2.3.3 Other benefits of characterizing COGNITION

I said earlier that the problem of cognition is that there is no clear account of how cognitive scientists in general understand cognition, though there is a felt need for such an account. If this is what the problem of cognition is, then demarcation is not the only—or even the most interesting—purpose of addressing the problem. Since the scientific concept of cognition is undergoing change and, indeed, progress, there are other practical and theoretical problems that we might hope to address by resolving it. In particular, if an explication of COGNITION can illuminate what we have learned about cognition by doing cognitive science, and if it can put certain counterintuitive claims about the nature of cognition in perspective, then it may also serve a public relations function and a metaphilosophical function. These two functions are best served by an approach that seeks primarily to make explicit the as-yet implicit knowledge that cognitive scientists have accumulated.

Regarding public relations, the inclusive concept of cognition is poorly understood by those who do not work in the cognitive sciences. Novice students and non-experts often struggle to understand how something called *cognitive* science can possibly deal with phenomena like perception, emotion, or dreaming. Furthermore, the controversies of the border wars, over e.g. extended cognition or plant cognition, are often considered absurd by laypeople and certain philosophers, although they are taken seriously by many cognitive scientists. A successful explication of COGNITION that aims to make explicit the implicit commitments involved in scientific work would, even if only in the form of a slogan, make the nature of cognition and the enterprise of cognitive science more accessible to non-specialists.

Finally, increased clarity about the nature of cognition, at least as conceived in the scientific image, has the potential to shed light on other recalcitrant philosophical problems. If, as is commonly thought, mental states are a subcategory of cognitive phenomena, then solutions to the problem of cognition stand to contribute to disputes about the place of minds in nature. Cognitive phenomena are also frequently at the center of controversies about scientific reductionism, scientific abstraction, multiple realizability, and the nature of computation and representation in

the cognitive sciences. An explication of COGNITION would have the metaphilosophical benefit of serving as a reliable resource for appeal regarding arguments in these literatures. At the moment, some form of Putnam- or Lewis-style functionalism usually serves as such a resource for appeal, despite widespread misgivings about the functionalism's adequacy for this purpose (see Appendix). If science is to be a guide to addressing these philosophical questions, however, we are safest in using as a resource for appeal a characterization of cognition that makes explicit the implicit commitments of scientists, but does not incorporate speculation about the answers to questions that are, as far as cognitive scientists are concerned, empirically open.

So there are four aspirational benefits of a successful characterization of cognition: First, it would have normative practical implications for cognitive science, specifically on the legitimacy of certain research programs (Table 2.1 I). Second, it would reflect what contemporary cognitive science knows (or perhaps merely believes) about cognition (Table 2.1 II). Third, if compact it would enable clear communication across disciplines and to non-experts (Table 2.1 III), and finally it would serve as a resource for appeal regarding other philosophical topics of interest (Table 2.1 IV). These four ends should also serve to dispel a concern expressed by Robert Wilson, who worries that the demand for an explication of COGNITION is unbecoming for

> philosophers who take one of the chief lessons of the failure of logical positivism in the philosophy of science, the collapse of the analytic-synthetic distinction along Quinean lines in the same, and the limitations of conceptual analysis to be a deep suspicion of the search for such principles. (2010)

That is, Wilson worries that many attempts to produce a conceptual explication of COGNITION will be inconsistent with a righteous rejection of the analytic-synthetic distinction. I believe this worry can be overcome. First of all, since the explication demanded is one that draws explicitly upon empirical knowledge, and whose adequacy depends on its responsiveness to that knowledge, it does not presuppose a strong distinction between analytic or meaning-constitutive claims and synthetic

or extra-conceptual claims (see § 1.4). Furthermore, by acknowledging the four functions above, the problem of cognition can be distanced from a demand for faithfulness to some ill-defined "meaning." Instead, an explication that resolves the problem of cognition can be judged by its potential to serve these four ends. Even if no single explication can simultaneously serve all four of these functions, any proposed explication of COGNITION might be measured against its potential to yield some combination of these four benefits (see § 7.3–7.4).

2.4 HOW NOT TO CHARACTERIZE COGNITION

Some philosophical approaches are promising ones for achieving these four benefits, and some are not. For example, it should be clear that an account of cognition along traditional lines, that makes judgment or rational thought central, is not what is called for in this context. The highfalutin notion of cognition is not the one that serves to demarcate the bounds of cognitive science, nor the one whose usage encodes the most interesting conceptual maturation, nor is it the one that eludes non-specialists or is needed as a resource for appeal regarding philosophical discussions that take the scientific conception of minds to be highly germane. The intended object of description here is the novel scientific concept of inclusive cognition wrought by cognitive scientists—the one with the surprising consequence that it makes the border wars of the 1980s and '90s intelligible. This much is agreed upon by the major contributors to what I have called the problem of cognition. Some of them construe their proposals as empirical hypotheses (e.g. Adams and Aizawa 2001; Rupert 2004; Adams and Garrison 2013; Buckner 2015). Some justify their proposals by appealing to the structure of contemporary theories (e.g. Rowlands 2009). All agree that contemporary or future cognitive science is the measure of the "mark of the cognitive."

William Ramsey is an apparent exception. He argues that we should allow our "intuitive, pre-theoretical image of the mind" guide us in roughly demarcating the subject matter of cognitive science. This pre-theoretical image is furthermore "in all likelihood a cluster concept with fuzzy

boundaries, with some prototypical processes in the center and more obscure or atypical processes on the periphery" (Ramsey forthcoming, MS 12). But since the scientific concept of cognition is changing, as I argued above, Ramsey's strategy faces a dilemma. Either the "pre-theoretical" conception that guides our characterization should be insensitive to the ongoing conceptual change, or it should accommodate that change. If our pre-theoretical conception is insensitive to conceptual development in cognitive science, this strategy would require us to backtrack in ways that may obscure implicit insights won over the decades, rather than illuminating them. (This is almost certainly not a view that Ramsey actually supports). If our pre-theoretical conception is sensitive to conceptual developments in the science, then Ramsey does not really disagree with the guideline that contemporary or future cognitive science is the measure of the "mark of the cognitive," and it is at best awkward to describe our conception of cognition as "pre-theoretical," even if it developed from such a notion.

However, contra the major contributions to the literature so far, it is also probable that a characterization of cognition will not take the form of a traditional definition in terms of individually necessary and jointly sufficient conditions. First of all, it is now commonly thought that few concepts submit to such characterizations (Ramsey 1992; Machery 2009, Ch. 3). Whereas some concepts in physics may have such definitions (though even this is not clear; see Wilson 2006 for extended discussion), few important theoretical concepts in the biological or behavioral sciences do. Even extremely well-characterized concepts like NATURAL SELECTION resist traditional definition (Godfrey-Smith 2009a) (see also §§ 3.4.1, 4.3.2). Though no proponents of any extant "mark of the cognitive" claim to offer such a definition, they each seem to envision such a definition as the ultimate goal. Thus Adams, Aizawa, and Garrison offer various necessary conditions on cognition, and Rowlands offers sufficient conditions.[6] What they offer, in effect, is

[6] Cameron Buckner's (2015) proposal is the exception, offering an attractive account on which COGNITION is a cluster concept in the style of Richard Boyd's homeostatic property clusters (Boyd 1991). However, Buckner's account amounts less to a traditional "mark of the cognitive" than to a field guide for inferring the existence of cognitive mechanisms based on clustered properties. Buckner anticipates that in the future we will discover a common kind of mechanism that exhibits the clustered properties, and that the description of this mechanism will offer the kind of classical clarity sought by other authors (2015, 324–325).

progress toward a traditional definition of COGNITION, i.e. hypotheses about its properties qua criterion. They do not offer ways of understanding a demarcation criterion that are alternatives to traditional definition.

Nevertheless, COGNITION poses a special difficulty apart from resistance to traditional definition, a difficulty that motivated the current interest in characterizing it: the fact of widespread expert disagreement (§ 2.2). The border wars are a manifestation of this disagreement. Of course, we may respond to expert disagreement by picking sides in the border wars. That is, instead of somehow accommodating disagreement in our characterization, we may explicate COGNITION in a manner that is agreeable to some and disagreeable to others. Call this conception of success *sectarian extensional adequacy*, and its pursuit the *sectarian strategy* of conceptual explication. All of the extant accounts are sectarian explications.[7] However, this strategy compromises on the second, epistemological benefit, that we should make explicit the new way that cognitive scientists think about their object of study. It is clear that COGNITION is presently an embattled scientific concept. To take sides in that battle is to make bets about the future—either about what we will discover about cognition, or about how cognitive scientists will decide to describe and explain cognitive phenomena. This is a worthwhile activity, but it has two drawbacks that impede its promise for yielding the benefits described earlier. First, it is a speculative project, about what cognitive science might look like in the future rather than about what cognitive science is like for us now. (Put another way: the sectarian strategy tries to predict the shape of cognitive science at the end of inquiry, rather than to faithfully describe our current practice. It is therefore not an anthropological approach of the kind that I pursue here; viz. § 1.3.) The sectarian strategy thus has limited practical application, which detracts from its ability to achieve the first, boundary-policing benefit, which is its primary aim. We may speculate that certain research will be unfruitful in the end, but if we eliminate the research on that basis then we eliminate our ability to correct our

[7] Even Buckner, whose goal is to demarcate cognition from association rather than to take a side in the border wars per se, takes his account to have alarming revisionary consequences, e.g. that cases of associative learning, such as taste aversion, are not cognitive phenomena (2015, 315). This consequence may be appropriate for highfalutin cognition, but is alarming for inclusive cognition.

speculation. This is a form of bad faith unless we are highly confident in our speculation, and the fact of widespread disagreement should prevent us from being so confident. The second drawback of a speculative strategy is that by speculating it disguises matters on which cognitive science is still ignorant, thus limiting its ability to achieve the fourth, metaphilosophical benefit. By downplaying rather than acknowledging our current ignorance, we would undermine the value of an explication of COGNITION as a secure resource for appeal regarding other philosophical projects.

A mostly-unexplored alternative to the sectarian strategy is to explain the disagreements of the border wars, rather than to take sides. For example, one might explain the fact of widespread expert disagreement by appealing to a conceptual structure that comes in degrees, or has sub-kinds, or something. The simplest traditional metaconceptual tools for disentangling conceptual complexity are appeal to vagueness and ambiguity, but it is not clear that these tools help much. First, COGNITION is not vague or graded in the manner of CLOUDY or SUNNY. Degrees of cloudiness can be compared according to a measure, and in a given context a standard can be set for being sufficiently cloudy to be worth calling "cloudy." But the disputes of the border wars do not seem to be about where to set the standard for being sufficiently cognitive. For example, proponents of plant cognition do not claim that plants are *slightly* cognitive, or *sort of* cognitive according to some measure, and that our standards of sufficient cognitiveness should be lowered to include them. Rather, they suggest that plant processes meet a criterion for cognition that, while less demanding than a criterion based on human performances, is nevertheless thoroughly apt for cognitive science (Calvo Garzón 2007). Likewise, proponents of extended cognition do not claim that a notebook can be a little cognitive (Clark 2010a). Rather, they argue that a notebook can be part of a cognitive system based on the same criteria that justify thinking of neurons as parts of cognitive systems. Advocates for these non-classical theoretical perspectives argue that we have learned that cognition is not the kind of thing we thought it was, and that it is the kind of thing that is relevant to the study of plants, microbes, and certain forms of tool use. There may plausibly be some metaconceptual gradation applicable to the use of COGNITION; for example it may be describable as a prototype concept (Rosch and Mervis 1975) like MAMMAL that applies more

comfortably to cats and rodents than to whales or monotremes. However, this would be gradation with respect to how comfortable we are in COGNITION-ascriptions, not gradation with respect to how cognitive a state or process is. Whales are mammals no less than squirrels are, and likewise (according to some) the processes that explain microbe taxis are cognitive no less than the processes that explain human language proficiency.

Similarly, "cognition" is not simply ambiguous in the fashion of "bank" and "bank." It is a natural instinct of the analytic philosopher to notice multiple standards for applying a word, and to distinguish between multiple senses in which the word might be used. Thus we might suppose that the enactivist means COGNITIVE_1 by "cognitive," and that plants have cognitive_1 processes, but that the traditional representationalist means COGNITIVE_2 by "cognitive," and that while human minds have cognitive_2 processes, plants do not. The project is then to spell out what the difference is between cognition_1 and cognition_2. It would neatly explain the disagreement of the border wars if this were the case: classical and anti-classical cognitive scientists simply have different concepts invoked by homophonous words. There is, I think, something to this, but it is a mistake to think that mere bank/bank ambiguity is the proper diagnosis. If "cognition" is merely ambiguous and refers to distinct and unrelated phenomena, then the border wars would not only be explained away but revealed to be merely verbal disputes (q.v. § 1.2.3). Extended cognition theorists and classical theorists would merely have distinct objects of study, objects which might be related somehow but which are not species of a common theoretical kind cognition. While this embarrassing scenario could conceivably obtain, a methodological commitment to interpretive charity demands we entertain alternative understandings of the practice before we reject it as misguided. And there are less disastrous alternatives in the vicinity: namely, that the scientific concept of cognition exhibits some more complex form of polysemy.

The various ways that cognitive scientists invoke the word "cognitive" may have different standards of application, but they are not unrelated. Furthermore, cognitive scientists are engaged in theoretical disputes about which conception of cognition is *best*. Consider that value theorists might engage in disputes over how to understand JUSTICE or AGENCY, and feminists might

disagree about how we should understand gender terms like WOMAN;[8] it is inappropriate to observe that various parties to these disputes mean different things by their words and leave it at that. Matters of social justice depend on understanding these concepts correctly, and the disputes are not mere confusions but precisely disputes about how one *should* use an expression or categorize phenomena, supposing of course that the expression has a regulatory function concerning how to think about politics, action, social justice, &c. The border wars may not be a matter of social justice, but they are a matter of scientific importance, and should be seen inter alia as disputes over what we *should* mean by "cognitive," supposing it has consequences regarding the demarcation of the cognitive sciences and makes explicit some scientific gains of recent decades. Put another way, the cognition border wars are an instance of what is sometimes lately called "conceptual ethics" (Burgess and Plunkett 2013a, 2013b), or what Sally Haslanger calls an "analytical approach" (2000).

2.5 ECUMENICAL EXTENSIONAL ADEQUACY

In the previous section I discussed some unsatisfactory ways of responding to the fact of widespread expert disagreement. The dominant approach is the sectarian strategy, which entails taking sides in the border wars and embracing revisionary consequences for cognitive scientific practice. This strategy is at best a straightforward means to the boundary-policing function, but I argued that it is an imperfect means to that end, and that it is also a poor means of accomplishing the epistemological the metaphilosophical goals. An alternative strategy is to explain away expert disagreements through appeal to vagueness or ambiguity, but I argued that these approaches are also unsatisfactory. In this section I shall describe another possible strategy: that of describing the character of disagreement exhibited in the border wars through an explication that countenances

[8] Bihui Li and J.S. Hancox offer something like an ecumenical treatment of WOMAN (Li and Hancox MS).

polysemy. I call this conception of success, accounting for rather than correcting the conflicting ascriptive practices of scientists, *ecumenical extensional adequacy*.

Traditional conceptual explication aims inter alia to render the extensions of concepts more precise than in actual judgments (§ 1.4.2). This is also what is demanded by sectarian adequacy; sectarian explications embrace revisionary consequences and serve to express, vindicate, or undermine the theoretical assumptions of various research programs. By contrast, in order to be ecumenically extensionally adequate an explication must instead reproduce the patterns of fuzziness and disagreement exhibited in actual judgment, even when such judgments are inconsistent as in cases of recalcitrant disagreement. [9] If philosophers are interested in characterizing the concept of cognition that lives in the practices of contemporary scientists, then the disagreements of the border wars are data in need of accommodation. A successful explication of COGNITION, if it is to be true to these data, must not draw a bright line that legislates which of these items[10] are cognitive and which are not. Rather, it must have an extension that is fuzzy, distinguishing which phenomena are clear cases of cognition, which are clearly not cases of cognition, and which are controversial.[11] Such an ecumenical characterization would require that membership in the extension of COGNITION not be all-or-nothing, but a property that can vary somehow without being merely graded in the manner of CLOUDY.

---

[9] As Adam Marushak says regarding contextualist approaches in philosophy of language (and echoing Lewis, quoted above, p. 18):"If the house is going to shake, you want the foundations to sway, too" (Marushak, personal communication; cf. Lewis 1973, 92).

[10] I use the term "item" to denote something of arbitrary ontological category—e.g. entities, events, properties, activities, processes, &c. It is a useful locution when talking of explications in the abstract, but a particular explication will usually apply to a restricted category of items, e.g. events, persons. See Andy Clark (2010a) on the infelicity of talking of "cognitive items" rather than more specifically of cognitive systems, cognitive processes, parts of cognitive systems, &c.

[11] A scholar of journalism or political science might be reminded here of Daniel C. Hallin's characterization of objectivity in media coverage (1986). Hallin suggests that a journalist's claim may fall into either of the "sphere of consensus," the "sphere of legitimate controversy," or the "sphere of deviance." These spheres describe the boundaries between claims that may be taken for granted by journalists, those that call for epistemic distancing or "balancing" evidence, and those that are generally considered unworthy of serious attention. My suggestion is that, in ecumenically characterizing cognition, we make similar distinctions regarding membership in the extension of COGNITION.

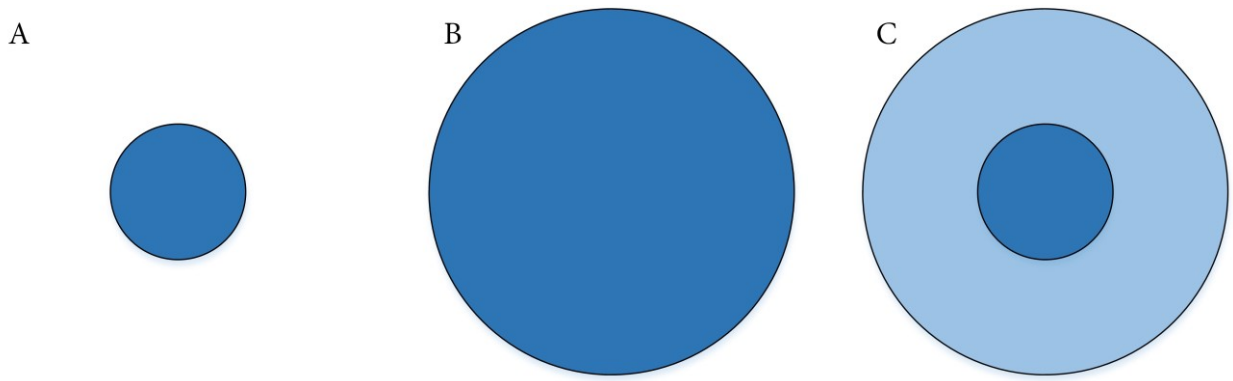

Figure 2.1: Venn diagrams representing classical and ecumenical extensions. A: A conservative classical extension. B: A liberal classical extension. C: An ecumenical extension.

Plausibly, this will be accomplished by being explicit about which commitments cognitive scientists seem to share, and which commitments are the objects of ongoing dispute (Ch. 3, 4). For example, there is ongoing dispute over how to understand the notion of REPRESENTATION as it features in cognitive science, and many extant disagreements about the nature and extension of cognition are cashed out as disagreements about the nature and extension of natural representation. Thus, Francisco Calvo Garzón argues that plants have cognitive processes because they represent their environment (2007), whereas Ramsey argues that plants do not have cognitive processes because nothing plants do is worth calling "representation" (2007). An ecumenically adequate explication of COGNITION may invoke the notion of REPRESENTATION without answering such questions, thus elucidating the commitments that are common and divergent between Calvo Garzón and Ramsey (though Ramsey [forthcoming] argues that an explication of COGNITION should not appeal to the notion of REPRESENTATION, and I agree; see § 6.3).

We might model this state of affairs with a non-classical extension. A classical extension, represented by a classical set, includes some items as members and excludes all others without differentiating any grades or variations in its members. A non-classical extension may be represented by means of a non-classical set (e.g. a fuzzy set), whose members have degrees of membership (Figure 2.1). In Figure 2.1 C, an ecumenical set with two grades of membership is

represented by two concentric circles. The members within the inner circle are paradigmatically cognitive items, such as neural memory processes and visual processes. The outer circle also contains those items, along with controversially cognitive items such as extended cognitive processes, plant cognitive processes, and robot cognitive processes. Thus the controversial cases are represented by the set of members of the outer circle that are not also members of the inner circle. So an ecumenical characterization requires a more complex structure than a sectarian characterization that appeals only the devices of necessary or sufficient conditions (or expressions whose extension is modeled by single classical sets), but the cost in theoretical and expository complexity could be made up for in other advantages, e.g. by modeling conflicting expert judgments, or making explicit what scientists have learned about cognition, or serving as a tool for explaining cognitive science to non-specialists, or serving as a non-sectarian resource for appeal in related philosophical arguments.

Though the ecumenical approach I advocate here is underexplored with respect to the problem of cognition, there are several "off-the-shelf" models for explications of other concepts that produce the right sort of graded extensions. For example, Cameron Buckner (2015) and William Ramsey (forthcoming) have suggested that COGNITION might be fruitfully modeled as a cluster concept. Prototype or exemplar-based characterizations of concepts (Rosch and Mervis 1975) also provide resources for explaining variable judgments, as do characterizations like Griffiths, Machery and Linquist's (2009) account of the lay concept of INNATENESS in terms of additive conditions. However, these models fall short of the desiderata of ecumenical explication insofar as they cannot be understood as specifying which commitments are shared and which are at issue in the border wars (§ 2.3.2). Perhaps more promising for this purpose are Daniel Dennett's intentional stance (Dennett 1987) and Stellan Ohlsson's "Darwinian explanatory strategy" (Ohlsson 1993), which provide conceptual characterizations in terms of abstract relations relativized to explanatory goals. These accounts produce non-classical extensions, where membership in the extension is indexed to the explanatory goals of an inquirer. John MacFarlane's (2014) assessment-sensitive semantics allow for differential ascriptions of epistemic modal claims and evaluative terms

like BRAVE and TASTY indexed to "contexts of assessment," so that licit ascriptions of predicates can vary with evidence or preferences. Explications on the model of Dennett, Ohlsson, or MacFarlane are promising, for they each involve an abstract characterization of shared commitments—belief-desire explanation, evolution through selection, and assessment-sensitive rules—that have consequences that vary according to other commitments or attitudes. I will develop a version of this strategy in subsequent chapters.

The primary virtue of ecumenical adequacy is that it has the right structure to achieve the second, epistemological benefit: that it should make explicit what the scientific community has learned about cognition since the 1980s. Given the ongoing disagreements of the border wars, and a charitable presumption that the counterintuitive beliefs of so many scientists should at least be taken seriously, it is desirable that an explication of COGNITION account for the disagreement. In particular, it would be valuable to acknowledge, rather than obscure, the limitations of our current knowledge. That such an account could serve the epistemological function of a solution to the problem of cognition contributes to its potential to serve the public-relations and metaphilosophical functions, as well. Insofar as it is also compact it would enable clearer expression of cognitive scientific theories and practices to non-specialists, and it would serve as a reliable resource for appeal in philosophical arguments. After introducing an ecumenical explication of COGNITION (Ch. 5), I will assess its potential to serve these functions (§ 7.3). The main drawback of an ecumenical characterization, of course, is that it would not clearly serve the first, boundary-policing function: that of providing clear answers to the open questions left by the border wars. After all, an ecumenical account is precisely meant *not* to settle the boundary disputes which were the original theatre of deployment for the mark of the cognitive, but rather to explain how cognitive scientists can disagree so much about a concept so fundamental to their inquiry.

Whether this is a tolerable drawback depends, of course, on one's goals. If, like most contributors to the current literature on the "mark of the cognitive," boundary-policing is a primary goal, then ecumenical explication may seem a non-starter. However, the ecumenical approach may be offer indirect promise for boundary-policing. An ecumenical explication aimed

at serving the epistemological function should make clear what implicit commitments are shared by cognitive scientists on various sides of the border wars, and which are not. In doing so, an ecumenical explication isolates matters of agreement and disagreement, thus suggesting different terms in which to more fruitfully continue the border wars, while allowing sectarians to see each other as, if not all correct, then intelligibly disagreeing. That is, if through an ecumenical explication of COGNITION we achieve new clarity about the structure of this embattled concept, we may hit upon new ways for philosophers and scientists to settle open questions about the nature and extension of cognition without talking past one another (§ 4.4). If so, an ecumenical characterization would provide normative guidance about how to practice cognitive science, after all. If this possibility were made good then ecumenical explication would be strictly superior to a sectarian explication not only with respect to the epistemological, public relations, and metaphilosophical benefits mentioned; it would also be a less contentious way to promote convergence between rival conceptions of COGNITION, thus opening new avenues for settling the boundaries of cognition, which is the main concern of the extant contributors to the "mark of the cognitive" literature (§ 7.4). Nevertheless, even if the ecumenical approach were to prove ineffective for policing the boundaries of cognitive science, it would be worth pursuing for its other potential benefits.

2.6 SUMMARY

It is a historical accident that the current century's treatment of the problem of cognition has aimed primarily to answer the open questions of the border wars, rather than to serve other functions. The open questions of the border wars motivate further questions about the development of the scientific concept of cognition (§ 2.3.2), about the implicit conceptual knowledge acquired through training in cognitive science (§ 2.3.3), and about the place of minds in the scientific image, as well about other matters that are of broad philosophical interest (§ 2.3.3). The dominant approach to the

problem of cognition is one that aims for clarity about the open questions at the expense of these further questions. I have recommended different strategy toward explicating COGNITION in which we aim to make the conceptual change explicit, even at the expense of clear answers to the open questions. Because of the fact of widespread expert disagreement, this ecumenical strategy requires the accommodation of many apparently inconsistent research perspectives, presumably by being clear about which commitments are shared by experts in the science and which commitments are controversial (§ 2.5; see also § 5.5 below). It has been my contention in this chapter that an ecumenical explication of the concept of COGNITION, though unexplored in the contemporary literature, has substantial promise at achieving most of what we would hope for in a solution to the problem of cognition. In the next two chapters, I will describe logical tools that can achieve ecumenical adequacy, before describing my solution to the problem of cognition.

3.0 CONSTRUAL-BASED CONCEPTUAL EXPLICATION

3.1 OVERVIEW

In the last chapter I described the problem of cognition and motivated an ecumenical criterion of adequacy for conceptual explication. An ecumenically adequate explication accurately reproduces the variance in expert judgments rather than treating that variance as noise, or correcting it by taking sides in disputes among experts. In this chapter, I will describe a general strategy for achieving ecumenical adequacy: making an explication answerable to judgments rather than to first-order facts. If the aim of an explication is to capture patterns of judgment, rather than facts about the world independently of how anyone may think about it, then conceptual explication can be a useful tool for modeling recalcitrant conceptual disagreement of the sort exhibited in the border wars. In this chapter, I will discuss a distinctive form of judgment I call a "construal." In the next chapter I will develop a more analytically precise strategy for construal-based explication—parameterization—which will permit me to introduce an ecumenical explication of COGNITION. The focus in this chapter, however, is on the properties of construals and construal-based explications.

I will begin with an orienting detour through some familiar territory regarding conceptual explication (§ 3.2), highlighting some common methodological assumptions that I will tinker with in order to make space for construal-based explications (§ 3.3). I take this route in part to make my commitments regarding conceptual explication more explicit, and in part because I take the construal-based strategy to be quite general, with potential applications for many philosophical topics. I thus expect there to be value in describing construals and construal-based explication in

some generality. For the same reason, I will not focus on the case of COGNITION in the main discussion of the chapter. After describing the approach, I will discuss several potential objections to its application in scientific contexts (§ 3.4). In particular, I will discuss objections that construal-based explications are insufficiently objective for science, that they are problematically aprioristic, and that they are problematically anti-realist. I will conclude that construal-based explication is fully consistent with a thoroughgoing and reasonable scientific realism. In fact, it is a commitment to realism that motivates construal-based explication over a nearby alternative, merely ascriptional explication (q.v. § 3.3.4).

3.2 TRADITIONAL CONCEPTUAL EXPLICATION

The traditional form of a conceptual explication of *F* is an expression that can be substituted, at least for certain purposes, for a familiar expression denoting Fs. The substituting expression is sometimes taken to denote items that are identical to Fs, but at any rate the substituting expression is generally taken to articulate necessary, sufficient, or individually necessary and jointly sufficient conditions for being F.[12] One criterion of adequacy on such explications is that they be extensionally adequate. An extension is a category that includes all and only the referents of a word or concept, and is often understood as a set that includes all and only these referents as elements. The notion of an "intension" is occasionally introduced to contrast extensions in the actual world to extensions in other (or all) possible worlds (e.g. Carnap 1947). Since that is an unusual use of the word "intension," however, and since the present discussion does not turn on contrasts between the actual and other possible worlds, I will simply use the notion of an extension to refer to referents

[12] A fuller account of conceptual explication might treat the differences between explicanda that are expressed as nouns, modifiers, verb phrases, &c., rather than assimilating them all to predicates. There are some niceties, e.g. that typically only nominal explicanda are analyzed in terms of identities, whereas any explicandum can be explicated in terms of necessary and sufficient conditions (to varying degrees of adequacy). The complexities involved in being precise about the differences are beyond the scope of the present discussion, and I assume here that little hangs on attention to the differences between explicating COGNITION vs. COGNITIVE. (But see Figdor 2014 for a dissenting opinion concerning nouns vs. verbs.)

in general. For a substituting expression in an explication of *F* to be extensionally adequate, it must pick out all and only the things that are Fs in actuality or in possible worlds.

For example, consider a familiar conceptual explication:

CLASSIC BACHELOR:

A bachelor is an unmarried adult man.

The claim of this explication is that BACHELOR can be explicated as *unmarried adult man*, i.e. that being unmarried, being an adult, and being a man are individually necessary and jointly sufficient conditions of being a bachelor. By convention, the left-hand side features the *explicandum*, the concept to be explicated, and the right-hand side articulates the *explicans*, the substituting or explicating expression. Depending on the use to which the explication is put, it is sometimes also important that the explicans be expressed in a privileged vocabulary, e.g. natural scientific vocabulary, observational vocabulary, non-normative vocabulary, or what have you. An explication expressed purely in terms of such a privileged vocabulary is sometimes called a *reductive* explication. The above explication of BACHELOR might be reductive if the relevant privileged vocabulary is, for example, predicates whose instantiation can be gleaned from available census data, and if we are studying the demographics of bachelors using census data we may want an explication that is reductive in this way.

In order for CLASSIC BACHELOR to be extensionally adequate, it must be the case that all bachelors are unmarried adult men, and that all unmarried adult men are bachelors. The claim can be shown to be false if a counterexample is provided that satisfies one side of the explication (e.g. "is an unmarried adult man") but not the other (e.g. "is a bachelor"). For example, it may be alleged that the Pope is an unmarried adult man, but no bachelor. If that fact is accepted, then the explication is defective because it is not extensionally adequate; it identifies an extension (unmarried adult men) that does not coincide precisely with the extension of BACHELOR. Explications with counterexamples like this, that belong to the extension of the right-hand side but

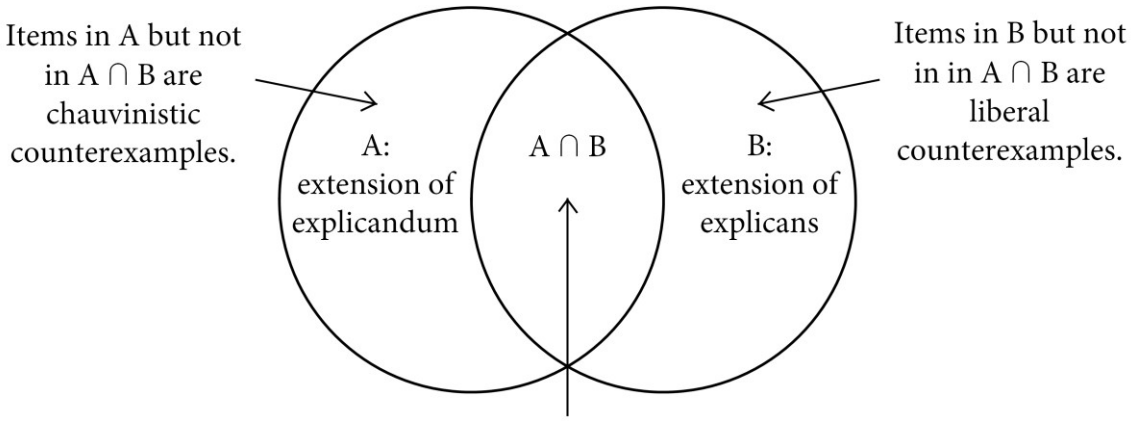

Figure 3.1: Venn diagram illustrating liberality and chauvinism for conceptual explications.

not the left-hand side, are sometimes called *liberal* because they articulate an extension that includes items that do not belong to the extension of the explicandum. By contrast, explications whose right-hand sides have extensions that fail to include items in the extension of the explicandum are called *chauvinistic* (Figure 3.1). An explication can be both liberal and chauvinistic. Note that liberality and chauvinism objections require that we know, independently of entertaining an explication, whether at least some items belong in the extension of the explicandum. Such explication-independent attitudes about whether a particular item belongs to the extension of an explicandum are sometimes called "intuitions." (I remain officially agnostic as to whether intuitions are a species of judgment, whether they have non-inferential etiology, and other mysteries. In the present context they merely denote whatever means we have for acquiring explication-independent criteria by which to judge the extensional adequacy of an explication.) Many speakers of English will have the intuition that the Pope is not a bachelor. I call this framework for conceptual explication the *traditional scheme*.

The gold standard for a conceptual explication of COGNITION, sometimes called the "mark of the cognitive," is a conceptual explication like this. However, no extant proposal claims to have achieved this; rather, most claim only to approximate an adequate explication, usually by identifying necessary conditions or sufficient conditions but not both. Adams and Aizawa (2001,

2008) claim to have identified two individually necessary conditions, but no sufficient conditions. Adams and Rebecca Garrison (2013) claim to have identified a different necessary condition. Mark Rowlands (2009, 2010) claims to have identified four jointly sufficient conditions, but no necessary conditions (Buckner claims to have found neither). Articulating only necessary or only jointly sufficient conditions selectively protects these partial accounts from liberality and chauvinism objections, respectively. However, such partial explications, if they are true, have the consequence that some items can be definitely declared non-cognitive (if the conditions supplied are individually necessary, and not satisfied by the items in question) or declared cognitive (if the conditions supplied are jointly sufficient, and satisfied by those items). Adams and Aizawa introduced the expression "mark of the cognitive" and the project of identifying it to do just this: to argue that particular items, e.g. plants or microbes or elements of the environment external to the body, are or are not cognitive systems or parts thereof. That is, while I characterize the problem of cognition as a puzzle about how to explicate a scientific concept (§ 2.5), it has historically been a setting for vindicating intuitions in intramural theoretical disputes. This is not to say that intuitions about cases are used primarily as premises in arguments about the mark of the cognitive—the arguments are not that homogeneous—but extensional adequacy is nevertheless taken to be a signal virtue in a proposed explication, and explications are in turn appealed to for their implications concerning the extension of COGNITION. Since the primary dialectic use to which putative "marks of the cognitive" are put is to demarcate cognitive from non-cognitive items, it is fair to say that extensional adequacy has been treated as the most important virtue of such explications (even though ironically the main question at stake is "which extension is it that an account must be adequate to?").

Accountability on the traditional scheme is possible when we have independent knowledge of the extension of the explicandum, whether that knowledge is accomplished through linguistic competence, non-inferential judgment, acquaintance with truisms, empirical investigation, or other means. However, in the cognition border wars, whether an item belongs to the extension of COGNITION is what is at issue. While empirical considerations play a role in justifying judgments,

the border wars concern inter alia how the concept COGNITION should be used, not whether an antecedently-understood predicate applies in particular cases (i.e. they are not factual disagreements, q.v. § 1.2.3). It is a fact, apparent from even a cursory examination of the theoretical literature from the mid-1980s to the present, that the practice of ascribing cognitive states and processes is controversial among experts. I argued in the previous chapter that it is appropriate in the case of COGNITION to aspire to ecumenical adequacy that reproduces this variance in expert judgment. It is well and good to undertake the project of achieving ecumenical extensional adequacy, in which membership in the extension of the explicandum varies without being merely graded, but it is not immediately clear how this can be accomplished with an explicans composed of necessary and sufficient conditions. I turn now to a recommendation: that the explicans feature construal-sensitive terms.

3.3 CONSTRUALS AND CONCEPTUAL EXPLICATION

I now describe a variation on traditional conceptual explication that can achieve the kind of structure demanded in ecumenical adequacy. The general strategy I adopt is to articulate in the explicans the conditions of ascription for (say) F-hood, rather than to articulate the conditions for being an F. Put another way, I seek an explication that describes what it is to think about something as F, rather than what it is for something to be F. As I will put it here, I am interested in judgment-based explications rather than object-based explications. In many cases these two strategies may not differ appreciably. If I know that my friend is an unmarried adult man, I may judge that he is a bachelor precisely because that is simply what it is to be a bachelor (assuming that CLASSIC BACHELOR is adequate). However, the judgment-based and object-based strategies diverge at least for a peculiar class of judgments that I refer to as "construals."[13]

[13] I borrow the language of "construal" from Robert C. Roberts (1988), though I use the term in a more expansive way so as to refer to judgments that engender interpersonal disagreement in addition to those that engender inconsistent semantic evaluations by the same person on a single occasion.

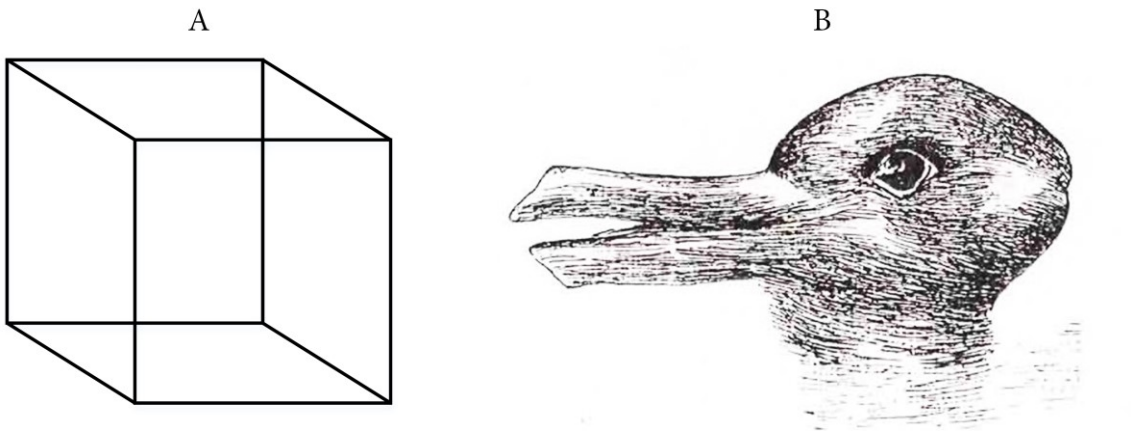

Figure 3.2: Ambiguous figures. A: The Necker cube. B: The duck-rabbit (duck-rabbit image from Jastrow 1899).

I shall begin by introducing construals. Others have written on the psychological character of construals (e.g. Wittgenstein 1953; Roberts 1988), but my main interest here is their semantic character. I choose not to commit myself to whether construals should in general be understood in terms of truth or warranted assertibility or some other semantic good-making property, but their semantic properties vary in ways that more straightforward judgments do not. After describing some semantic properties of construals (§ 3.3.1) I will describe how they can be incorporated into conceptual explications (§ 3.3.2). I conclude this section by discussing the distinctive way that one evaluates counterexamples to construal-based explications (§ 3.3.3), and by discussing the reasons one might prefer construal-based explications to a nearby alternative, merely ascriptional explications (§ 3.3.4).

3.3.1 The semantics of construals

Construals are a kind of judgment whose semantic properties (e.g. truth) vary either inter- or intra-subjectively but independent of the situation, i.e. independent of the state of affairs that the judgment is about. I borrow the notion of a "situation" from situation semantics (Barwise 1981;

Barwise and Perry 1983) because it involves a convenient and familiar departure from the notion of a possible world or a the notion of a world-time pair. Namely, a situation is a partial possible world such that some states of affairs are fixed by the situation but other states of affairs are left open (i.e. whereas at a particular possible world all facts about that world are true or false, a situation is consistent with several possible worlds that differ with respect to facts outside of the situation). However, I do not mean to adopt a strong commitment to situation semantics; one might substitute states of affairs for situations, as well as structured propositions or in some cases Davidsonian events (Kratzer 2016). I am not presently concerned with formal natural language semantics, but rather only with the informal description of an unusual semantic phenomenon and its applications in scientific conceptual explication. The language of situations is hereafter adopted for instrumental reasons.

The construals most familiar to philosophers are probably those described as "seeing as" ("Aspektsehen" or "Sehen-als"; cf. Wittgenstein 1953, pp. 165ff). The phenomenon is most easily identifiable in ambiguous figures. For example, the duck-rabbit figure made popular by Wittgenstein can be seen either as a duck or as a rabbit. The Necker cube can be seen as if it were viewed either from above or from below (Figure 3.2). The same phenomenon is involved in conceiving of an object or a situation in terms of something else. Robert C. Roberts writes:

> One "sees an aspect" of a face by construing it "in terms of" another face, in something like an act of imagination. […] Construing seems to involve dwelling on or attending to, or at a minimum holding onto, some aspect, for example, the duckiness of the duck-rabbit or one's triumph. It seems to mean bringing some perceived paradigm, or some concept or image or thought, *to bear*. (1988, 187)

Thus one may construe a friend's face as like her father's, or a political defeat as crushing, or a Tory government as vicious. The semantic properties of a construal may change according to mental states and commitments of the construer independently of changes in the referent of the construal.

The figures of the duck-rabbit and the Necker cube do not change, but the truth or assertibility of a viewer's construals of those figures may change according to her attention or other factors. My friend's face may not change, but I may alternately construe her face as like her father's or as unlike her father's. The effects of a Tory government may be agreed upon, but one may construe them as vicious or as principled. Furthermore, since there are judgments whose semantic properties can vary this way intersubjectively as well as intra-subjectively, I shall not distinguish here between those kinds of case. Thus while I may construe an election as a crushing defeat, my neighbor may construe it as a terrific success, or if I am ambivalent about party politics I may construe it both ways over the course of my morning commute.

The kind of semantic variation particular to construals is not the kind that arises from indexicals or some other familiar contextual shifts. For example, my thought that "I am the First Minister of Scotland" and Nicola Sturgeon's thought that "I am the First Minister of Scotland" are sometimes said to have the same "character" (Kaplan 1989). A thought with this character is true when it is Ms. Sturgeon's and the time is now, whereas when such a thought is mine it is false at every time in the actual world. However, this indexical variation in semantic value is distinct from the variation exhibited by construals. The semantic properties of the proposition "I am the First Minister of Scotland" change only when the situation is varied, i.e. according to changes in the state of affairs that the proposition is about. The semantic variability characteristic of construals corresponds not to variations of the situation, but to variations in the mental states or commitments of the construer, independently of her situation. The obvious complication is that one may make judgments about one's own mental states or commitments, so that one's own commitments become part of the situation. For example, I may judge truly that I am sleepy because I have become sleepy, and that judgment will cease to be true when I have ceased to be sleepy. This judgment is not ordinarily a construal[14] because it is not a judgment whose truth-values change independently of the situation; if its truth value changes because I have ceased to be sleepy then it changes only

[14] At least, the judgment that "I am sleepy" is not a construal when understood in the simplest terms, where "I" and "am sleepy" are not interpreted so as to be construal-sensitive. The notion of a construal is really a pragmatic one, as I will discuss shortly.

because of a change in the relevant situation. Where my judgment concerns my mental states or commitments any change in those states or commitments are changes in the state of affairs that my judgment is about. On the other hand, Yasmin may construe herself as generally morose or not depending on her present mood or the salience of various episodes in her life. Since the relevant situation—namely the general historical facts concerning her demeanor—is not altered by her present mood or by which episodes in her life are presently salient, such a judgment is a construal.

The essential point about construals is that their semantic properties can vary licitly though the situations they describe do not. Put another way, the semantic properties of a construal depend on a *context of assessment* in addition to depending on facts about their objects.[15] When it comes to construals, the important features of a context of assessment are just the concurrent mental states of the judge, e.g. what she is attending to, her background beliefs, her evaluative commitments, and sometimes her occurrent thoughts. The ascription of certain predicates, such as TASTY or GREENISH, is typically a construal, but epistemic modal claims are not. Insofar as the ascription of a predicate or use of a concept makes a containing judgment into a construal, I will call the predicate or concept *construal-sensitive*. Likewise, I will call the exemplification of such predicates or concepts (i.e. tastiness or greenishness) construal-sensitive phenomena. The referents of construal-sensitive noun phrases are construal-sensitive phenomena.

3.3.2 Construal-based conceptual explication

What I call a *construal-based conceptual explication* is an explication that exploits construal-sensitive terms on the right-hand side. In particular, whether the right-hand side of a construal-

[15] I borrow the term "context of assessment" from John MacFarlane. The kind of variation described here resembles some of the variation that John MacFarlane aims to account for with his assessment-sensitive semantics, e.g. that exhibited in matters of taste (2014), though MacFarlane's semantics are in fact more general than those required for construals. For example, MacFarlane extends his view to account for epistemic modal judgments and judgments of future contingents, and these judgments do not seem to be (or to necessarily involve) construals. The relativization to context here does not distinguish between a relativist and a contextualist semantics, since no rule has been provided for assessing truth-ascriptions concerning construals, and therefore retraction and rejection behavior is underdetermined (see MacFarlane 2014; Ch. 5).

based explication is instantiated in a particular case depends on a construal by some judge. This can be contrasted with an *object-based conceptual explication*, which features no construal-sensitive terms in the explicans, or is used in such a way that the dependence on finer-grained criteria than worlds is suppressed. CLASSIC BACHELOR (p. 52 above) is an object-based explication. The right-hand side features only predicates that either apply to an item or do not. Judgments that someone is a man, is an adult, or is unmarried are not typically semantically variable in the sense peculiar to construals.

I say only "typically," however, since intra-subjectively variable judgments, i.e. seeing something as something, are easy to come by. For example, one may construe someone as adult-like in certain respects, or stereotypically man-like or unmarried in certain respects, whether or not he or she is in fact a man or has reached the age of majority or is legally (or socially or as a matter of religious practice) unmarried. Imagine Brenda, who is married but whose spouse frequently travels for work so that Brenda is frequently to be found out with friends or at home, surrounded by pizza boxes, playing video games. One may be tempted to say of a woman like Brenda that she is "a real bachelor," or "totally a bachelor."[16] One response to such temptations is to deny that they are literally true. Brenda is not literally an unmarried adult man, and she is therefore not literally a bachelor. If such appeals to literal ascription are unsatisfactory, however, then almost any term will be construal-sensitive in some circumstances. Nevertheless, many explications are meant to be used in such a way that the construal-sensitivity of their terms is suppressed. The appeal to CLASSIC BACHELOR that justifies the utterance that Brenda is "a real bachelor" is not a use that suppresses the construal-sensitivity of its terms. On the other hand, the reductive use of CLASSIC BACHELOR, on which one of its virtues is that the terms on the right-hand side can be gleaned from census data, is a use that does suppress the construal-sensitivity of its terms. Since Brenda's bachelor-like qualities are not discernible from census records, she does not satisfy the conditions of this reductive use of CLASSIC BACHELOR. As long as either the

[16] Edouard Machery makes a similar point about about ascriptions of "bachelor" (Machery MS, Ch. 6 § 3.2).

appeal to literal ascriptions is satisfactory or there are construal-suppressing explications, there are object-based explications.[17]

A construal-based explication, on the other hand, is one on which the peculiar semantics of construals is exploited in order to account for variability in the ascription of an explicandum. (Notably, some general inferential properties of an explication change depending on whether the explicandum term and the construals in the explicans appear in intensional rather than extensional contexts. I will return to this complication in § 3.3.4.) To take a toy example, consider the following explication of the concept BRAVE:

BRAVE BEHAVIOR:

An action ϕ is brave iff ϕ is dangerous for its agent, and ϕ is admirable.

Assume we have a suitably precise understanding of what it is for an action to be dangerous for its agent, but allow judgments of whether an action is admirable to vary according to the evaluative commitments of various judges. In other words, let judgments of the form *φ is admirable* be treated as construals. Now the right-hand side of BRAVE BEHAVIOR features a construal-sensitive predicate, *admirable*, whose semantic value may vary independently of a situation. Whether an action is admirable depends on whether the relevant judge construes it as admirable. Since the semantic value of the right-hand side of the explication depends on a construal by a judge, BRAVE BEHAVIOR is a construal-based explication. I will illustrate with an example. Suppose that Marilyn likes to go skydiving, and that skydiving is dangerous for the skydiver. Chandana thinks that Marilyn is brave for skydiving, and according to BRAVE BEHAVIOR we can infer that Chandana thinks that skydiving is admirable. Kasson, on the other hand, thinks that there is nothing admirable about skydiving for fun. He therefore does not think Marilyn is brave for skydiving; he thinks she is merely foolhardy. Let us also suppose that Chandana and Kasson are epistemic peers; it is not the case that (as far as we are concerned) one of them is a better judge of

[17] Thanks to Jennifer Corns for pressing me on this point.

admirableness than the other. They merely have different standards of assessment for admirableness. BRAVE BEHAVIOR allows us to account for Chandana's and Kasson's differential judgments about whether Marilyn's behavior is brave by appealing to a single explication that illuminates which commitments about bravery they share, and which commitments they do not share.

3.3.3 Counterexamples and construal-based explications

The goal of a construal-based explication is to articulate the conditions of ascription for a concept according to competent users—be they right or wrong—rather than the conditions of instantiation according to some more objective standard. Because of this, construals are sensitive to counterexamples in a distinctive way. They go wrong not simply by giving incorrect verdicts about the combinations of properties of items, but by giving incorrect verdicts about the combinations of judgments relative to a context of assessment. For example, consider CLASSIC BACHELOR again. Above I noted that one might object to this explication on the grounds that the Pope is an unmarried man but no bachelor. The explication is therefore liberal, as illustrated by the counterexample of the Pope. By contrast, consider an objection to BRAVE BEHAVIOR. Chandana approved of Marilyn's skydiving. Suppose that Chandana thinks that Marilyn is brave for skydiving because she cannot imagine herself skydiving and having fun—leaping from a plane, enjoying the thrill of the fall, and not terrified by the thought that that her parachute could malfunction. She thinks that Marilyn is liberated, confident, and made of wonderfully strong stuff. However, suppose that Chandana also believes that it is never praiseworthy to put oneself in mortal danger except in order to avert greater harm. Let us also suppose that Marilyn's hobby has no purpose other than Marilyn's amusement. Now, so described Chandana may vacillate between admiration and disapproval about Marilyn's hobby. Chandana might, then, object that BRAVE BEHAVIOR is chauvinistic because skydiving is a brave action, but not one that is both dangerous and admirable.

So described, this objection may be confused (it may not be, but only if it avoids the pitfall I am about to describe). As a construal-based explication, BRAVE BEHAVIOR articulates the conditions of being brave in terms of a judge's other judgments, i.e. relative to a context of assessment that specifies which actions are admirable ones. One's construals may vary because one's context of assessment changes. In the example above, Chandana construes skydiving variably as admirable or foolish depending on what features of the activity she attends to. What Chandana may not do is make a charge of chauvinism on the grounds that skydiving is not admirable from one context of assessment, and that skydiving is brave based on another context of assessment. If she does this she equivocates over contexts of assessment. Her two judgments cannot constitute a counterexample to BRAVE BEHAVIOR so long as they are relativized to different contexts of assessment. The same is true for interpersonal combinations of judgment; one cannot object to BRAVE BEHAVIOR on the grounds that skydiving brave according to Chandana and that it is not admirable according to Kasson. The mistake here is akin to the mistake involved in claiming the duck-rabbit's face is on the back of its head while equivocating over whether one sees it as a duck or as a rabbit.

What I am describing is a notable complication to the schemata for counterexampling described above in Section 2. An object-based explication makes claims of the following form:

OBJECT-SCHEMA:
(x) x is F iff x is G

where *F* stands for an explicandum and *G* stands for an explicans. The object-based schema is subject to counterexamples in the following ways, where *a* is a name denoting an item in the domain of quantification for *x* in OBJECT-SCHEMA:

OBJECT-LIBERALITY:
a is not F, and a is G.

OBJECT-CHAUVINISM:

a is F, and a is not G.

A construal-based explication, by contrast, conforms to the following general form:

CONSTRUAL-SCHEMA:

(x) x is F in c iff x is G in c.

Once again, *F* stands for the explicandum and G stands for the explicans. Here, however, *F* and *G* are construal-sensitive predicates. They are relativized to a context of assessment *c* in CONSTRUAL-SCHEMA because, as I described above, the semantic value of a construal is underdetermined by world-time pairs or situations; the context of assessment is required to fix the semantic value. This relativization was implicit in BRAVE BEHAVIOR, though it is made explicit here (and it will be explicit below in BRAVE-ASCRIPTION and WITCH-ASCRIPTION, but I prefer to keep this relativization implicit when possible; cf. § 3.3.4 below). In the careful consideration of counterexamples, relativization to a context of assessment must be made explicit and the context must be preserved between the left- and right-hand sides. Counterexamples to a construal-based explication must have the following form:

CONSTRUAL-LIBERALITY:

a is not construed as F in c and a is construed as G in c.

CONSTRUAL-CHAUVINISM :

a is construed as F in c and a is not construed as G in c.

Now reconsider the defective skydiving counterexample to BRAVE BEHAVIOR. Let the context of assessment be a standard of evaluation that fixes whether Ga and whether Fa. That is, whether

skydiving is brave and whether skydiving is dangerous and admirable, respectively. There need not be a unique, correct answer to the question of whether skydiving is brave; it is enough if it is considered brave relative to contexts of assessment where it is made to seem admirable, and not brave relative to contexts of assessment where it is not. There is a counterexample to BRAVE BEHAVIOR just in case there is a context of assessment *c* and an item *a* such that either CONSTRUAL-LIBERALITY or CONSTRUAL-CHAUVINISM is true.

One cannot object to CLASSIC BACHELOR by citing the facts that the Pope is an unmarried adult man and that Barack Obama is not a bachelor—the individuals (i.e. the Pope and Barack Obama) must be identical or OBJECT-LIBERALITY is not satisfied. Similarly, CONSTRUAL-CHAUVINISM is not satisfied unless one identifies a case in which the contexts of assessment are identical. Chandana can substantiate her objection only if she (or someone) judges both that skydiving is brave and that it is not admirable without equivocating over contexts of assessment. She might, for example, claim that being admirable is not a condition on bravery—that she thinks Kamikaze bombers were brave for facing death although they sacrificed themselves for a war that could not be won, and denies that they were admirable because she deplores the Empire of Japan's conduct in the war, and it is not admirable to promote imperialist or depraved regimes. (Though often, those who disapprove of suicide bombers call them "cowardly" rather than brave.[18] But whenever competent judges of bravery judge some action brave, if they simultaneously failed to find it admirable then the combination of judgments constitutes a valid counterexample to BRAVE BEHAVIOR.

3.3.4 Construal-based vs. ascriptional explication

A construal-based explication features at least one construal-sensitive term in the explicans, so that the satisfaction of the explicans varies according to a judge's context of assessment. In an ascriptional explication, by contrast, construal-sensitive terms appear only in intensional contexts.

[18] Thanks to Robert Brandom for drawing this example to my attention.

Consider that we might have accounted for Chandana's and Kasson's differential ascriptions by proposing a generalization like the following:

> BRAVE-ASCRIPTION:
> A judge j in c construes an action ϕ as brave iff in c, j judges that ϕ is dangerous for its agent and j construes ϕ as admirable.[19]

This explication also allows us to account for Kasson's and Chandana's different judgments about Marilyn's skydiving. However, it is not a construal-based explication. Since the construal-sensitive terms ("brave" and "admirable") are safely isolated in intensional contexts, it is not the case that either instantiation of the left-hand side or the right-hand side of the explication depends on construals by a user of the explication. The instantiation of either side depends only on facts about the judgments of the mentioned judge j. This is a merely ascriptional explication. However, ascriptional explication is not a less-weird version of construal-based explication. I distinguish construal-based explications like BRAVE BEHAVIOR from ascriptional explications like BRAVE-ASCRIPTION because they have distinct inferential properties.

First of all, BRAVE-ASCRIPTION has weaker ontological commitments. Commitment to BRAVE BEHAVIOR allows us as inquirers to conclude that there are brave actions so long as there are actions that are dangerous and admirable, whereas BRAVE-ASCRIPTION leaves open that there are no brave actions even if there are dangerous, admirable actions. This is because BRAVE-ASCRIPTION and other ascriptional explications permit us to generalize about how a concept is ascribed while endorsing an error theory about that concept. Second, ascriptional explications leave us with odd inferential gaps in cases of first-person reasoning. For example, if I think that skydiving is dangerous and admirable, BRAVE BEHAVIOR licenses my inference that skydiving is brave. BRAVE-ASCRIPTION licenses only my inference that I judge skydiving to be brave. Surely, there are circumstances in which we might prefer a construal-based explication like

[19] On the right-hand side, the scope of "in c" ranges over the remainder of the sentence.

BRAVE BEHAVIOR to a merely ascriptional explication like BRAVE-ASCRIPTION, as well as circumstances in which we would prefer the reverse, and circumstances in which we are indifferent. Consider the following pair of explications:

> BAD WITCH:
> A witch is a woman who is evil and possesses magical powers.
>
> WITCH-ASCRIPTION:
> (x) A judge j in context c believes that x is a witch iff in c, j believes that x is a woman and j believes that x possesses magical powers and j construes x as evil.

As with admirableness above, we might suppose that judges may disagree without fault regarding evilness-ascription. If we are interested only in practices of concept-ascription, and not in ontological commitment regarding witches, we may be indifferent between BAD WITCH and WITCH-ASCRIPTION, or we may endorse both. If we believe that there are no witches, then we may endorse WITCH-ASCRIPTION for a restricted domain of judges—those who believe in witches—while withholding commitment from BAD WITCH. Finally, if we are committed to the existence of witches and interested in making use of the extra inferential power of construal-based explications, then we should endorse BAD WITCH. For example, if we want to decide for ourselves whether someone is a witch, we can appeal to BAD WITCH in order to justify such judgments whereas WITCH-ASCRIPTION does not allow us to infer that anyone is a witch. It allows us at best to infer that we judge someone to be a witch. If, on the other hand, we believe in witches but also that most people have false beliefs about the attributes of witches, we might commit to BAD WITCH and withhold commitment from WITCH-ASCRIPTION (though we might endorse a deontic modal version of WITCH-ASCRIPTION, i.e. one in which "should believe" replaces "believes").

Nevertheless, BRAVE-ASCRIPTION and WITCH-ASCRIPTION may generally seem preferable to some on the ground that they are less metaphysically "queer" (in the sense of Mackie 1977). While ascriptional explications mention construals, they do not use them. Therefore, we may endorse an ascriptional explication instead of a cognate construal-based explication without admitting construals into our metalanguage. Ascriptional explications are therefore object-based explications, in which any construal-sensitive predicates are safely contained by intensional contexts where they do not complicate our—the inquirers'—inferences. Or, some moderately sympathetic readers might be tempted to say: "This is all well and good for concepts like BRAVERY or EVIL or GREENISHNESS, but COGNITION is a scientific concept, and scientific concepts should carve nature at the joints; they are not so wishy-washy." In the next section, I will argue that the practice-oriented philosopher of science should not be anxious about construal-based accounts.

3.4 HOW TO STOP WORRYING AND LOVE CONSTRUAL-BASED EXPLICATION

Though construal-based explication has not been described in quite this way before, it is not a radical, new idea. It has been proposed before (under different names). Dennett's (1987) intentional stance, for example, might be understood as a construal-based account of minds. What it is to have mental states, on Dennett's view, is for one's behavior to be explainable by reference to certain kinds of explanatory strategy, namely, explanation in terms of beliefs and desires that rationalize action or behavior. To take the intentional stance toward a system is to *construe* it as explainable by means of such a strategy. Insofar as that strategy is successful, Dennett claims not merely that that system is apt for mental state-ascription, but that it *has mental states*. In other words, to have a mental state is to be successfully construed as apt for intentional explanation. Now, mental or intentional states of the sort Dennett accounts for do not exhaust the realm of the cognitive, and Dennett's view is no answer to the problem of cognition. Nevertheless, a construal-based explication of COGNITION

might be thought of as an account of cognition in terms of taking a "cognitive stance" (Ch. 6). But since there are some differences between the structure of Dennett's view and what I have in mind, and since Dennett-hermeneutics would be a distraction from the problem of cognition, I shall illustrate construal-based explication with another example.

3.4.1 Darwinian explanation as construal-based

Consider Darwin's theory of natural selection. Darwin's theory explains the distributions of traits in a population of biological organisms that persist or reproduce in such a way that they tend to pass on their own traits, and where the possession of particular traits makes a significant difference with respect to how many organisms or their offspring persist. The theory is classically taken to apply to biological organisms, especially in contexts where reproductive opportunities are not intentionally controlled by human agents. Richard Dawkins (1976) suggested that similar processes might apply to cultural units—memes—that exhibit differential survival or adoption in the environments of human minds, institutions, societies, &c. There is controversy about precisely how similar memetic selection is to natural selection, and whether a study of "memetics" can be a science (see e.g. Godfrey-Smith 2012b), but there are nevertheless distinctive similarities between the processes of natural and memetic selection. Similarities also extend to evolutionary algorithms, in which candidate solutions to an optimization problem are differentially selected based on a fitness function. If we were to characterize selectional phenomena generally, rather than by enumeration, we would naturally look to what they have in common and what distinguishes them from phenomena that are not selectional. However, these phenomena are extremely heterogenous—they (arguably) include distributions of organism phenotypes, DNA codon-types, some beliefs, slang expressions, features of regional accents, customs, fashions of dress, architectural motifs, management structures, as well as the success of specific firms in an economy, and the discovery of problem solutions that have been optimized by evolutionary algorithms. Nevertheless, selectional phenomena share a common structure: they concern populations in

which there is variation of traits or properties between individual members of the population, this variation is at least partly heritable or persistent, and the variation differentially affects the fitness of individuals, and fitness contributes to the probability an individual will persist or pass on its traits or properties. A common, abstract explanatory strategy accounts for the distribution of traits or properties in such a population (see Ohlsson 1993; Godfrey-Smith 2009a). Call that strategy selectional explanation. We may explicate SELECTIONAL PHENOMENA as a concept denoting phenomena that are apt for successful selectional explanation. Call this explication SELECTION (for a more precise treatment of selection, see § 4.3.2, p. 90).

The determination that a phenomenon is apt for selectional explanation is a construal. This is, in part, because the judgment that something is a suitable member of a population is a construal. For example, whether one sees slang trends as selectional phenomena depends on one's willingness to see slang expressions as individuals in a fitness landscape of human discourse, and whose persistence consists in continued usage. Many critics of memetics have been unwilling to see memes as individuals of the relevant kind. Moreover, controversy of this kind exists even in biology where there is consensus that there are selectional phenomena, but dispute over what the relevant individuals are (see §§ 4.3.2, 4.4).

Perhaps countenancing these disagreements about the theory of natural selection seems dreadful. However, I am not arguing that we adopt this pluralism as a regulative ideal. Perhaps there is a unique true way to think about natural selection and these disputes are best resolved. However, the current practice of biology includes several inconsistent ways of partitioning the space of individuals, and as practice-oriented philosophers of science we may use a construal-based explication of selection even if we regard the present state of disagreement in biology as a fault. If we approach these disputes as practice-oriented philosophers of science with an eye to the ecumenically adequate explication of SELECTIONAL PHENOMENON (as I am suggesting we do with respect the explication of COGNITION), then we may wish to characterize the concept in such a way as to explain the variance in biologists' conflicting judgments rather than eliminating the variance by justifying a choice of a side. Indeed, we may wish to do so for reasons analogous to the reasons

for doing so with COGNITION (§§ 2.3–2.4). This does not preclude us from also weighing in on first-order debates; philosophers have contributed and should by all means continue to contribute arguments that some ways of understanding the details of natural selection are better than others. Nevertheless, the diversity of expert opinion about how to understand concepts like COGNITION and NATURAL SELECTION is an interesting feature of scientific discourse. In describing the current state of theory, we should acknowledge the construal-sensitivity of DARWINIAN INDIVIDUAL, FITNESS, &c. A consequence of this acknowledgment is that SELECTION is construal-based. If SELECTION is a satisfactory explication of a scientific concept, then we should countenance the construal-sensitivity of SELECTIONAL PHENOMENON (at least in the mode of doing what I called "logical anthropology"; q.v. § 1.3).

Pared down, the argument of the previous paragraph is this: there is recalcitrant disagreement among biologists about the nature of the individuals that are selected, and sometimes inconsistent ways of partitioning individuals across contexts. If we aim for an ecumenically adequate explication of a concept, we should account for such expert disagreement about its ascription where such disagreement is recalcitrant. Construal-based explication provides resources for describing the divergence of expert judgments whereas object-based explication does not (it is better suited for the sectarian strategy, q.v. § 2.4). Therefore, if we aim for ecumenical adequacy concerning SELECTIONAL PHENOMENON then we have reason to favor construal-based explication as a strategy. Importantly, I do not discharge the antecedent of this conclusion. That is, I do not argue that philosophers of science should never take sides in theoretical disputes, or that they should only aim for ecumenical adequacy. However, under the guise of philosophers engaged in logical anthropology, we should also wish to produce explications that allow us to reason about concepts and practices while leaving the disputes, for the moment, as they are. We might do this as a matter of anthropological impartiality, or we might do this in order to distinguish matters of agreement and disagreement, or to have a non-partisan resource for appeal in other arguments, or as a non-question-begging step in a partisan argument.

The preceding discussion should go some way toward dispelling the temptation to think that construal-based accounts are inappropriate—too queer, or wishy-washy, or something—for philosophy of science. As long as SELECTION is a plausible sketch of an explication of a respectable scientific concept like SELECTIONAL PHENOMENON, then since SELECTION is construal-based we have some reason to accept that construal-based explications are respectable in philosophy of science. In order to further motivate my claim that construal-based explications are respectable, there are two specific worries that I shall address in this section: that construal-based accounts are aprioristic, and that they are problematically anti-realist.

3.4.2 Apriorism and construal-based explication

With respect to apriorism, one might worry that construal-based accounts are insufficiently sensitive to evidence. I have argued that construal-based accounts are a means to achieve ecumenical adequacy, which aims to characterize a concept so that substantive theoretical disputes are left in place, rather than resolved (§ 3.3.2). One way for an explication to be consistent with all competing claims is for it to be insensitive to potential counterexamples, and therefore vacuous. Invulnerability to counterexamples would of course make for a bad explication. Adams and Aizawa take pains to avoid such a charge with respect to their "mark of the cognitive," insisting that it is an empirical hypothesis. That is, it is no explication at all, and therefore there can be no objecting to it on the basis of conflicting expert intuitions. Rather, their proposed "mark" is an empirical hypothesis disconfirmable only by first-order scientific evidence of the sort gathered by scientists (this is particularly clear in their 2010).[20] As I discussed above, however, construal-based explications are *not* insensitive to counterexamples. Insensitivity to empirical evidence *of the sort gathered by scientists* does not leave an explication vacuous. Many non-vacuous explications are insensitive to first-order evidence. One does not disconfirm CLASSIC BACHELOR, for example,

[20] This "empirical hypothesis" approach is associated with what I call a "speculative mark of the cognitive"; cf. §§ 2.5, § 6.4. I describe my own view as a hypothesis, as well (§ 5.4), but it is a hypothesis about discursive norms, not about first-order matters of fact that that discourse is meant to describe.

by hauling unmarried, adult men into a laboratory and testing whether they are bachelors.[21] Rather, one examines the discourse of competent users of the concept BACHELOR to determine whether they apply the predicate "bachelor" to things they believe are not unmarried, adult men, or withhold applying the predicate to things they believe are unmarried, adult men. Likewise, with respect to COGNITION, I contend that an ecumenical explication should be tested against the discourse of cognitive scientists rather than against their data. This is so even though, over longer timescales, changes in the way a concept is used may be driven by responsiveness to evidence (cf. § 2.3.2).

This strategy of testing is not inconsistent with the Quinean rejection of analyticity because there is no pretense that conceptual explication is answerable only to non-empirical facts (§ 1.4.1). Furthermore, as logical anthropologists we may acknowledge the possibility of change in the relevant patterns of discourse. Recall that ecumenical adequacy involves explicating a concept as it is used at a time. It does not aim to establish the true nature of the concept's referent, except insofar as use by contemporary experts can be taken as a reliable guide (cf. § 1.3.1). Achieving ecumenical adequacy is no guarantee that one has settled, once and for all, what cognition is; it is merely evidence that one has characterized what contemporary science takes cognition to be. Scientific concepts are always subject to potential revision or elimination: for example, atoms are not indivisible; the aether is not real.

Nevertheless the conceptual practices of current science are a strong (if fallible) guide to our best scientific ontology, given our current evidence. In fact, the scientific concept of COGNITION has undergone significant change since the beginnings of cognitive science in the mid-20th Century. Early cognitive science had relatively strict presuppositions about what cognition would be, e.g. the classical computational theory of cognition (§ 1.2.1). Much cognitive science since the connectionist critique of the 1980s does not retain those assumptions, and the phenomena and theories of cognitive science have become much more varied, as evidenced by the arguments of

[21] I take this to be a plausible and widely-accepted premise, with apologies to Edouard Machery who insists that it is false (personal communication; see also Machery MS, Ch. 1 § 3.2, Ch. 6 § 3.2).

liberal partisans of the cognition border wars and the empirical research that they cite (§§ 1.2.2, 2.2, 5.3). It is plausible that these changes in practice reflect developments in the implicit commitments of cognitive scientists regarding the nature of what they study (§ 2.3.2). If our goal is to make current practice explicit, rather than to anticipate the shape of a future theory, then it must be possible to say something about the nature of cognition as we understand it now, and it must be possible to do so without placing bets on what the science of the future will say (§ 2.5). After all, cognitive scientists do make judgments about what is and is not a suitable object of study, and they do it without any spooky power to predict the results of a completed science. One cost of this approach, of course, is that it may not tell us what cognition really is. But I submit that it is better than any alternative approach, and that it is arrogant to think that philosophers should not take seriously the results of such an approach (§ 1.3.1, 2.5). Another cost of this approach is that we may be forced to tolerate an account that, when made more determinate in various ways, makes apparently incompatible determinations about its object. This second cost raises the specter of anti-realism.

3.4.3 Anti-realism and construal-based explication

There is some justice to the worry that construal-based explications are anti-realist, but there is anti-realism, and then there is anti-realism. Suppose, with some simplification, that the correct ascription of a concept (i.e. the proper use of a category) can depend sometimes on what I have been calling a situation—properties of the object being categorized—and sometimes to some extent on the mental context of assessment of the categorizer. Let concepts exhibiting the former kind of dependence be called "descriptive," and concepts exhibiting the latter form of dependence be called "mind-dependent." Of course concepts can be both descriptive and mind-dependent.

For example, if correctly categorizing an object as composed of calcium carbonate depends only on the properties of the object, i.e. only on the situation described, COMPOSED OF CALCIUM CARBONATE is a fully descriptive concept. Correctly categorizing an object as "helpful," by contrast, depends on the context of assessment of the judge—in particular, it depends on what it is

with respect to which the object is supposed to be helpful. HELPFUL, therefore, is to some extent mind-dependent. However, while COMPOSED OF CALCIUM CARBONATE may be a fully descriptive concept, CHALK is not. This is because which substances count as chalk depends on the suitability of those substances for human purposes. Chalks are soft, porous rocks that can be used in athletic contexts to increase friction or to make temporary markings on slates, blackboards, fabric, and so on. The substances generally included under this heading include otherwise chemically dissimilar substances: calcium carbonate, calcium sulfate, magnesium silicate, and titanium dioxide (see Weiskopf 2012 for discussion). While matters of chemical composition may be fully descriptive, being chalk is mind-dependent in this thin sense since inclusion or exclusion from the category depends upon relations to human projects and history. Construing a sample of magnesium silicate *as chalk* involves conceiving of the substance in terms of its potential use to tailors, or its similarities to other chalks. Dominic Murphy makes a similar point about the concepts WEED, VERMIN, and MENTAL DISORDER (2006). On some understandings, to be an anti-realist about F is just to take F to be "mind-dependent" in this sense (As discussed in § 3.3.1 above this is obviously complicated in cases where one makes judgments about one's mental states).[22] Based on this criterion, construal-based explications do imply anti-realism about their explicanda, since the goodness of construals depends on the context of assessment of the judge. For instance, whether a phenomenon is selectional in the sense articulated by SELECTION depends on whether we can get ourselves to construe certain items as individual members of a population subject to selection pressure.

There are two things to note about this kind of anti-realism, however. First, to acknowledge that judgments or claims are anti-realist in this modest sense does not imply that they are unconstrained by facts or situations. It is not sufficient for being chalk that a substance be conceived of as chalk; it must also be a soft substance that either leaves visible marks on surfaces by adhering to them, or a powdery substance used by athletes or workers to reduce slipping, &c. Only some

[22] E.g. Mark Sprevak has said of my construal-based explications that "For my money, the view is anti-realist in that it makes the facts about what is and isn't cognitive depend not only on the system itself, but on our reactive dispositions as enquirers" (personal communication).

substances have properties such that they can serve these purposes. Similarly, not any phenomenon can be selectional just because it is construed as such. In the construal-based explications of scientific kinds discussed above, construing an item as F does not imply that it is F. Rather, construing items in certain terms (e.g. as G) is a feature of judging that they are F. In cases suitable for scientific explication, these construals depend on inquirer-independent facts about the target phenomena. It might be the case that such construals play an important role in the context of discovery, or in novel cases of generalization. William Bechtel and Adele Abrahamsen (2005), for example, argue that much generalization in the biological sciences is based on judgments of salient "similarity," which judgments could be cashed out as construals (cf. § A.4). Robert's paradigmatic cases of construals are similarity judgments.

The second thing to note about this kind of anti-realism is that it does not imply a stronger form of anti-realism. The more worrying notion of anti-realism is one that encompasses instrumentalism and constructive empiricism, and the suggestion that theoretical entities in science are mere fictions or inference tickets that enable us to reason from some observable state of affairs to another observable state of affairs, without corresponding to any facts about the more detailed structure of the target systems. On prominent versions of scientific anti-realism, scientific theories or claims are (often or always) not literally true, or should not be believed (e.g. van Fraassen 1980). It may be tempting to think that acknowledging a place for construals in serious scientific discourse opens the door to this form of anti-realism, since construals blur the lines between literal and non-literal uses of terms (§ 3.3.2). However, as I claimed above there are construals that depend only on taking certain real relations to be more salient than others, or on taking certain items to be grouped together when alternative groupings are possible. These relatedness or groupiness judgments are appropriate when, as in the case of chalk, the properties of the target phenomenon can be viewed in terms that make some relations more salient rather than others. Or, as in the case of selectional phenomena, conceiving items like slang utterances as grouped together in a certain way when they could well be grouped in different ways, or where the suitability of the grouping is contentious among experts. Furthermore I have taken pains to describe the sense in which

endorsing a construal-based explication can furnish us with ontological commitments about the phenomenon being explicated, as opposed to the "less queer" strategy of ascriptional explication (§ 3.3.4). My claim here is certainly not that all construals have a place in science; there may be no need to find a place in scientific discourse for a construal that President Erdoğan resembles a character from a popular fantasy novel. My claim is merely that some scientific thinking depends upon construals, and that this dependence can be exploited for various purposes by philosophers of science.

3.5 SUMMARY

The main purpose of this chapter was to describe some logical resources for realizing ecumenical adequacy in conceptual explication. To that end, I described the semantic features of a distinctive form of judgment I called "construal," and described how construals can feature in conceptual explications with unusual properties. The appeal to construals contributes to our descriptions of certain kinds of judgments whose semantic values vary interpersonally or intra-personally but independently of events affecting the object of judgment. In particular, the semantic value of a construal varies according to factors more find-grained than situations; it also varies according to a context of assessment where the context is determined by ancillary mental states of the judge (§ 3.3.1). Of course, construal-based explications can be used to provide criteria for the ascription of blatantly construal-sensitive explicanda (like BRAVE), but they can also be used to illuminate the construal-sensitive use of expressions like "bachelor" that are sometimes non-literal or are sometimes extended to cover non-paradigmatic circumstances (§ 3.3.2). I described the distinctive structure of valid counterexamples for construal-based explications (§ 3.3.3), discussed the inferential properties of construal-based explications (§ 3.3.4), and argued that construal-based explication has a place in anthropological philosophy of science (§ 3.4).

In order to make more plain how construal-based explication can be applied to the problem of cognition, however, I shall need to introduce another bit of logical machinery: parameterization. In the next chapter, I will describe this device and its advantages as a strategy for achieving ecumenical adequacy. Having done this, I will be in a position to provide an example of an ecumenical, construal-based, parameterized explication of cognition.

4.0 THE PARAMETERIZATION STRATEGY

4.1 OVERVIEW

I argued above (Ch. 2) that in circumstances like the problem of cognition, which feature pervasive and entrenched disagreement about the extension of a concept, it would be fruitful to aim for ecumenical extensional adequacy. That is, rather than aiming (like Quine) to produce an explication whose extension is more precise than that of the explicandum term, we should sometimes aim to produce an explication whose extension varies in precisely the ways that the explicandum's extension varies. In the last chapter I described one way of achieving such variation: the exploitation of predicates that vary in their semantic properties based on the mental context of the ascriber (i.e. construals). However, although I described some semantic properties of construals, I have not provided an analytically clear method for evaluating them. I imagine there are many such methods, but in this chapter I will describe one method that suits contexts of scientific conceptual explication: parameterization.

I will begin by describing parameterization abstractly, and motivating it by reference to a familiar philosophical context (§ 4.2). I will then provide some simple examples of parameterization, discussing some of the kinds of use to which parameterization might be put (§ 4.3). I will conclude by reflecting on the advantages that parameterization provides for understanding discourses mired in controversy (§ 4.4); namely that they localize topics of controversy, structure opportunities for progressively less controversial resources for appeal ("Quinean retreat"), and afford a "divide and conquer" approach to achieving convergence in judgments.

4.2 PARAMETERIZED EXPLICATION

In contexts amenable to mathematical or logical description, one distinguishes variables (or arguments of functions) from constants and parameters. Take, for example, a simple mathematical formula: the form of a linear equation in the slope-intercept form:

LINEAR EQUATION:

$y = mx + b$

The equation specifies a relationship between the variables x and y. However, the value of y given x is not determinate unless values are also assigned to the parameters, m and b. In the case of a particular linear relationship, e.g. the one expressed by "$y = 2x + 3$," the values for the parameters are fixed so that they function as constants. In the general form of the equation, however, they are abstracted so as to be capable of being assigned a variety of determinate values. Altering the value of the parameter alters the input-output characteristics of the expression (i.e. the relationship between the variables).

Or consider a logical context. The general form of an explication is a relationship (implication, equivalence, or identity) between the explicandum and the explicans, and both the explicandum and the explicans can be thought of as functions from arguments to semantic values. Take, for example, the general schema for a construal-based explication (p. 64), expressed in a more formal notation:

CONSTRUAL-SCHEMA 2:

$(x)[\ F_c\ (x)$ iff $G_c\ (x)\ ]$

The variable x takes as its domain any item in the relevant discourse context. However, in order to semantically evaluate an expression such as $F_c\,(a)$, where a particular item denoted by a is

substituted for x, and F is a construal-sensitive predicate, we require some specification of the ancillary mental states of the judge, denoted by the context of assessment c. Thus, an expression featuring construals such as "skydiving is brave" can be evaluated only relative to a context of assessment that determines which sorts of things are brave. The manipulation of the context of assessment alters the input-output characteristics of the logical expression, i.e. alters the mapping from arguments of construal-sensitive predicates to semantic values. In this way, contexts of assessment function in a manner analogous to parameters.

Construal-sensitive predicates, i.e. those whose extension varies along with a context of assessment, can be called *parameterized predicates*. An expression with parameterized predicates has a determinate semantic value given a determinate value given a context of assessment. I will sometimes refer to these values as "values," and often (since it is more natural) I will refer to them as "interpretations." The introduction of this language gives us a new way to express the distinction between object-based and construal-based explications (§ 3.3.2). Object-based explications are those in which all predicates function as constants, i.e. in which the input-output characteristics of the predicates do not vary according to their interpretations. In construal-based explication, some predicates are parameterized; i.e. their input-output characteristics do vary according to their interpretations. We may call an explication featuring parameterized predicates a *parameterized explication*, and the construction of such explications may be called *parameterization*.

4.2.1 Parameterization and ecumenical extensions

In certain cases, when there is a core of agreement concerning the extension of an explicandum, but a penumbra of controversy, the strategy of parameterization can yield ecumenical extensions of the sort described in Section 2.5 above. Recall that ecumenical extensions are graded in such a way as to perspicuously represent cases that engender agreement and cases that engender disagreement. Let us call the extension of a parameterized explication with determinate interpretations for all parameters a *sectarian extension*. We may discuss the extension of a

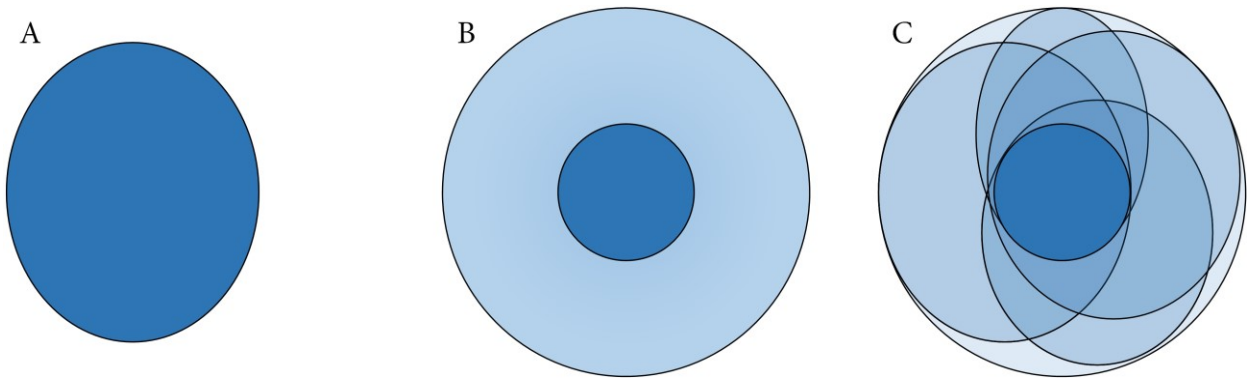

Figure 4.1: Venn diagrams representing parameterized extensions. A: A sectarian extension. B: An ecumenical extension, distinguishing paradigmatic and controversial items (cf. Figure 2.1 C). C: An ecumenical extension achieved through parameterization.

parameterized explication with determinate interpretations, but no particular interpretations assigned, by examining the set of sectarian extensions. Call the set of sectarian extensions an *ecumenical extension*. Where there is some intersection and some divergence between sectarian extensions, we may identify cases that engender relative agreement and those that engender relative disagreement, respectively. If there is a core of items that belong to most or all sectarian extensions, we may call such items paradigmatic.[23] Other items, that belong to at least one sectarian extension but not all or most, we may call controversial (Figure 4.1).

To illustrate, let us revisit Chandana and Kasson (§ 3.3.2), along with a parameterized version of our toy explication BRAVE BEHAVIOR (p. 61):

[23] I hedge by saying "most or all" because in open parameterization (q.v. § 4.3.2) the available interpretations are not exhaustively enumerated. The hedge is adequate in the context of my current project because of my simplifying assumptions: an interpretation is available if it is avowed in a prominent publication, or suggested by a careful reading of the same (§ 1.3.2). More granularity is not required. In a more careful treatment of paradigmaticity, one would want to characterize interpretations not simply as "available" or "unavailable," but in terms of their degrees of acceptance by participants in a discourse. In such a case, interpretations (and therefore sectarian extensions that feature them) could contribute more to paradigmaticity when they are more widely accepted in the discourse. Then one could produce measures of paradigmaticity more fine-grained than "paradigmatic" and "controversial." However, I eschew this complication because it is difficult to glean from a qualitative literature review, and introduces significant additional complications for little gain. In other contexts where parameterization is suitable, and perhaps in further work on COGNITION, this approach could be explored.

BRAVE BEHAVIOR 2:

An action ϕ is brave iff ϕ is dangerous for its agent, and ϕ is *admirable*.

The construal-sensitive predicate in the explicans, *ADMIRABLE*, functions here as a parameter and is italicized above for perspicuity. The other elements of the explication—the variables, the logical connectives, and predicates like ACTION and IS DANGEROUS FOR ITS AGENT that have constant interpretations, are unparameterized. Let us also consider three actions: protesting injustice, skydiving, and brutally oppressing the masses, each of which is (let us say) dangerous for its agent. Recall that Chandana finds skydiving admirable and brave, but Kasson does not. Suppose that Chandana and Kasson both agree that protesting injustice is admirable and brave, and that brutally oppressing the masses is not admirable or brave. The term BRAVE has two sectarian extensions corresponding to the contexts of assessment determined by Chandana's evaluative commitments and by Kasson's. According to Chandana it includes protesting injustice and skydiving, but according to Kasson it includes only protesting injustice (Table 4.1). In our mini-discourse, we may divide actions into three categories just as ecumenical extensions do: paradigmatically brave actions, controversially brave actions, and paradigmatically non-brave actions. The ecumenical extension of BRAVE BEHAVIOR 2 classifies our three actions according to these categories in the following way: protesting injustice is paradigmatically brave, since it belongs to every sectarian extension of BRAVE. Skydiving is controversially brave, since it belongs to some but not all or most sectarian extensions of BRAVE. Brutally oppressing the masses is a paradigmatically non-brave action, since it belongs to no sectarian extensions of BRAVE. Thus, the ecumenical extension of BRAVE, according to BRAVE BEHAVIOR 2 and our mini-discourse, correctly classifies actions as paradigmatically or controversially brave, and is ecumenically extensionally adequate.

The strategy of explication that concerns me in this chapter, however, is a strategy that is slightly more complex than the one suggested by CONSTRUAL-SCHEMA 2. In particular, I am presently concerned with cases that feature more than one parameterized predicate in the explicans, and where there is no strong relationship between the interpretative values assigned to those

Table 4.1: Variation in the sectarian extension of BRAVE according to BRAVE BEHAVIOR.

Variation in the sectarian extension of BRAVE according to BRAVE BEHAVIOR

Interpretations of parameter	Sectarian extension of BRAVE includes
ADMIRABLE: Chandana's interpretation	Protesting injustice Skydiving
ADMIRABLE: Kasson's interpretation	Protesting injustice

predicates. In BRAVE BEHAVIOR, for example, there was only one construal-sensitive predicate in the explicans, "admirable," and its interpretation was ex hypothesi tightly correlated with a conception of the explicandum, BRAVE. This was plausible enough for a toy example because judging an act to be brave and judging an act to be admirable depend on similar ancillary mental states, namely commitments concerning the moral evaluation of actions. In the following, I will be concerned with explicantia in which the interpretations of parameterized predicates depend on relatively independent mental states.

4.2.2 Parameterized explications as themes

Extensional adequacy, even ecumenical extensional adequacy, is a somewhat minimal criterion for success. Consider, for a moment, an alternative explication of BRAVE:

> BOURGEOIS BEHAVIOR:
> An action ϕ is brave iff ϕ is dangerous for its agent, and ϕ is not *bourgeois*.

In this explication, the predicate *BOURGEOIS* functions as a parameter. Imagine now that Chandana and Kasson both agree that protesting injustice is not bourgeois but that brutally oppressing the masses is, and furthermore Chandana doesn't think of skydiving as bourgeois but that Kasson does. In our mini-discourse, BOURGEOIS BEHAVIOR is also extensionally adequate. I do not object to the possibility that there may be many distinct but equally good explications of a concept—I am after all committed to articulating norms of use, not meaning (§ 1.4.1), and there may be many ways to express norms or to characterize patterns of behavior. But a parameterized explication is most interesting if it can be used for more than characterizing a state of disagreement in a discourse at a time.

So what factors other than ecumenical adequacy make a parameterized explication interesting? I will suggest three below in my discussion of controversy (§ 4.4). In this subsection, I recommend a broad one: that a parameterized explication, in making explicit a rule for using a concept that is subject to varying interpretations, articulates a theme that is recognizable in each more determinate (sectarian) interpretation. This is perhaps clearest when the uninterpreted explication commands broader assent than any of its interpretations.

Consider, for example, the following familiar explication of KNOWLEDGE:

JTB:

Knowledge is *justified*, *true belief*.

It is an interesting feature of this explication that, Edmund Gettier's (1963) famous counterexamples notwithstanding, it enjoys broader assent than any more specific explication of KNOWLEDGE, or of the terms in the explicans. That is, it is generally agreed among Western philosophers that JTB delivers correct verdicts on most cases even though it is not generally agreed among Western philosophers how belief is to be ascribed, how truth is to be ascribed, or how justification is to be ascribed. (And philosophers aside, it is not agreed among people in general which propositions express someone's beliefs, which kinds of propositions are true, or which kinds of propositions are

justified.) Each of these terms is construal-sensitive; the ascription of each term may vary between philosophers based on a complex collection of ancillary mental states, and the pattern of ascription for each can be largely independent of the pattern for the others. In the context of logical anthropology, we could therefore treat JTB as a parameterized explication in which *JUSTIFICATION*, *TRUTH*, and *BELIEF* are all parameterized predicates. Understanding KNOWLEDGE this way affords us a frame on which we can hang various more specific epistemological views and disputes. Most clearly, many pre-Gettier disputes turn on what one should mean by *JUSTIFICATION*—e.g. internalists about justification vs. externalists about justification, or reliabilists vs. other externalists, or process reliabilists vs. virtue reliabilists. The discussion of BRAVE BEHAVIOR 2 above was, for purposes of expository clarity, limited to a small set of interpretations for the parameter *ADMIRABLE*, but in some parameterizations it may be left open which parameters are available, as in the case of JTB. I will discuss "open" parameterization in more detail in Section 4.3.2.

There is also another way that parameterized explications can articulate a theme that is recognizable when developed in different forms. Recall that the variable ascription of construals reflects not only interpersonal variability in judgment, but intra-personal variability over time (§ 3.3.1). It is an extra benefit if a parameterized explication accounts not only for variability in judgments between individuals at a time, but variability in judgments over a career of conceptual development (cf. §§ 2.3.1–2.3.2 on the development of COGNITION). A parameterized explication, especially with open sets of interpretations for its parameters, may be robust not only between people but diachronically independent of changes in use. Thus, parameterized explications reveal their value when they form the basis of what Mark Wilson calls "analytic prolongation" for their explicanda, the repurposing of a concept to be of use in situations where it was not used before (Wilson 2006, e.g. 312–319). I am not claiming here that parameterized explications are of particular use in *predicting* how a concept will be analytically prolonged in the future, even the near future. I claim only that they sometimes have value in making clear how new uses of a concept are indeed

new uses of a familiar concept, rather than entirely new concepts used for purposes embarrassingly similar to the uses for old concepts expressed by the same term.

4.3 USES OF PARAMETERIZATION

At this point it may be useful to expand the repertoire of examples for parameterization. I will first (§ 4.3.1) describe one fanciful example about goats which is simple, but more structured than BRAVE BEHAVIOR 2. In particular, each parameter has a fixed set of interpretations with two members, and the ecumenical extension exhibits the paradigmatic structure I described above (§ 4.2.1). This example serves to illustrate the pragmatic criteria by parameterizations might be judged. Second (§ 4.3.2), I will expand on my earlier discussion of natural selection (§ 3.4.1), modeling it as a case of what I call open parameterization, in which there is not a fixed number of interpretations for the parameters. I will conclude the section (§ 4.3.3) by addressing some worries that might be raised about the arbitrariness of open parameterizations.

4.3.1 Goats of quality and the uses of parameterization

Consider the following open sentence explicating the ad hoc concept GOAT OF QUALITY:

> QUALITY GOAT:
> *x* is a goat of quality iff *x* is a goat, *x* is *handsome* and *x* has a *good disposition*.

Let *HANDSOME* and *GOOD DISPOSITION* be parameters, and for the sake of simplicity let each admit of only two interpretations. First, on a permissive interpretation *HANDSOME* applies to anything that is pleasing to look at. On a restrictive interpretation, "*x* is handsome" is satisfied if and only if *x* refers to an animal of a recognized breed that meets the breed standards of the

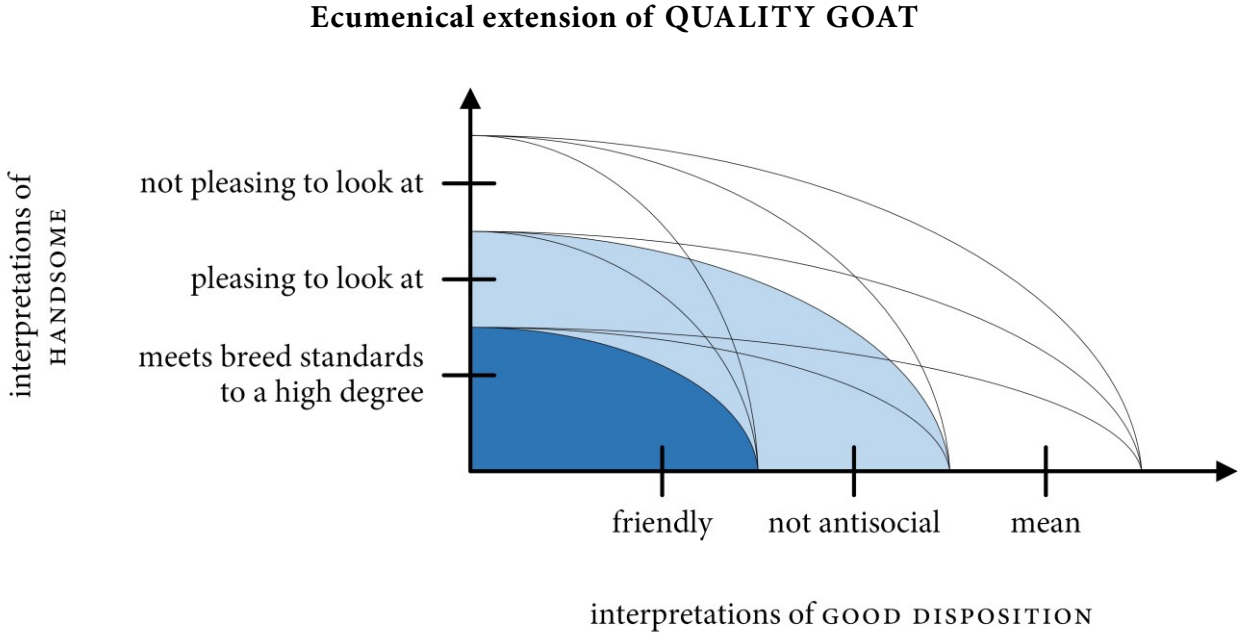

Figure 4.2: Ecumenical extension of QUALITY GOAT. The nine arcs correspond to nine ways of categorizing goats according to the explication. The dark region on the graph includes paradigmatic goats of quality; the pale region includes controversial goats of quality; the white region includes goats of poor quality.

American Goat Society to a high degree. Suppose also that any animal meeting the breed standards to a high degree is also pleasing to look at; i.e. the sectarian extension of *HANDSOME* on the restrictive interpretation is a subset of the partisan extension on the permissive interpretation. Regarding *GOOD DISPOSITION*, suppose that on a permissive interpretation the predicate applies to any animal that is not actively belligerent or antisocial; on a restrictive interpretation the predicate applies only to an animal that is particularly friendly to humans. The rest of the open sentence—the predicate GOAT and other features of the sentence including its connectives—is not parameterized.

This parameterization exhibits the paradigmatic structure I described above, with an umbra of paradigmatic cases and a penumbra of controversial cases (Figure 4.2). On one assignment of interpretations to parameters, the partisan extension of the formula contains only goats that are

friendly and excellent specimens of recognized breeds. These are the paradigmatic cases, as they belong to every sectarian extension of the explicans. On another assignment, the partisan extension contains friendly goats of mixed breeding. On yet another assignment, the partisan extension contains any attractive goat that is not belligerent, even if it is not friendly. Goats that are well-bred and pleasant but not friendly are in some but not all sectarian extensions; goats that are friendly and attractive but not well-bred are also in only some sectarian extensions. These goats are all controversial examples of quality goats. Goats that have a disquieting countenance are in no sectarian extensions, and are thus not quality goats. Goats of recognized breeds that meet the American Goat Society's breed standards but that are downright mean are also not quality goats. The open sentence, then, independently of assignments of particular interpretations to parameters, has an ecumenical extension of the right form.

The example is admittedly fanciful, but reflection on its uses may illustrate the way that explications can articulate themes. QUALITY GOAT might, for example, articulate a rough guideline for judging goat competitions at, say, the Lorain County Fair, where there is some disagreement between judges as to the nature of the competition (as far as I know Lorain County, Ohio does have a goat competitions but is not renowned as a center of fine goat breeding). Perhaps excellent show goats satisfy the restrictive interpretation for *HANDSOME*, but need only satisfy the permissive interpretation of *GOOD DISPOSITION*, whereas excellent companion goats exhibit the reverse pattern: satisfying the permissive interpretation of HANDSOME and the restrictive interpretation of *GOOD DISPOSITION*). The explication then models the state of contention between judges who believe they should be giving awards for show goats and judges who wish to give awards for companion goats. The explication may not serve other purposes so well; for example, it might not serve as an appropriate norm for identifying excellent dairy goats, meat goats, fleece goats, or pack goats. However, handsomeness might be a proxy for endurance, or indeed for dairy, meat, or fleece quality. And goodness of disposition might be a desirable quality, as well, since goat farmers must interact with their goats. Importantly, it is empirical facts that settle whether

QUALITY GOAT is a useful guideline for selecting goats to farm, not anything resembling an "analytic" truth.

4.3.2 A parameterized treatment of selection

QUALITY GOAT is an example of *closed* parameterization. That is, each parameter has a fixed number of interpretations (specifically, there are two interpretations for each parameter). It is admittedly messier but more useful to consider *open* parameterizations, in which there may not be a fixed or determinate number of well-defined interpretations for the parameters. The available interpretations for each parameter in the model are constrained not by exhaustive enumeration, but by the commitments of parties in the discourse being modeled.

In order to illustrate open parameterization, I will return to my discussion of selection from the previous chapter (§ 3.4.1). Consider the following parameterized explication of selectional phenomena, based on Richard Lewontin's (1970) famous characterization:

> SELECTION:
>
> Selection occurs where
>
> 1. There is a population of *individuals*,
> 2. There is variation in the traits of those individuals,
> 3. This variation in traits causes variation in the *fitness* of individuals, and
> 4. Possession of the traits is partly *heritable*.

Let us call phenomena that meet this description "selectional phenomena." In the paradigmatic cases of evolution by natural selection, the individuals are organisms of a common species inhabiting a geographic area. Individuals are fitter when they reproduce more, so that they cause more of their own traits to be represented in the population at future times. The result of their reproduction must be offspring that inherit some of the traits that contribute to their fitness,

typically through the germ-line transmission of genetic structure. Call this the paradigmatic conception of SELECTION. Of course, natural selection takes many forms, not all of which are aptly described by the brief characterization I just offered; my characterization is simplified for the sake of expedient exposition.

What other kinds of phenomena are selectional phenomena? Evolution by natural selection is the paradigmatic case, but controversial cases include (1) breeding by "artificial" selection, (2) the development of social structures by memetic selection, and (3) the selection of solutions by evolutionary algorithms. In order for SELECTION to be ecumenically adequate, these phenomena must be included in some sectarian extensions of SELECTION, and excluded from others. I will briefly discuss each of these three cases in turn, noting which predicates must be parameterized in order to accommodate them. In order to simplify the discussion, I will mostly characterize interpretations of the parameters simply in terms of which entities are members of their extensions, rather than in terms of different criteria for ascribing a parameterized predicate.

The simplest case is that of breeding by artificial selection, discussed by Darwin but distinguished from natural selection (1859). The main difference between natural and artificial selection is the interpretation of *FITNESS*. In cases of artificial selection, the reproductive opportunities of individuals in the population are limited by human actions that are motivated by the goal of promoting certain traits. In paradigmatic cases of natural selection, by contrast, reproductive opportunities arise freely, unrestricted by this specific category of human action. Thus, the difference between natural and artificial selection can be accounted for by assigning different interpretations to a single parameter: *FITNESS*. Any interpretation of *INDIVIDUAL* or *HERITABLE* that is appropriate for cases of natural selection will also be appropriate for a construal of SELECTION that has artificial selection in its sectarian extension.

Dawkins is credited with coining the word "meme" (1976), and contributing to the popularization of conceiving of cultural change and development as processes analogous to natural selection. In order to understand cultural evolution as a phenomenon described by SELECTION, one must interpret *INDIVIDUAL* so that it encompasses not only organisms or genomes, but items

such as artifacts, the behaviors and other linguistic performances of people (or dispositions to behave), and institutions like clubs, governments, firms (or perhaps the organizational structures and policies of such institutions). Fitness in cultural evolution does not generally involve the process of physical replication by means of mechanisms internal to the individual; rather, it involves being successfully reproduced by mechanisms of other entities such as craftspeople, language users, or members of institutions (in this way the interpretation of *FITNESS* for cultural evolution is similar to an interpretation of *FITNESS* that corresponds to viral reproduction, in which reproductive processes crucially involve mechanisms internal to other organisms). In the case of institutions, fitness may consist in mere persistence, rather than reproduction (and in this way the interpretation of *FITNESS* may resemble interpretations of fitness suitable for describing various "biological organisms" in Godfrey-Smith's sense, q.v. his 2012a and discussion below). Similarly, heritability for cultural evolution depends not on germ-line transmission of genetic structure, but processes that create individuals that resemble the "parent" individual, e.g. artifacts with similar design features, linguistic performances featuring the same lexical items or other formal features, or institutions with the same organizational structure or policies. And where fitness for institutions consists in persistence, heritability may consist in the conservation of organizational structure or policies, in addition to the inspiration of those traits. Thus, on certain interpretations of the three parameters in SELECTION, selectional phenomena may be construed so as to include cases of cultural evolution (but see e.g. Godfrey-Smith 2012b for a review of worries about whether these are fruitful construals).

Evolutionary algorithms borrow inspiration from Darwin in pursuit of solutions to computational problems, usually optimization problems. The algorithms manipulate populations of candidate solutions, so in order to bring evolutionary algorithms into a sectarian extension of SELECTION one must interpret *INDIVIDUAL* so that it includes such candidate solutions. For the sake of simplicity we may imagine here that candidate solutions are represented as ordered n-tuples of values. The solutions are judged according to a fitness function that selects better-performing solutions and discards poorly-performing solutions. Selected solutions "survive" and their

components (i.e. values) are recombined or mutated to create "offspring" solutions, and parents and offspring make up the individuals in a population at the next time step. Then the process is repeated: solutions are selected by a fitness function, recombined to create offspring, and a population is created for the next time step. So evolutionary algorithms differ from the paradigmatic conception in the extensions of *FITNESS* and *HERITABLE*, as well. As in the case of artificial selection, fitness is determined by a special class of human actions rather than by circumstances that exclude the effects of such actions. Specifically, fitness is selection by a human-made decision algorithm for inclusion in the population at subsequent time steps. And the inheritance of traits is accomplished through the preservation of solution components across time steps, rather than through biological germ-line transmission.

Disputes over the proper interpretation of these parameterized terms carry over into theoretical biology, as well. Does the extension of *INDIVIDUAL* really include whole organisms, or just their genomes? Or just parts of genomes—e.g. DNA codons, whole protein-coding sequences, or phenotype-coding sequences? Or can it include groups of biological organisms in a population of other such groups? Does *FITNESS* refer to a propensity toward reproductive success, or a ratio of offspring in a population? Do fitness characteristics belong to individuals, or groups of individuals with particular traits, or to traits themselves? Must inherited traits be expressed in phenotypes? Must they be absolutely similar or can they involve degrees of similarity? Though biologists have reached significant agreement on answers to some of these questions, they have all been matters of dispute and some remain contentious. For example, Ellen Clarke (2013) argues that the criteria of identity for Darwinian biological individuals simultaneously license distinct and inconsistent partitions of individuals in marginal cases like ant colonies and aspen groves. Peter Godfrey-Smith (2012a) cites similar cases, and argues we may distinguish between individual biological organisms—biological systems that maintain themselves and persist over time, but for which reproduction and evolution are conceivably optional—and Darwinian individuals—or the units of selection. Godfrey-Smith's conclusion amounts to the suggestion that disputes over the individuation of biological organisms should be divorced from the role that *INDIVIDUAL* plays in

SELECTION. And both BIOLOGICAL ORGANISM and DARWINIAN INDIVIDUAL are notions subject to further marginal cases and disagreement. But as was the case with JTB, Lewontin-inspired explications of SELECTION are valuable because they command greater assent than more specific characterizations. It is in such cases that parameterization can be used to model the state of a discourse concerning a concept, so that an ecumenical theme can be articulated which finds expression in sectarian norms of use.

4.3.3 Free parameters in explication

An open parameterization like SELECTION is apt to describe cases in which there is not an exhaustively enumerated set of interpretations for each parameter. A part of the value in a concept like selection is its availability as a schema for novel applications to phenomena (Ohlsson 1993), as in the cases of cultural evolution and of evolutionary algorithms (though such applications may pretty awful, as in the case of social Darwinism). An open parameterization is a tool for characterizing a concept such that its potential for Wilsonian "analytic prolongation" is made perspicuous.

It might be objected that open parameterization seems a rather ad hoc method of characterizing a discourse. After all, if a term can receive any interpretation, changing the extension of the explicans, then an explicans can be manipulated into determining an arbitrary extension so as to accommodate any pattern of cases. So, it might be worried, parameterizations are immune to counterexamples after all and make no determinate claims. I would resist this objection. In particular, it is not licit to invoke just any interpretation for parameters. Open parameterizations have sets of interpretations that are not exhaustively enumerated, but that does not imply that any interpretation is permissible. There are reasonable constraints on the permissible interpretations of a parameter; in particular, each interpretation should correspond to a context of assessment—a set of ancillary mental states or theoretical commitments—that is occupied by participants in the discourse. Ideally, there should be independent evidence that particular judges have commitments

regarding the interpretation of a given parameter. For example, there might testimonial evidence by particular judges concerning their commitments, which evidence constrains the plausible attribution to those judges of an interpretation for a parameter. Such evidence can confirm (defeasibly) or disconfirm the claim that such an interpretation is considered available in a discourse. The invocation of interpretations that do not correspond to the actual commitments of participants in a discourse is at best empirically risky, and at worst demonstrably false. Furthermore, parameterization is a form of construal-based explication, and as such parameterizations are subject to the norms of counterexampling I described in the previous chapter (§ 3.3.3). In particular, an appropriate context of assessment determines an assignment of interpretations, and parameterizations make claims about the relation between an assignments of interpretations and a sectarian extension.

I have also taken care in the foregoing examples to clearly identify a limited number of parameterized predicates in each explication, and to keep that number small. Where there is no evidence concerning which interpretation a judge assigns to a parameter, the assignment of a value to that parameter must be treated in a manner analogous to "free parameters" in mathematical models. An abundance of free parameters in such a model weakens the empirical purport of the model, and likewise an abundance of free parameters in a parameterized explication weakens the empirical purport of the explication.

I have recommended parameterizations for two fairly distinct roles: the modeling of states of controversy, and the modeling of avenues for analytic prolongation. However, this second role is best fulfilled retrospectively, rather than prospectively. In cases where novel uses of a concept are precisely those that inspire controversy, as in the case of the cognition border wars, it is appropriate to use a common expressive tool to model both phenomena. However, my primary aim is to model states of controversy. I turn presently to the discussion of the advantages of parameterization for this task.

4.4 CHARACTERIZING CONTROVERSY

Parameterization affords three distinct advantages for understanding controversy. First of all, if an explication is adequate it distinguishes points of agreement and disagreement, localizing disagreements in the assignment of interpretations to parameterized predicates. Thus, in our toy examples BRAVE BEHAVIOR 2 and QUALITY GOAT, it was in the ascription of parameterized expressions such as "admirable," "handsome," and "good disposition" that disagreement lay. The ascription of unparameterized expressions like "dangerous" and "goat" were not topics on which the judges in our toy discourses disagreed, and likewise the general structure of the explications (e.g. that predicates in the explicans were conjoined rather than disjoined) was not a matter on which judges disagreed. The same might be said for JTB and SELECTION. Parameterized explications thus might have the appearance of being truistic, but that is to be expected. In order to serve the ecumenical function of localizing points of agreement and disagreement, the unparameterized elements of the explication must draw broad or (ideally) universal assent among participants in the discourse being modeled. They should articulate common ground between the viewpoints being accommodated.

Nevertheless, the defense of a particular parameterization is not trivial. It is often difficult to see, in a particularly tangled discourse like that arising from the cognition border wars, precisely where the disagreements are or how to describe them succinctly. In messy cases such as these, there is often a risk that parties to a dispute talk past each other, and fail to understand each other. That is, where the norms of use for a concept differ between disputants, disputants might fail to realize that these norms differ. And even if they do realize that the norms differ, they may not understand what the alternative norms are (i.e., disputants may take themselves to be in the situation I described in my discussion of ambiguity, § 2.4). A successful parameterization can function as an intermediary for counteracting these failures of understanding, inspiring in their place what might be called "conceptual sympathy." That is, parameterizations may allow parties to a conceptual

disagreement to understand each other as engaged in a common enterprise, even if they are not moved by each other's positions.

A second advantage conferred by parameterizations is the structuring of opportunities for progressively less controversial resources for appeal. For example, if two judges at the Lorain County Fair disagree about whether a specimen is a goat of quality, they may appeal to QUALITY GOAT: is the entity in question a goat? Is it handsome? Does it have a good disposition? Robert Brandom refers to the behavior as *Quinean retreat* (Brandom 1994, 221–224; cf. Quine 1960, 42–44). By identifying points of disagreement, a parameterized explication offers guidance on how Quinean retreats may be conducted. The matters most likely to inspire continued disagreement are those described by the parameterized predicates (e.g. what counts as being handsome) rather than by unparameterized elements of the explication (e.g. how to determine whether something is a goat, or whether it must also be handsome). If there is no agreement on which interpretation should be assigned to a parameter in a given circumstance, disagreements can be conducted productively about which interpretation is appropriate, forestalling a scenario in which judges talk past each other. If there is agreement on an assignment of interpretations to parameters, then the Quinean retreat may be conducted even further. For example, according to the American Goat Society a face shape that is slightly "dished" (concave) is preferable to a straight nose in pygmy goats, and far preferable to a Roman nose (convex face shape). The opposite is true of a Nubian or a Boer goat. Disagreements about the curvature of goat noses are more tractable than disagreements about goat quality per se.

By facilitating Quinean retreat a parameterized explication directs us efficiently toward the identification of less controversial resources for appeal in adjudicating cases. If disagreement about an embattled term is not hopeless, there will be some well-traveled lines of retreat that should be incorporated into its parameterization. Thus in JTB, the proper interpretation of *JUSTIFIED* in particular is a long-standing source of contention between epistemologists. In SELECTION, the proper understanding of *INDIVIDUAL* is one source of perennial friction that manifests in debates over the "units of selection" (see Lloyd 2012 for review). Even apart from the units of selection

debate, the proper individuation of individuals is a continuing source of friction in the philosophy of biology.

So parameterizations should incorporate well-traveled lines of Quinean retreat. However, an insightful parameterization may also identify less-traveled lines of retreat and will prove its value as a piece of intellectual labor if it directs disputants toward them, either effecting the productive resolution of disputes or at least the promotion of conceptual sympathy.

By facilitating Quinean retreat, parameterization eases the adjudication of particular cases that arise in a synchronic state of general conceptual disagreement that the present disputants are not in a position to resolve, because they are merely participants in a larger discourse. However, parameterization may also facilitate the diachronic resolution of conceptual disagreements. In particular, a third advantage of parameterization is that, where there are multiple parameters, it is possible to take a "divide and conquer" strategy toward resolving conceptual disagreements. Parameters represent areas of disagreement, and if disagreements are eventually resolved, either on an occasion of dispute or in general for a linguistic community, then parameterization becomes an unnecessary complication in modeling conceptual norms. It is possible to resolve disputes regarding one parameter even if the proper interpretation of the others remains contentious. Parameterization may encourage such incremental progress by dividing a complex topic of disagreement into several more manageable topics of disagreement.

To take an example, let us suppose that judges at the Lorain County Fair have been arguing for years about how to judge the goat competition. One year they decide in a meeting to adopt the American Goat Society standards for determining whether a goat is handsome. The more permissive interpretation of *HANDSOME* is therefore extinguished in their competition discourse, and where there is only one available interpretation for a parameter, it functions as a constant (like GOAT does). This may happen while the dispute over how to determine whether a goat has a good disposition is preserved. In this case, although the standard for goat quality remain contentious, there has been some conceptual refinement for the fair judges in Lorain County. Progress like this

on recalcitrant disputes is facilitated where there is clarity on which matters are sources of contention, and where these matters can be distinguished clearly from each other.

In the sensitive management hypothesis, which I describe below (§ 5.4), disputes about the nature and extension of cognition can be played out in terms of disputes about how to interpret the terms *BELONGING TO AN ORGANISM*, *SENSITIVITY TO CIRCUMSTANCES*, and *BEHAVIOR*. I contend that as a matter of description these three parameters trace the major fault lines of disagreement in the cognition border wars, just as the parameters in SELECTION trace fault lines of disagreement between theoretical biologists. These disputes, while by no means simple, may be a great deal more tractable than the problem of cognition, or at least more incrementally tractable. Making clear distinctions between sources of contention may help cognitive scientists to eventually converge on common interpretations of some of these parameters, as biologists have for *FITNESS*, though they have not for *INDIVIDUAL*.

4.5 SUMMARY

In this chapter, I described parameterization as a way to accomplish construal-based explication with more analytic clarity. In particular, the provision of multiple interpretations for parameterized predicates provides a way to state more explicitly which commitments are involved in construing an item in terms of a concept, especially where an explicans includes multiple construal-sensitive terms whose interpretations are not correlated with each other. The parameterization strategy is a framework that enables the construction of models that are ecumenically extensionally adequate. It also permits the representation of themes in conceptual norms that remain stable, though competent users of the concept may employ different, more specific criteria depending on their context of assessment. These themes are at least potentially stable diachronically, as well, across some episodes of conceptual change. Finally, I described three benefits that parameterization affords for the modeling of discourses where the norms of use for a concept are controversial:

parameterized explications localize areas of agreement and disagreement, they structure opportunities for Quinean retreat, and they afford a "divide and conquer" strategy for the resolution of recalcitrant conceptual disputes.

To review the criteria of success for a parameterized explication, then: a parameterization should at least be ecumenically adequate; that is, it should successfully reproduce the patterns of disagreement exhibited in a discourse. Where there is agreement on central cases—as was true for our examples of BRAVE, KNOWLEDGE, QUALITY GOAT, and SELECTION above, as well as for COGNITION—it should accurately classify items as paradigmatic instances, controversial instances, and agreed-upon non-instances of the explicandum concept. It is a secondary desideratum of parameterization that the parameterized predicates should correspond to concepts whose interpretations are well-traveled lines of Quinean retreat when these are apparent, and perhaps some roads not taken in the case of particularly fraught discourse where the sources of disagreement are not clear.

However it is not always apparent which differences in commitments can be invoked to explain divergent judgments. Parameterizations are analytical hypotheses about the norms for using a concept, and must be tested against the discourse they purport to model.[24] As hypotheses, parameterizations should make specific claims about the state of the discourse they represent. One way to make the claims specific is to introduce a closed parameterization, in which there is a fixed number of interpretations for each parameter, though even open parameterizations can make specific claims. Another way to sharpen the empirical purport of a parameterization is to confirm that a set of interpretations for the parameters corresponds to the correct sectarian extension for the explicandum concept. Where there is not evidence concerning the interpretation of each parameter, the number of "free" parameters should be limited as much as possible.

[24] I am reminded of Quine's language: "The method of analytical hypotheses is a way of catapulting oneself into the [object] language by the momentum of the home language. It is a way of grafting exotic shoots on to the old familiar bush… until only the exotic meets the eye" (1960, 70).

I note here, as a confession of my philosophical inclinations rather than as a matter for argument at present, that I am inclined to yield in the face of complicated and messy conceptual disputes like those that arise in theoretical biology over the criteria of individuation for *INDIVIDUAL*. The world is complicated and messy, and our best scientific concepts must reflect these complications or else fail to describe the world accurately. There is a place for simplified representations in science—for abstraction, idealization, and computational shortcuts (e.g. of the sort Mark Wilson calls "physics avoidance," Wilson MS)—and I do respect the Quinean impulse for austere formalization. But there is clarity that can be provided in contexts unsuitable for such treatment, as well. Parameterization is one format in which the project of making explicit what is implicit in our concepts can be carried out. I do not claim, of course, that every concept will yield gracefully to treatment by parameterization, and indeed those that do yield to parameterization may do so only within certain limits—for certain communities and not others, or relative to particular activities or interests, or only for certain periods of time. But for some concepts and for a particular discourse at a time, parameterized explications can outline themes that can be developed in distinct but recognizably similar forms. I contend that these conditions apply to the scientific concept of COGNITION right now, and it is to that topic that I now return.

5.0 COGNITION AS THE SENSITIVE MANAGEMENT OF ORGANISMAL BEHAVIOR

5.1 OVERVIEW

In this chapter I will introduce and defend a conceptual explication of COGNITION. I will begin by briefly reviewing the problem of cognition and the logical tools I described in previous chapters, and which I make use of in my explication (§ 5.2). I will then describe a series of cases—paradigmatically cognitive phenomena, controversially cognitive phenomena, and non-cognitive phenomena. These will serve as test cases for the explication I describe (§ 5.3). Then I will describe my explication, the sensitive management hypothesis (§ 5.4). In the last sections I will discuss the parameters of the sensitive management hypothesis, reflecting on how they bear on the test cases introduced earlier (§ 5.5) before concluding (§ 5.6).

5.2 THE STORY SO FAR

Having completed a long detour concerning conceptual explication, I am now prepared to describe a novel explication of COGNITION. Before proceeding, let us review the main considerations of the foregoing discussion. I described the problem of cognition—that despite over fifty years of cognitive science, no account of the nature of cognition has proven popular, though increased clarity about the nature and extension of cognition is desired (§ 2.1). The matter is complicated by the fact that since the 1980s, a proliferation of anti-classical perspectives has fractured the

theoretical landscape of cognitive science. These anti-classical perspectives included connectionist and situated approaches to cognitive science; the embodied, embedded, enactivist, and dynamical traditions; as well as claims that there is extended, social, plant, and microbe cognition. This proliferation of perspectives illustrates disagreement not only about the nature of cognition in paradigmatic cases, but about the extension of cognition. Where adherents of these perspectives have clashed with each other about the nature and extension of cognition, I referred to these disputes as the cognition border wars (§§ 1.2.2, 2.2). In the present century, Fred Adams and Ken Aizawa reintroduced the problem of cognition, under the guise of the search for the "mark of the cognitive" (2001).

As Adams and Aizawa would have it, the mark of the cognitive would be a set of necessary and sufficient conditions for being a cognitive item. Adams and Aizawa do not claim to have found such conditions, but in their aspirations they adhere to the traditional scheme for conceptual explication: an explicandum concept is explicated by articulating necessary and sufficient conditions (the explicans) for being an instance of the concept, where the main criterion of success is extensional adequacy (§ 2.4, 3.2). However, there are several facts about cognition that frustrate its treatment by means of the traditional scheme. First, it is widely acknowledged to be polysemous. I distinguished between "highfalutin" cognition, a scientific and philosophical analogue for though or belief, and "inclusive cognition," the subject matter of cognitive science (§ 2.3.1). Other philosophers (e.g. Prinz 2004; Orlandi 2014) have attempted to clarify the nature of highfalutin cognition, but only inclusive cognition concerns me here because it is only the explication of inclusive cognition that can deliver the four benefits of resolving the problem of cognition (Table 2.1; §§ 2.2–2.3). Nevertheless, it is not always clear that the highfalutin and inclusive senses of COGNITION are distinguished in discussions of the problem of cognition. The second fact is the fact of widespread expert disagreement (§ 2.2): there is recalcitrant disagreement between contemporary cognitive scientists about the nature and extension of inclusive cognition. Since the extension of cognition is contested, along with most of its properties, it is not clear what criterion

of adequacy should govern a conceptual explication. We cannot infer COGNITION's extension from its properties, nor its properties from its extension.

Therefore, I introduced three alterations of the traditional scheme for conceptual explication. First, I described the notion of ecumenical extensional adequacy. An explication is ecumenically adequate if it reproduces the pattern of judgments found in the discourse being modeled. In particular, rather than dividing phenomena into those that are genuinely cognitive and those that are not genuinely cognitive, it divides phenomena into those that participants in a discourse generally agree are cognitive, those that engender dissent regarding whether they are cognitive, and those that participants in a discourse generally agree are not cognitive (§ 2.5).

Second, I described a form of judgment I call "construal," and demonstrated how the notion of a construal can be used to produce conceptual explications with unusual properties. The semantic value of a construal (i.e. a judgment featuring a construal-sensitive predicate) depends not only on the situation the construal is about, but the mental states of the construer. I subsume these mental states under the heading of "context of assessment," following MacFarlane (2014) and as a special case of his notion. If the construal-sensitivity of terms in an explication is acknowledged, then putative counterexamples to an explication can only be judged relative to a context of assessment. A construal-sensitive explication of COGNITION categorizes phenomena into those that are cognitive and those that are not relative to a context of assessment, rather than categorizing phenomena into those that are cognitive and those that are not simpliciter. This relativity to a context of assessment makes it possible for a construal-based explication to be ecumenically adequate, since it produces different extensions for different judges, and can therefore capture states of disagreement (§ 3.3).

Third, I introduced parameterization as a way to make construal-based explications more analytically precise, and in order to make the ecumenical potential of construal-based explication more determinate. In parameterized explication, construal-sensitive predicates in the explicans can be assigned various interpretations that have different inferential properties. When each parameter is assigned a particular interpretation, the explicans has a particular extension, called a sectarian

extension. Different assignments of interpretations to parameters can result in in different sectarian extensions. This structure allows us to reproduce an extension of the form demanded by ecumenical adequacy. In a parameterized explication of COGNITION, phenomena that belong to all or most sectarian extensions may be considered paradigmatically cognitive phenomena. Phenomena that belong to some sectarian extensions but are not paradigmatic may be considered controversially cognitive phenomena. Phenomena that belong to no sectarian extensions may be considered agreed-upon non-cognitive phenomena (§ 4.2).[25] Thus, parameterization is a form from which we can cast a Lewisian foundation (q.v. § 1.4.2).

5.3 TEST CASES

I will now proceed to describe a series of cases that shall serve as test cases for my explication of cognition. Each of these cases is either a paradigmatically cognitive phenomenon, a controversially cognitive phenomenon, or is generally agreed not to be cognitive phenomenon. With the target thus fixed firmly, I will proceed to introduce my explication, the sensitive management hypothesis, and show that it correctly classifies the cases.

5.3.1 Paradigmatically cognitive phenomena and non-cognitive phenomena

Despite the entrenched controversies of the border wars, there are a number of cases that engender agreement among cognitive scientists concerning whether they involve cognitive phenomena. For example, Inga's capacity to navigate a city by biological, brain-based memory is an uncontroversially cognitive phenomenon (cf. Clark and Chalmers 1998). The capacities to see color, recognize pitches, feel disgust, and control one's limbs are also paradigmatically cognitive

[25] But see note 23 (p. 79 above) for a complication that I ignore in this document.

phenomena, at least when these capacities belong to animals that are not using external aids or modern technology.

The human capacity to visually recognize particular faces is also a cognitive phenomenon. It is universally agreed that cortical structures in the brain play a crucial role in explaining this capacity, and in particular a region of cortex called the fusiform face area (FFA) which exhibits high activation selectively when a subject is viewing faces (see e.g. Kanwisher 2010), as measured by functional magnetic resonance imaging (fMRI). Lesions in FFA are also implicated in pathological deficits in face-recognition (e.g. Barton et al. 2002). It is not clear precisely how FFA contributes to capacities for face-recognition, and there is some dispute as to whether the function of FFA is face-recognition specifically or whether it is involved in any or most capacities for expert visual recognition (Haxby et al. 2001). Nevertheless, there is little dispute that it is crucial for face-recognition.

In addition, there are many examples of phenomena that cognitive scientists agree are not cognitive. A rock warming in the sun, for example, is not a cognitive phenomenon. Nor is the erosion of a stream bed, or the relationship between the time it takes for light to reach the earth and the distance of the light's point of origin, nor is the way light plays on the floor when it comes in through the windows on a summer afternoon.

5.3.2 Controversially cognitive phenomena

There are many cases that fall under the heading of "controversially cognitive phenomena," though I will briefly describe only six here which should illustrate their variety. Each of these claims is associated with theoretical frameworks and empirical research programs, and has vigorous proponents and opponents, both in philosophy departments and in other disciplines (though I will largely cite philosophers in what follows, since philosophers tend to be the most articulate about abstract commitments).

Table 5.1: Test cases for an explication of COGNITION.

Test cases for an explication of COGNITION

A. paradigmatic cases (§ 5.3.1)	A1. urban navigation in Inga
	A2. visual face recognition with FFA
B. non-cases (§ 5.3.1)	B1. rock heating in the sun
	B2. erosion in stream beds
C. controversial cases (§ 5.3.2)	C1. group cognition in ants
	C2. robot cognition in Herbert
	C3. extended cognition in Otto-notebook
	C4. embodied audition in crickets
	C5. off-line phototropism in *Lavatera cretica*
	C6. magnetotaxis in bacteria

First, proponents of *group cognition* contend that the coordinated activity of multiple corporeal agents constitutes a kind of cognition over and above individual cognition (see Hutchins 1995). Since ants perceive, navigate, &c., they are generally understood to have cognitive processes. However, the behavior of individual ants is partly regulated by policies that promote the good of the whole colony, e.g. the use of pheromone markers that promote performances the efficient acquisition of food resources, the adaptive layout of nests, the regulation of aggressive behaviors, &c. Indeed, colonies sometimes thrive precisely because of the behavioral variability of individual ants (Modlmeier and Foitzik 2011). Since ant colonies are capable of distinctive performances whose adaptive character seems to exceed the adaptive character of the behaviors of individual ants, they are a simple case of putatively group cognition.

Second, Rodney Brooks designed an autonomous robot called Herbert that once inhabited the MIT robotics lab. Herbert moved on its own, avoided obstacles, and sought unattended soda

cans which it would collect at a central point (Brooks, Connell, and Ning 1988; Brooks 1991).[26] Arguments about the possibility of *robot cognition* are as old as cognitive science and older (Turing 1950; Putnam 1967a), but cognitive traditionalists insist that Herbert has no genuine cognitive processes. Maybe some machine could think, but Herbert is a *mere* machine, since its behavior is explained by appeal to Brooks' reasons rather than by appeal to its own reasons (Adams and Garrison 2013).

Third, Otto is a fictional man whose memory is impaired by Alzheimer's disease. Otto carries a notebook with him wherever he goes. When he learns or decides something, he writes it in the notebook. He also consults his notebook constantly in the course of his day. Thus Otto uses his notebook to augment the shortcomings of his biological memory. Proponents of *extended cognition* hold that Otto's cognitive processes occur in part outside his skin, in the notebook (Clark and Chalmers 1998); opponents argue that while the notebook must be involved in the explanation of Otto's behavior, the notebook itself is not part of a cognitive mechanism (see e.g. Adams and Aizawa 2001; Rupert 2004).[27] I also discuss this case in greater detail in the Appendix.

Fourth, proponents of *embodied cognition* argue that cognition occurs outside of brains in bodies (Varela, Thompson, and Rosch 1991; Gibbs 2005; Chemero 2009). For example, female crickets have a tympanum (eardrum) on each foreleg, connected to spiracles (openings on its body) via tracheal tubes. The spiracles are spaced like openings on a woodwind so as to resonate with the mating calls of conspecifics. Sounds at a particular frequency are especially amplified on the side of the cricket nearer to the call, and diminished on the side farther from the call. Thus, healthy female crickets are capable of unusually precise egocentric spatial discrimination regarding the locations of potential mates, based on their calls (Webb 1994). Proponents of embodied cognition

[26] Brooks does not defend Herbert as a case of genuine cognition per se, but he does suggest that cognition (or "intelligence") need consist in nothing more than Herbert's capacities, produced by Herbert's mechanisms. And since Adams and Garrison take Herbert up as a case, and since Herbert is an oft-discussed example, I include Herbert in this list.

[27] In Clark and Chalmers' original paper, Otto is an example of the extended *mind* hypothesis, not of extended cognition per se. However, the example of Otto is nevertheless the most discussed example in the extended cognition literature.

claim that the network of spiracles and tracheal tubes are components of the cricket's cognitive system, whereas opponents argue that the cognitive action, so to speak, is entirely in the cricket's neural tissue.

Fifth, proponents of *plant cognition* maintain that plants behave, though at longer time scales than animals, and that this behavior is coordinated by cognitive processes just as animal behavior is. John Haugeland (1981a) once supposed that if there were such a thing as a "super-sunflower" that could track the direction of the sun in the absence of photostimulation, it would be an organism with cognitive capacities. Francisco Calvo Garzón (2007) argues that there are indeed such plants. For example, *Lavatera cretica* orients expectantly toward the sun even when deprived of sunlight for several days. Proponents of plant cognition argue that this and other complex, adaptive plant behaviors are evidence that there is plant cognition.

Sixth, there are even defenders of *microbe cognition*. Perhaps the most familiar example is one made famous by Fred Dretske (1986). There are anaerobic bacteria that possess organelles called magnetosomes which respond differentially to magnetic fields, including the Earth's magnetic field. Bacteria that use this sensitivity to direct their locomotion, e.g. away from oxygenated water, are called "magnetotactic." Some enactivists argue that all living things possess cognitive capacities (Thompson 2010), and the case of magnetotactic bacteria is an example with intuitive appeal.

These are six contentious cases of cognition: group cognition in ants, Herbert the autonomous robot, the extended Otto-notebook system, the embodied cricket auditory system, plant phototropism, and bacterial magnetotaxis. These cases do not represent all of the disputes of the border wars—in particular, they do not represent disputes that concern the proper method of modeling or explaining cognition, e.g. disputes over dynamicism—but they illustrate a substantial amount of the flavor and variety of controversial proposals.

5.4 THE SENSITIVE MANAGEMENT HYPOTHESIS

Even with these cases in hand, however, the correct explication of COGNITION cannot simply be "read off" from the patterns of disagreement exhibited in the border wars. Explications are analytical hypotheses; they are risky proposals and must be tested for adequacy against cognitive scientific discourse. I contend that these (and other) cases can be accommodated by the sensitive management hypothesis, an ecumenical, construal-based, parameterized explication. The slogan for this view is that *cognition is the sensitive management of organismal behavior*. More precisely, the view is that

> SMH:
>
> Cognition is the operation of mechanisms, where
>
> 1. the mechanisms *belong to an organism*,
> 2. the operation of the mechanisms is *sensitive to the organism's circumstances*, and
> 3. the mechanisms manage the *behavior* of the organism.

Of course, this explication requires some unpacking. SMH contains three parameters, italicized above for perspicuity: *BELONGING TO AN ORGANISM* (or *ORGANISM*), *SENSITIVITY TO CIRCUMSTANCES* (or *SENSITIVITY*), and *BEHAVIOR*. When these parameters are assigned interpretations, the explication determines a sectarian extension. With different assignments of interpretations to these parameters, the sectarian extensions of this formula variously contain or exclude each of the six test cases described above. I will discuss the parameters in more detail in the next section; first I will unpack some unparameterized features of SMH.

The sensitive management hypothesis can be used to define a number of other specific expressions. *Cognitive mechanisms* are mechanisms that satisfy conditions 1–3. *Cognitive processes* are the component activities and operations of cognitive mechanisms. *Cognitive states* are functional states of cognitive mechanisms. A *cognitive system* is an exhaustive collection of

cognitive mechanisms belonging to a particular organism. *Cognitive behaviors* are behaviors that are managed by cognitive mechanisms. *Cognitive capacities* are capacities to exhibit cognitive behavior, or capacities of components of cognitive mechanisms, and so on. As an illustration of these definitions, consider visual face-recognition. The mechanism through which humans recognize faces—crucially involving FFA—is a cognitive mechanism. The operation of that mechanism can be said, abstractly, to be "cognition," and specifically to be a cognitive process (or cognitive processes). States of that mechanism (e.g. a representation that identifies a face as Nicola Sturgeon's face) are cognitive states. That mechanism together with all the other cognitive mechanisms belonging to the same organism is a cognitive system. The mechanism is sensitive to circumstances of the organism (e.g. the visual presence of faces) and manages the organism's behavior (by causing the organism to behave differently depending on whether faces are present or absent, and depending on which faces if any are present). Such behavior is cognitive behavior. The capacity to recognize faces is a cognitive capacity, as are the capacities to exhibit the behaviors managed by the FFA mechanism, as are the capacities of the mechanism and its components. These formulations are generally consistent with the judgments of most cognitive scientists, e.g. that "cognition" does not refer to any behavior itself, but to the processes that produce it (Shapiro 2013; Aizawa forthcoming).

To "manage" behavior as I mean it here means to cause a pattern of behavior, i.e. certain individual behaviors in some circumstances and other individual behaviors in others. (The circumstances in question are the same circumstances to which cognitive mechanisms are sensitive.) The disposition to simply cause motion or articulation of the body is not sufficient to qualify as the management of behavior, unless it is a disposition to cause such motion differentially relative to circumstances. For example, path integration in ants is a cognitive phenomenon in which the ant's locomotion is managed. By this I mean that some bodily movements are caused rather than others, and the difference is ascribable to the ant's circumstances (e.g. its destination being at such-and-such a bearing, or such-and-such a distance away). A muscle spasm, by contrast, is an unlikely candidate for a behavior, if it cannot be explained by appeal to any adaptive character. The

explanation of spasms is furthermore not normally considered to be a proper subject of cognitive science, rather than of physiology or some other branch of the sciences. The on-line control of motor processes, like grasping performances in humans, is a well-studied topic in cognitive science (see e.g. Jeannerod and Biguer 1982; Jeannerod 1984). In the case of on-line control of grasping, the behavior in question is not the mere contraction of muscles, but the guidance of these processes in response to perception and proprioception.

Second, I describe cognition in terms of "mechanisms," in line with recent trends in philosophy of science. The new mechanists (Machamer, Darden, and Craver 2000; Bechtel 2008; Bechtel and Abrahamsen 2005; Craver 2007) claim that the main explanatory aim of many special sciences is not to articulate universally applicable, exceptionless natural laws (the nomological view, q.v. Appendix § A.3), but to describe naturally-occurring structures—mechanisms—that exhibit phenomena we want to explain. I do not mean to undertake all of the commitments of the new mechanists, if indeed there are commitments they all share (see Levy 2013 for discussion). Instead, I endorse the trend of abandoning nomological views of explanation for cognitive science, and I endorse the claim that one of the most prominent explanatory practices of cognitive scientists is to describe natural structures called mechanisms, whatever their nature (See Appendix, below). These more modest commitments are undertaken by some philosophers of science who do not identify as new mechanists (e.g. Cummins 2000).

A cognitive mechanism, for my purposes, is a structure of component entities and component operations that are organized such that they manage behavior (adapted from Machamer, Darden, and Craver 2000; Bechtel and Abrahamsen 2005; Craver 2007). The component entities that figure in cognitive mechanisms are things like representations, modules, brain areas, populations of neurons, or idealized "neurons" in artificial neural networks. Characteristic operations of cognitive mechanisms include processing operations on or between those entities: transformations of representations, computational interactions between modules and brain areas, activation and inhibition of neuron populations, and interactions between artificial neurons as specified by connection weights. The organization of these entities and operations into

mechanisms is usually represented by graphs, but can be specified more or less completely by equations or descriptions of relations between components.

Neither "mechanism" nor "management" is a parameter of SMH. While the correct interpretation of these terms certainly admits of controversy, that controversy is limited to philosophical contexts. For example, the disagreement about the nature of mechanisms is primarily disagreement among philosophers about what scientific practice is or should be, not disagreement between cognitive scientists about the merits of particular research programs. The existing practice and research goals of scientists are the standards against which various accounts of mechanism are to be judged (see e.g. Chirimuuta 2014; Ross 2015). At any rate, I am not convinced that differing interpretations of "mechanism" correlate with different judgments concerning the extension of COGNITION. More expressions could be parameterized on a finer-grained view that aimed to articulate points of controversy beyond those of the border wars, but since explanatory power is inter alia a result of effective parameter reduction, I have limited the number of parameters as much as possible (see § 4.3.3).

I submit that the disagreements of the border wars, at least those about the extension of COGNITION, can be gathered under three headings: which critters that are candidate possessors or beneficiaries of cognitive processes, which ways cognitive processes must exhibit sensitivity to their circumstances, and which if any of the critter's performances count as behaviors. The questions left open by these headings are subsumed under three open parameters: *BELONGING TO AN ORGANISM*, *SENSITIVITY TO CIRCUMSTANCES*, and *BEHAVIOR*, respectively. Thus, if it is successful, SMH accomplishes the first of the three benefits I described for characterizing controversy (§ 4.4): it distinguishes points of agreement and disagreement by expressing matters of disagreement as parameters.

In the previous chapter I claimed that parameterization could, in addition to characterizing a synchronic state of disagreement, illuminate processes of conceptual change. Specifically, I claimed that a parameterized explication can express a theme that is developed in particular assignments of interpretations to its parameters, and this theme might survive the process of

analytic prolongation if the parameterization is diachronically robust; i.e. if it is adequate across periods in which the norms of use for the explicandum concept change (§§ 4.2.2, 4.4). In the case of COGNITION, the demonstration of this added value is straightforward, since more traditional (i.e. pre-1980s) conceptions of cognition are conserved in contemporary discourse as the most restrictive contemporary conceptions. If it is shown that SMH has one sectarian extension that includes only the paradigmatically cognitive phenomena, this is strong evidence that SMH is robust between pre-1980s cognitive science and the present. (This demonstration is open to the objection that at some point between the 1980s and the present SMH would not have been adequate to cognitive scientific discourse, and I welcome such empirical challenges, but I will not engage in a fine-grained historical study here.)

5.5 REPRODUCING PATTERNS OF EXPERT DISAGREEMENT WITH PARAMETRIC VARIATION

I will now discuss each of the parameters in turn, describing how the parameters can be manipulated in order to selectively include or exclude various controversial cases, while including paradigmatic cases and excluding non-cases. In general, I will begin by considering relatively restrictive interpretations of parameters, discussing more liberal interpretations in turn. As I go, I will remark on SMH's potential to achieve the three advantages I described above for characterizing controversy (§ 4.4; see also § 7.4). In particular, I will be concerned to show that these parameters subsume topics of disagreement.

5.5.1 Belonging to an organism

Cognitive processes, states, mechanisms, &c. are not generally thought to float free; they belong to or subserve some critter. All parties to the border wars seem to agree that it is objectionable to

ascribe "un-owned" cognitive processes (Adams and Aizawa 2001; Rupert 2009; Clark 2010a; Rowlands 2010), though only Rowlands seems to discuss the details of "belonging" to a subject at length (2010, Ch. 6). Rupert calls the owner or container of a cognitive process a "cognitive system." Rowlands calls it a "representational subject." Let us call that critter, whatever its nature, an "organism." Appeal to particular interpretations of *ORGANISM* is a less-traveled line of Quinean retreat than appeal to *SENSITIVITY* (q.v. below § 5.5.2), but there are some well-discussed examples, particularly in the literature on extended cognition.

The most conservative intuitions about the extension of COGNITION attribute cognitive processes only to animals. However, such an interpretation for *ORGANISM* is almost certainly too restrictive. Writers like Adams and Garrison who deny plant cognition and robot cognition do not believe in a magic cognitive power of animal systems, or a magic branch of the phylogenetic tree. They deny that plants and robots have cognitive processes because of some other difference between those systems and animals—in their case, because plants and robots cannot act for reasons (see § 6.2.1 for discussion). Moreover, they suggest that there could conceivably be robots with genuinely cognitive processes, if there were robots that were capable of acting for reasons (Adams and Garrison 2013). Thus interpreting *ORGANISM* as ranging only over animals is excessively restrictive, for it does not correspond to a dimension of disagreement among cognitive scientists about COGNITION. Such a narrow interpretation is not supported by a charitable interpretation of cognitive scientists' practice of ascribing the concept COGNITION or their justifications for such ascriptions. (Although John Searle might have suggested an interpretation like this when, in his famous criticism of functionalist cognitive science, he suggested that there is something special about the causal properties of brains; Searle 1980. However, this may be a misreading of Searle).

A more plausible conservative interpretation of *ORGANISM* restricts suitable organisms to biological individuals, on whatever basis justifies the biological usage. Thus humans are cognitive organisms, whereas stones and stream beds are deemed unsuitable subjects of cognitive processes, along with hunks of iron, tectonic plates, mountains, &c. Furthermore, nothing that happens in Herbert the robot can be a cognitive process of Herbert's, since Herbert is not a biological creature

(Table 5.1 C2). The common view that non-animal organisms do not have cognitive processes is usually supported by appeal to the different representational capacities of animals and other organisms. However, even this interpretation leaves room for variable interpretation and negotiation. For example, Ellen Clarke characterizes biological individuals as those entities that possess both (1) mechanisms that inhibit natural selection between their parts, and (2) mechanisms that increase or maintain natural selection between them and other entities (Clarke 2013). Both of these requirements break down and become difficult to adjudicate in a variety of problem cases. Clarke discusses several fascinating examples, but a more familiar example is that of ant colonies, which are composed of individual ants that have separate bodies but share a common reproductive fate. Individual ants develop from single cells, creating a genetic-developmental bottleneck so that the ant's cells share a common genome, thus inhibiting within-ant selection. By contrast, the corporeal separation between ants means that ant individuals with different traits may have differential rates of survival, thus maintaining some between-ant selection. However, new ant colonies are formed by individual queens, who create a second genetic bottleneck at the level of the colony (thus inhibiting within-colony selection), and distinct ant colonies compete with each other for resources (thus maintaining between-colony selection). Further complicating matters, in some species (like the yellow crazy ants, *Anoplolepsis gracilipes*, that have invaded Christmas Island) colonies cooperate, forming supercolonies. Thus, even this conservative interpretation of *ORGANISM* inherits some construal-sensitivity from theoretical disputes in biology. This construal-sensitivity may be reduced by appealing to Peter Godfrey-Smith's (2012a) notion of BIOLOGICAL ORGANISM, understood in contrast to his notion of a DARWINIAN INDIVIDUAL (see § 4.3.2), or a blunter interpretation that requires organisms to have continuous bodies composed of organic compounds and that maintain a controlled internal environment through established homeostatic mechanisms.

If cognitive processes are attributable either to individual ants or to ant colonies, then an explication that parameterizes *ORGANISM* in this way is in a position to explain at least some variation in expert judgment concerning the existence of group cognition, the first of our six

examples (Table 5.1 C1). That is, there can be cognitive organisms that are composed of other cognitive organisms, and whose activity is mediated by cognitive processes that consist in the coordinated activity of its component organisms. Furthermore, the complexity of this interpretation offers direction in how to conduct Quinean retreat. Disagreements about whether the behavior of an ant colony is a cognitive phenomenon can play out, in some circumstances, through arguments concerning whether a colony is a suitable organism in the biological sense. Such disagreements might turn on appeal to Clarke's explication of *ORGANISM*, or they might involve appeal to other explications (Table 5.2 α, β).

The case of Otto and his notebook is also complicated. Otto is a biological organism, so the interpretation of *ORGANISM* as denoting biological organisms does not exclude cases of extended cognition, unless there is a more precise restriction on what it is to *belong* to an organism. For example, an organism might be said only to possess those mechanisms that are parts of its body, or that it possesses normally (Table 5.2 α, β). However, if Andy Clark is to be believed, extended cognition is a normal and this second criterion does not rule it out (2008). Robert Rupert suggests that the only parts of a cognitive system are those that are realized by token physical structures that are integrated into the body (2004, 37–40; 2009). However his take on extended cognition is difficult to discern precisely from the literature; he takes a regularity criterion to exclude most putative cases of extended cognition, though perhaps not the case of Otto). A more liberal criterion like Clark's might identify any mechanisms subserving the goals of an organism as mechanisms belonging to the organism, regardless of whether they are parts of the organism's biological body (Table 5.2 γ).

Table 5.2: Selected interpretations of *BELONGING TO AN ORGANISM*.

Selected interpretations of *BELONGING TO AN ORGANISM*	
α. corporeal elements of biological organisms	β. corporeal elements of Darwinian individuals
γ. mechanisms subserving Dawinian individuals	δ. mechanisms subserving systems apt for goal-ascription

An even more liberal interpretation of *ORGANISM* might require only that organisms be any critters to which we attribute goals—perhaps just survival and reproduction, or perhaps other goals such as constructing shelter, disabling yonder antelope, or following an experimenter's instructions—and mechanisms that promote the achievement of those goals (Table 5.2 δ). What matters here is that we attribute goals to a system, and that there be mechanisms that promote those goals, whatever biologists may say about these systems. This, like Dennett's intentional stance, is an ineliminably construal-based criterion for identifying organisms (see § 6.2.1). On this more liberal interpretation, Otto's notebook may realize some of his cognitive processes since its properties contribute to his successful achievement of his goals (Table 5.1 C3). Herbert the robot may also count as an organism so long as we attribute goals to Herbert, like collecting soda cans and avoiding collisions, and so long as Herbert has mechanisms that promote these goals (Table 5.1 C2). Since we do not attribute goals to rocks or stream beds, they are not cognitive organisms even on this liberal interpretation.

Since ant colonies, Herbert, and Otto's notebook (Table 5.1 C1–3) are in the gap between this more liberal interpretation and the more conservative biological/realizer-based interpretation of *ORGANISM*, they are ruled controversial cases on the basis of the *ORGANISM* parameter (assuming they satisfy the rest of the explication on at least one assignment of interpretations to other parameters). That is, they are in the sectarian extension of COGNITION on some interpretations of *ORGANISM*, and not on other interpretations. The brain of an undergraduate in

a psychology experiment, by contrast, is in the sectarian extension of COGNITION on all interpretations of *ORGANISM*: it is a token physical component of her body that houses mechanisms that promote her goals, whether we construe her as a biological individual or as a subject of goals and mechanisms. So the undergraduate is a paradigmatic cognitive organism, and her brain is a paradigmatic realizer of her cognitive processes. A hunk of iron satisfies none of the admissible interpretations of *ORGANISM* mentioned above. It is not a biological organism, does not undergo natural selection, and is not an entity to which we attribute goals or mechanisms subserving goals. It is therefore not a candidate possessor of cognitive processes.

5.5.2 Sensitivity to circumstances

Cognitive mechanisms must be sensitive to the circumstances of an organism. I do not intend *SENSITIVITY* to be understood in the sense generally meant by epistemologists (cf. Nozick 1981), but in the more colloquial sense of tracking selected states of affairs, perhaps imperfectly, but with non-accidental regularity. The language of "sensitivity" is intended to include accounts that rely on some conception of representation, as well as accounts that deny that "representation" so-called is necessary for cognition (e.g. Brooks 1991; van Gelder 1995; Chemero 2009). Even anti-representational accounts acknowledge that cognition involves sensitivity to an organism's circumstances; they simply deny that the forms of sensitivity we discover when we examine cognitive systems are always happily called "representation." Appeals to representation and its analogues are the most common strategy of Quinean retreat in the discourse of the border wars, and the *SENSITIVITY* parameter is meant to subsume most disputes that turn on such appeals.

I understand "circumstances" here to refer to states of affairs that bear on an organism's goals, either by being consistent with those goals, inconsistent with those goals, or by being such that an organism must modulate its behavior according to them in order to pursue its goals. Organisms are generally sensitive only to a proper subset of their circumstances. Sensitivity is paradigmatically achieved through perceptual, interoceptive, or inferential processes. However, it

Table 5.3: Selected interpretations of *SENSITIVITY TO CIRCUMSTANCES*.

Selected interpretations of *SENSITIVITY TO CIRCUMSTANCES*	
α. classical representations with non-derived content	β. classical representations
γ. pushmi-pullyu representations	δ. operation of adaptive mechanisms

is not the case that only perceptual, interoceptive, and inferential processes exhibit sensitivity to circumstances. It is sufficient for a mechanism's operation to be modulated distally according to such sensitivities. For example, "place cells" in parahippocampal cortex respond differentially to an organism's position in space (O'Keefe and Dostrovsky 1971), though they are rarely reckoned to be part of a perceptual system in the way that cells in visual cortex are. Nevertheless, their activity is sensitive to the organism's circumstances. This sensitivity is of course mediated by the connections between place cells and other brain regions.

5.5.2.1 Sensitivity as representation The most traditional understanding of *SENSITIVITY TO CIRCUMSTANCES* is that a mechanism is sensitive to an organism's circumstances only if it *represents* those circumstances (whatever "represents" means). For example, consider the FFA and its role in visual face-recognition (§ 5.3.1). Since patterns of neural activity in FFA are thought to be involved in facial recognition, it is often said to represent faces (Kanwisher 2010, for example, uses representational language freely). Representations typically involve changes in a representing medium—e.g. the pattern of neural activity in FFA—that reflect changes in the thing it represents—e.g. different faces. The pattern of activity in FFA tracks various states of affairs, and humans are capable of recognizing faces in virtue of these changes.

Activity in FFA is furthermore taken to be a case of representation with "intrinsic" or "non-derived" content—the sort of content had by mental representations but not by conventional

representations like printed words (Table 5.3 α). Representations with merely derived content are supposed to be assigned content in virtue of representations with non-derived content, like mental representations (see also § 6.3.1). The stipulation that cognitive sensitivity be cashed out in terms of representations with non-derived content is controversial, but nevertheless figures in many accounts of cognitive representation (e.g. Searle 1980; Adams and Aizawa 2001; Rowlands 2010). Defenders of extended cognition, of course, deny that cognitive sensitivity must be understood in terms of representations with non-derived content, and tend to be suspicious of the very notion (e.g. Clark 2005). Clark and Chalmers claim that Otto's memories are constituted in part by the state of his notebook, which is understood to have merely derived content. By denying that cognitive representations must have non-derived content, Clark and Chalmers make space in their conception of cognition for Otto's notebook to serve as part of a cognitive mechanism (Table 5.3 β).

5.5.2.2 Marginal cases of representation Classical conceptions of sensitivity in terms of representation, whether they involve non-derived content or not, generally impose further requirements on cognitive sensitivity. I mention a few here not to argue for any particular conception of REPRESENTATION, but merely to highlight the great variety of marginal cases encountered in the literature on representation and to highlight the degree to which the concept REPRESENTATION is embattled. It is in part because the concept of REPRESENTATION is so contentious that I choose to avoid using the term in SMH.

The sensitivity of a magnetotactic bacterium to its circumstances, for example, is not very robust; it takes magnetic fields as a proxy for the oxygen gradient of its environment, and this contingent relationship sometimes fails to hold (C6). Dretske describes an example in which ocean-dwelling bacteria are carried across the equator in the ballast tanks of modern ships. Once in the wrong hemisphere, they are misled by their magnetosomes and drive toward oxygen-rich water (and doom). Of course, most perceptual systems are maladapted for some range of scenarios, so it would be excessively restrictive to admit only interpretations of *SENSITIVITY* that never mislead in the way that bacterial magnetosensitivity does. A more plausible way to restrict interpretations of

SENSITIVITY so that they exclude the bacterial case is to require that cognitive mechanisms be sensitive to the possibility that they are maladapted or maldeveloped, perhaps by the comparison of different sensitivities. For example, airline pilots exhibit sensitivity to the fact that their vision and vestibular sense can mislead them about the orientation of an airplane, so they learn to trust their instruments. Call the sort of sensitivity exhibited by the airline pilot *integrative sensitivity*. The magnetotactic bacteria fail to represent their circumstances on conceptions of REPRESENTATION on which representations properly so-called are necessarily integrative. (Of course, not all conceptions require this; e.g. the teleosemantic notion of Millikan 1984). However, Calvo Garzón argues that plants do generally exhibit integrative sensitivity, coordinating information from different sensory organs in their root tips in order to exploit the soil most effectively (2011).

Jerry Fodor (1985) seems to suggest that "central" cognition must exhibit *globally integrative sensitivity*; that is, cognitive processes must be sensitive to information acquired via many or all perceptual modalities and inferential capacities. The beliefs of airline pilots exhibit globally integrative sensitivity, since they may be revised in light of any information available to the pilots, whereas commitments acquired through subconscious, automatic processes, e.g. an availability heuristic, are not. Availability heuristics are integratively sensitive, since they integrate information acquired through any sensory modality, but they are not sensitive to all information available to the subject, since they are insensitive to declarative beliefs. In general, "modular" cognitive processes do not exhibit globally integrative sensitivity. However, many paradigmatic instances of human cognition are not globally integratively sensitive. Visual illusions like afterimages and the Müller-Lyer illusion, which persist even when a subject has knowledge of the illusion and knows she is experiencing it, are failures of globally integrative sensitivity, yet are paradigmatically cognitive phenomena. Fodor's requirement, then, is at best a requirement for highfalutin cognition (which is probably what he means by "central cognition"), but too strong a requirement for even a conservative conception of inclusive cognition.

Another well-known failure of integrative sensitivity is found in the familiar story of the *Sphex* wasp, described by Henri Fabre and introduced to philosophical audiences by Dennett

(1984). The wasp is said to perform a highly stereotyped sequence of behaviors when it brings prey to its nest; if the wasp's sequence of behaviors is interrupted in certain ways, the wasp can be tricked into repeating a part of the sequence in an endless loop (this story may well be apocryphal, however; see Keijzer 2013). These behaviors fail to exhibit *ontogenetic adaptation*, i.e. management in light of learning, skill acquisition, or tuning that take place during the life of an organism rather than during the evolution of its organism-type. Adams and Garrison plausibly endorse an interpretation of *SENSITIVITY* that requires ontogenetic adaptation (see their 2013, pp. 347f). Most cognitive scientists would agree that these wasps possess cognitive capacities—they perceive, navigate, &c.—but perhaps stereotyped behavioral loops like this are merely elaborate, non-cognitive reflexes. In that case, while *Sphex* wasps would be cognitive systems in virtue of their other cognitive capacities, the mechanisms that manage the wasps' prey-caching behavior would not be cognitive mechanisms so long as *SENSITIVITY* is assigned an interpretation that requires ontogenetic adaptation. However, this demand does exclude many other paradigmatic cases of insect cognition such as honeybee dance or path integration in desert ants, since these capacities do not seem to be learned.

Some conceptions of REPRESENTATION require that representations manifest in the absence of their object. Haugeland's super sunflower represents the direction of the sun, because it can orient its face toward the sun even when the sun's position in the sky is obscured. Actual sunflowers, however, do not represent (in this sense) because they are only sensitive to the sun's position in the presence of a solar stimulus. Since *Lavatera cretica* (C5) behaves like a super sunflower, however, it does satisfy the requirement for representation-in-absence. Indeed, Calvo Garzón defends the claim that *Lavatera* has cognitive processes on precisely this basis, invoking the premise that "The sine qua non of representation-based competency is off-line adaptive behavior" (Calvo Garzón 2007, 210f), where by "off-line adaptive behavior" he means performances that reflect sensitivity in the absence of the relevant stimulus. Thus, Calvo Garzón claims that some plants satisfy this restrictive interpretation of *SENSITIVITY* as representation-in-absence.

5.5.2.3 Permissive interpretations of *SENSITIVITY* A more permissive interpretation of sensitivity to circumstances includes classical representations as well as *pushmi-pullyu representations* (Millikan 1995). These are representations that are primitively both descriptive and "directive," registering a state of the environment and causing appropriate behavior, without being part of an all-purpose representational system like human natural language. Pushmi-pullyu representations need not be integrative, and may or may not represent in absence of a stimulus (Table 5.3 γ). The magnetotactic bacterium (C6) exhibits sensitivity of this sort, so that the state of its magnetosomes alters in response to the local magnetic field and in turn alters the direction of the bacterium's locomotion, apparently without being sensitive to other circumstances of the bacterium or affecting its behavior in other ways. Of course, conditions of an explication interact with each other. Consider a full assignment of parameters to the explication above on which the *SENSITIVITY* parameter permits pushmi-pullyu representations as sensitivity enough for cognition. Such an assignment will only license the ascription of cognitive processes to magnetotactic bacteria if, on that assignment, the *BEHAVIOR* parameter is interpreted so as to include some activities of bacteria.

A still more modest interpretation of *SENSITIVITY* allows a mechanism to be sensitive to circumstances in case it is adapted to help an organism fulfill its goals in response to features of the organism's environment, even its configuration does not change in response to stimuli. For example, whereas the neurons in the FFA change their firing activity in response to faces, the structure of tracheal tubes in cricket forelegs does not change in response to auditory stimuli (C4). Nevertheless, proponents of embodied cognition argue that they are parts of an information-processing mechanism subserving cricket audition, in virtue of their special sensitivity to the frequency of sounds and the relative location of their source. Many would be uncomfortable calling such sensitivity representational (e.g. William Ramsey 2007), but embodied cognition theorists take such sensitivities to be sensitivity enough for cognition (Table 5.3 δ). However, stones and stream beds do not meet this minimal criterion, as they do not have goals.

Each of these interpretations of the *SENSITIVITY* parameter corresponds to avowed commitments of various researchers. The variation in these commitments mirrors complementary variations in which phenomena the researchers take to be cognitive phenomena. If SMH is an adequate ecumenical explication of COGNITION, then this is precisely the relationship we would expect to see. This is not a deductive argument for the sensitive management hypothesis, but it illustrates the relationship that parameterized terms (like *SENSITIVITY*) have to variations in willingness to ascribe the explicandum (i.e. COGNITION). Furthermore, variation in interpretation of the *SENSITIVITY* parameter incorporates variation in scientists' commitments about whether cognition is representational, and which varieties of representation (or near-relatives, as in the case of the operation of adaptive mechanisms in cricket audition) are necessary for cognition. Thus the *SENSITIVITY* parameter also traces a well-established line of Quinean retreat. Similarly, by interpreting the other parameters of the sensitive management hypothesis we can account for the variation in expert judgment about the other examples.

5.5.3 Behavior

Finally, cognitive mechanisms must manage the behavior of organisms. The sensitive management hypothesis requires only *mediated* management of behavior, just as it requires only mediated sensitivity to circumstances. Thus, while the on-line control of grasping movements is the management of behavior, I take it that face-recognition in FFA also manages behavior, since the recognition of a face (e.g. as Sanna's rather than Sam's) has effects that manifest as differences in behavior (e.g. uttering "Hi, Sanna," instead of "Hi, Sam"). In contemporary scientific practice, a mechanism's effect on behavior is less often an object of attention than its sensitivity. Nevertheless, I argue that such a connection is presumed for any putatively cognitive process, and the discovery that a process has absolutely no effect on behavior would imply that it is not a cognitive process. For example, consider a honeybee hive construed as a cognitive organism, much like our earlier example of the ant colony as an instance of group cognition. It was controversial for some time

Table 5.4: Selected interpretations of *BEHAVIOR*.

Selected interpretations of *BEHAVIOR*	
α. some performances of animals	β. some performances of robots
γ. some performances of plants	δ. some performances of microbes

whether foraging honeybees are guided by the waggle dances of their sisters (Munz 2005). Despite the robust correlations between the properties of bee dances on the one hand, and the relative positions of the hive and a discovered food source on the other, it was not always clear that witnessing a waggle dance affected the subsequent behavior of foragers. If it had been discovered that the dances had no effect on bee behavior, I submit that cognitive scientists would resist the claim that waggle dances are cognitive mechanisms for honeybee hives (although the dances may well be behaviors of individual bees). Rather, the dances would merely be fascinating performances that served no function in the life of the hive. To take another example, Susan Goldin-Meadow argues that gesturing helps us (i.e. humans) to learn and to think (Goldin-Meadow and Wagner 2005). Her evidence is often cited by proponents of embodied cognition, who claim that it shows that physical gesturing is itself a part of a cognitive mechanism—not simply a behavior. This is because gesturing manages the successful performance of other behaviors, e.g. the performance of memory tasks or new skills. Thus, some external performances can themselves be parts of cognitive mechanisms rather than merely effects of them.

It is a familiar saying that "psychology is the scientific study of behavior," but the proper interpretation of *BEHAVIOR* is not a well-traveled line of Quinean retreat, and while there is an ample literature distinguishing intentional action from behavior, there is little discussion in the literature of what distinguishes behavior from other performances of organisms. So for example, self-motion (e.g. walking but not being pushed) is generally regarded as behavior, but pumping

blood and healing damaged tissue are generally not regarded as behavior. Bodily motions resulting from muscle spasms are not behaviors. One might suggest that behavior at least involves some endogenously-caused articulation of the body (e.g. movement of the limbs) so as to alter the organism's position or the organism's environment, but this guideline is just as difficult to make precise as *BEHAVIOR*. Furthermore, many behaviors do not obviously accomplish the manipulation of the organism's position or environment. Gesturing, for example, or dancing for the hell of it (i.e. not as a social signal) seem to be behaviors that are initiated only for their psychological effects on the subject. One might suggest that behavior be understood methodologically, as performances that are generally observable in an experimental setting, but cognitive scientists also observe non-behavioral phenomena (e.g. neural activity). I submit, then, that the proper explication of *BEHAVIOR* is an undertheorized topic, and that, if the sensitive management hypothesis is true, attempting to explicate *BEHAVIOR* has the potential to yield interesting results.

There is extant controversy about the extension of *BEHAVIOR*, though much of it is flat-footed and does not invoke articulate criteria for judging a performance to be a behavior. In the absence of much extant discussion, I will limit myself at present to characterizing interpretations of *BEHAVIOR* solely by reference to their extensions (I will defend a more general criterion in § 6.2.3). Some proponents of embodied cognition hold that adjusting one's gait (i.e. walk, trot, gallop) so as to conserve energy is an appropriate object of cognitive scientific explanation, and therefore appropriately behaviorlike (Kelso 1995, 70–74), though other cognitive scientists consider this at best a liminal case. Adams and Garrison suggest that cognitive behaviors are bodily movements that are explained by the reasons of the behaving critter (2013). I am sympathetic to this suggestion (see § 6.2.3), but it is not entirely clear what Adams and Garrison take reasons to be. Nevertheless, it is clear that they take this condition to restrict the extension of cognition to animal performances (and perhaps performances of possible robots, but not Herbert and not any other actual robots). So at least we can say that they assign a restrictive interpretation to *BEHAVIOR*, on which its extension includes many performances of animals, but not other organisms (Table 5.4 α). It therefore is an interpretation on which our paradigmatically cognitive cases are included

Table 5.5: Variation in the sectarian extension of COGNITION according to SMH. Interpretations are drawn from Table 5.2, Table 5.3, and Table 5.4. Elements of extensions are drawn from Table 5.1.

Variation in the sectarian extension of COGNITION according to SMH

Interpretations of parameters	Sectarian extension of COGNITION includes
ORGANISM α: corporeal elements of biological organisms *SENSITIVITY* α: classical representations with non-derived content *BEHAVIOR* α: some animal performances	A1. urban navigation in Inga A2. visual face recognition with FFA
ORGANISM β: corporeal elements of Darwinian individuals *SENSITIVITY* β: classical representations *BEHAVIOR* α: some animal performances	A1. urban navigation in Inga A2. visual face recognition with FFA C1. group cognition in ants
ORGANISM δ: mechanisms subserving systems apt for goal-ascription *SENSITIVITY* β: classical representations *BEHAVIOR* β: animal and robot performances	A1. urban navigation in Inga A2. visual face recognition with FFA C1. group cognition in ants C2. robot cognition in Herbert C3. extended cognition in Otto-notebook
ORGANISM γ: mechanisms subserving Darwinian individuals *SENSITIVITY* δ: classical representations and operation of adaptive mechanisms *BEHAVIOR* α: some animal performances	A1. urban navigation in Inga A2. visual face recognition with FFA C1. group cognition in ants C3. extended cognition in Otto-notebook C4. embodied audition in crickets
ORGANISM γ: mechanisms subserving Darwinian individuals *SENSITIVITY* γ: classical representations and pushmi-pullyu representations *BEHAVIOR* γ: animal and plant performances	A1. urban navigation in Inga A2. visual face recognition with FFA C5. off-line phototropism in *Lavatera cretica*
ORGANISM γ: mechanisms subserving Darwinian individuals *SENSITIVITY* γ: classical representations and pushmi-pullyu representations *BEHAVIOR* δ: animal and microbial performances	A1. urban navigation in Inga A2. visual face recognition with FFA C6. magnetotaxis in bacteria

in the sectarian extension of COGNITION. My recognition of Sanna's face in virtue of activity in FFA is a cognitive process, since human utterances belong to a commonsense extension of *BEHAVIOR* and my linguistic behavior—e.g. saying "Hi Sanna," is managed by the mechanism that allows me to recognize her face. Likewise, Inga's able navigation of her home city is a behavior on a commonsense interpretation of *BEHAVIOR*, so the processes that guide her navigation meet a necessary condition on cognitive processes.

In general, the most contentious issue concerning the extension of *BEHAVIOR* is whether it includes any performances of non-animals. Some scientists, like Brooks, are willing to countenance at least some performances of robots (C2; Table 5.4 β). Recently there has been a significant degree of interest in plant cognition, accompanied by ascriptions of *BEHAVIOR* to plant performances (C5; Table 5.4 γ). There is a journal, established in 2006, called *Plant Signaling and Behavior*. Calvo Garzón freely characterizes certain plant performances as behaviors (2007), but offers no criterion distinguishing behaviors from non-behavioral performances. He does offer a representation-in-absence criterion for distinguishing "cognitive" behavior from "reactive" behavior, but this seems to be a criterion for distinguishing cognitive from non-cognitive phenomena, rather than a criterion for distinguishing behaviors from non-behaviors. Evan Thompson and others (Lyon 2006; van Duijn, Keijzer, and Franken 2006; Thompson 2010) have suggested that all (or almost all) living organisms have cognitive processes, so they must be willing to consider at least some microbial performances to be behavior (C6; Table 5.4 δ).

If the magnetotaxis of some bacteria, or the phototropism of *Lavatera cretica* is considered behavior, then the mechanisms that manage those behaviors can be considered cognitive mechanisms according to SMH. If they are not considered behavior, then those performances do not justify the ascription of cognitive processes to bacteria or plants. Since these performances fall in the gap between conservative and liberal sectarian extensions, they are appropriate ruled controversial cases by SMH. An ecumenical explication aims to articulate seams of disagreement, and so much the better for the explication if some of those seams have not yet received the attention

they merit. If SMH is right, then philosophers interested in vindicating or denying the possibility of plant cognition should divert more attention to the question of what behavior is.

5.6 SUMMARY

In this chapter, I introduced the sensitive management hypothesis as an ecumenical, construal-based, parameterized explication of COGNITION. In discussing its parameters, I argued that it classifies each of the test cases correctly. It includes the paradigmatically cognitive phenomena in every sectarian extension, excludes clearly non-cognitive phenomena from every sectarian extension, and includes each of the controversially cognitive phenomena in some sectarian extensions but not others. Manipulating the *ORGANISM* parameter allows us to account for different judgments on group cognition (C1), robot cognition (C2), and extended cognition (C3). Manipulation of the *SENSITIVITY* parameter can account for some variation in judgment on extended cognition (C3) and embodied cognition (C4), as well as on plant cognition (C5) and microbe cognition (C6) if the *BEHAVIOR* parameter is also interpreted liberally (see Table 5.5).

While it is necessary to assign an interpretation to each parameter of the view in order for the explication above to produce a determinate sectarian extension, disputes over the better and worse interpretation of parameters can be independent. Thus, the sensitive management hypothesis affords not only the localization of disagreements in parameters and direction for conducting Quinean retreat, but also the pursuit of a divide and conquer strategy toward eventually clarifying and resolving scientific disputes about the nature of cognition.

In the next chapter, I will discuss a particular assignment of interpretations to the parameters of the sensitive management hypothesis. The conception I will describe is extremely permissive—the most permissive conception I think is available—but it offers some suggestive resources for explaining various features of cognitive scientific discourse, and in particular facts about the way cognitive scientists us the concept of REPRESENTATION.

6.0 THE COGNITIVE STANCE

6.1 OVERVIEW

In the previous chapter I described the sensitive management hypothesis (SMH), a parameterized explication of COGNITION that is ecumenically adequate, and a model of conceptual norms that articulates the patterns of disagreement found in contemporary cognitive science. In this chapter, I will describe and defend the most permissive conception of cognition I am willing to—i.e. the conception with the most permissive sectarian extension. I have three reasons for this. First, my remarks above about the most permissive interpretations of SMH's parameters were brief and bear expanding. Second, permissive conceptions of cognition are often criticized as conceptually confused, perhaps more often in conversation than in print (although Adams and Aizawa in particular are steadfast critics in print). However, as a construal-based explication the sensitive management hypothesis affords a novel kind of permissive conception for cognition. Moreover, the permissive conception I defend here in turn has novel resources for understanding various features of scientific practice. My discussion will also serve to illustrate the value of construal-based explications in general, I hope, by demonstrating how they can contribute in novel ways to anthropological questions in the philosophy of science. And third, many consequences of the permissive conception are also consequences of SMH on any assignment of interpretations to its parameters. So the permissive conception's potential for modeling practice and its metaphysical implications are, many of them, the potential and implications of the sensitive management hypothesis per se.

Therefore, I will revisit the sensitive management hypothesis, articulating what I take to be the most minimal defensible interpretations of its parameters (§ 6.2). Doing this will reveal the coherence of the permissive conception, which turns out not to be a clear criterion for demarcating the bounds of cognition so much as an expression of what it is to think of something as a bit of cognition. It is, if you will, an explication of what it is to take a "cognitive stance." While the permissive conception is certainly unusual, it is not without precedent. It is a precisification of a number of vague statements sometimes made by liberal partisans in the border wars, like Michael Anderson's, that

> conceptualizing cognition as the process of shaping inputs to cause desired behavioral outputs rather than as primarily a matter of transforming inputs to create certain kinds of representational content, will turn out to be one of the keys to developing a new theory of the mind and science of the brain. (Anderson 2014, 88)

Unlike SMH in the previous chapter, the permissive conception is not meant to model disagreements about the extension of COGNITION. However, it does model a fair degree of indifference between various approaches to modeling. I do not mean to defend the permissive conception here in all its details, or to defend the claim that it is the "correct" assignment of interpretations to the parameters of SMH. The crux of this chapter will be found in the last two substantive sections. In the first of these (§ 6.3) I will discuss my reasons for eschewing appeal to representations in SMH, and argue that the permissive conception nevertheless has the potential to illuminate various features of representation talk. In particular, I will discuss disputes over the existence of non-derived content, William Ramsey's job description challenge, and the ubiquity of representation talk in cognitive science. In the second (§ 6.4), I will reflect on the anthropological and conceptual implications of taking the permissive conception seriously. The main upshot of this discussion will be that the permissive conception encourages us to countenance the continuities between cognitive and non-cognitive phenomena.

6.2 A PERMISSIVE CONCEPTION OF COGNITION

Whereas the point of the previous chapter was to argue that the sensitive management hypothesis adequately models the variance in expert judgments, the point of this discussion is to articulate the minimal requirements for what it is to think of a system as a cognitive system. The permissive conception I describe is ineliminably construal-based, like Dennett's intentional stance (1987). To be a cognitive system, on this conception, is to be thought of in terms of the sensitive management hypothesis, on minimal interpretations of its parameters.

By way of reminder, the sensitive management hypothesis (Figure 6.1) holds that

SMH:

Cognition is the operation of mechanisms, where

1. the mechanisms *belong to an organism*,
2. the operation of the mechanisms is *sensitive to the organism's circumstances*, and
3. the mechanisms manage the *behavior* of the organism.

I will describe an extremely permissive conception of cognition by discussing very broad interpretations for each parameter. The approach I take here is like what Pamela Lyon calls a "biogenic approach" (2006), in which relatively simple biological systems are taken to be the core cases and fancy cognitive systems, like humans, are considered special cases. She contrasts this approach with the "anthropogenic approach," in which humans are taken as paradigmatic instances of cognitive organisms. The previous chapter was organized in an anthropogenic fashion; in this chapter I will linger more on non-human and non-animal cases.

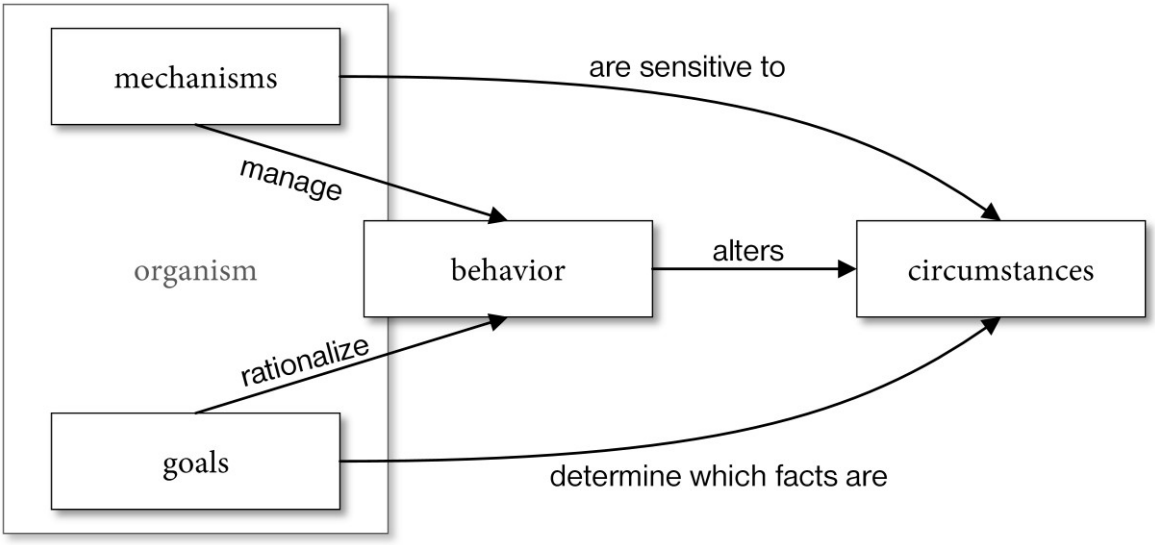

Figure 6.1: A permissive version of the sensitive management hypothesis.

6.2.1 Belonging to an organism

For the purposes of a maximally permissive conception of cognition, the notion of an organism can be fairly minimal. Thinking of a system as an organism involves at least construing it as a system that has, by our lights as inquirers, goals and mechanisms. Things go well for an organism insofar as it satisfies its goals (or is on track to do so), and poorly insofar as its goals are frustrated. Traditionally-ascribed goals for biological systems usually include survival and reproduction (or perhaps just the dissemination of genes), but on a permissive interpretation an organism might have any goals, including self-destruction. Thus the relation I propose between permissive organisms and their goals is less strong than the relation suggested by Nicholas Shea or Gualtiero Piccinini between organisms and their evolutionary functions, since I do not require organisms in this context to have Darwinian goals like survival and reproduction (Shea 2013; Piccinini 2015, Ch. 6). I leave it open what goals may be attributed to an organism because it is a matter for empirical or theoretical investigation. It is an open question, and anyway not a matter of much concern among cognitive scientists, whether these goals are subgoals of more basic biological goals

like survival and reproduction. Furthermore, I suspect that many goals of organisms will have tenuous or defeasible relations to Darwinian goals, given recent work on the development of "model-free" hypotheses about cognitive structure (e.g. Cushman 2013; "model-free" representations do not contain causal models of the world. The gist here is that an organism may treat instrumental goods as if they were intrinsic goods, because it does not or cannot keep track of the dependence relation between the goods).

Research programs in cognitive science are commonly structured around the implicit or explicit attribution of goals—the ends to which cognitive mechanisms are means. For example, the empirical research on word segmentation in linguistic perception (e.g. Saffran, Aslin, and Newport 1996; Johnson and Jusczyk 2001) and causal learning (Buchsbaum et al. 2012; Gopnik and Wellman 2012; Woodward 2012) are informed by a strong normative theory, so that the performance of actual cognitive systems can be compared against it.[28] It is also sometimes contentious how the goals should be described, as when Gerd Gigerenzer, Peter Todd, and colleagues (1999) argue that biases in human reasoning should be understood as effective heuristics for inference in conditions of limited information, time, and computational power, rather than as failures of rationality.

Organisms are furthermore structured so as to pursue their goals more or less effectively. I use the term "mechanism" to refer to the structures in virtue of which organisms are capable of pursuing their goals, in keeping with the common use of that term in the biological and behavioral sciences (Cummins 2000; Bechtel and Abrahamsen 2005) (see also Appendix § A.4). The ubiquity of the term "mechanism" in these contexts suggests that there is a scientific ontology of mechanisms, although causal, explanatory, and methodological claims about mechanisms are controversial (see Levy 2013). Actual (though perhaps not all conceivable) organisms have mechanisms that are physically realized.

Organisms can be identified with the physical system in which their mechanisms are realized, however its boundaries are demarcated, or they can be considered abstracta (like point

[28] Thanks to James Woodward for making this observation to me pointedly.

masses) that are associated with such systems, and are the possessors of goals and mechanisms. This second option is not unlike Rowlands', which requires that cognitive processes "belong" to a representational subject. On the permissive conception, an organism is anything that we construe as having goals and mechanisms. Thus, a hunk of iron rusting in a stream bed is an organism so long as I construe it as having goals, e.g. becoming reddish, and mechanisms that enable it to pursue those goals, e.g. a chemical constitution such that it becomes red in certain circumstances. However, a hunk of iron is a poor instance of an organism insofar as this construal is strained. And the hunk of iron does not count as a *cognitive* organism, not even on a maximally permissive conception, unless it meets the other conditions of SMH at least on maximally permissive interpretations of the other parameters.

The case of Brooks' robot Herbert is illustrative here (§ 5.3.2). Recall that Herbert is an autonomous robot that roams the robots laboratory at MIT (or rather, was—Herbert is retired now). Herbert is an autonomous mobile robot, capable of directing its own path (it mostly stays close to walls), avoiding obstacles, finding soda cans, carrying them home, and returning home to charge when its battery is low. Herbert's behaviors can be construed as goal-directed; it is enabled by its mechanisms (sensors, software, effectors) to engage in certain activities preferentially and to avoid certain circumstances in a flexible manner. The temptation to construe Herbert as possessing goals may be limited by the dissimilarity between Herbert's apparent goals and those of paradigmatic organisms: Herbert does not seek temperatures in a particular range, nor does Herbert possess capacities for self-defense or avoiding noxious stimuli, since Herbert was not designed for a hazardous environment. And Herbert certainly has no goal of reproducing. Nevertheless, I contend that if we succeed in construing Herbert as having *some* goals, it is tempting to that extent to construe Herbert as a potential subject of cognitive processes.

Adams and Garrison hold that cognition is the performance of behavior for reasons, and insist on that basis that Herbert's routine of collecting soda cans is not evidence of cognition. This is because "Herbert's movements… are not performed for Herbert's reasons. Herbert has no reasons" (Adams and Garrison 2013, 342). It becomes clear that Adams and Garrison hold this

opinion because they think Herbert is not the sort of entity that is capable of possessing goals, although it does possess mechanisms that can accomplish the goals of other genuine organisms, e.g. human designers:

> The programmer has constructed Herbert such that he will retrieve coke cans. If anyone's reasons are involved in what Herbert is doing, it is the reasons of the programmer, not of Herbert. At the level of the individual robot Herbert, there are no reasons… Herbert has no beliefs or desires nor intentions or plans of his own. (Adams and Garrison 2013, 347)

I confess that I do not fully understand what motivates Adams and Garrison to impose the particular interpretation they do of their own terms, and I will not attempt to exposit their view in detail here. I fully agree, however, that it is difficult to construe Herbert as a cognitive organism if we think of Herbert as a sophisticated recycling machine. My claim here is just that *insofar as we can be tempted* to construe Herbert as an autonomous system with an endogenous drive to collect soda cans, avoid colliding with obstacles, and so on, it is *just to that extent* that we can see Herbert as a system with its own cognitive properties (given that we can also construe Herbert as exhibiting sensitivity and behavior). The judgment that Herbert has cognitive processes of a sort varies along with our construal of Herbert as possessing goals. This covariation is all that is required of a construal-based explication.

6.2.2 Sensitivity to circumstances

Sensitivity to circumstances, on a maximally permissive interpretation, requires little more than the counterfactual (or subjunctive) dependence of states/operations of a mechanism upon circumstances. However, these dependences can only count as sensitivity if they meet constraints imposed by the interpretation assigned to *ORGANISM*. That is, by the goals and mechanisms

ascribed to the critter being considered. Circumstances, recall, are states of affairs that bear on an organism's goals (§ 5.5.2), so states of affairs that have no relation whatsoever to a goal of an organism are not circumstances for that organism, and circumstances are individuated on the basis of their consequences for the organism's ability to pursue its goals.

Sensitivity is also constrained by an understanding of the mechanisms possessed by an organism, insofar as these mechanisms constrain the ways in which a mechanism's operation can robustly depend on its circumstances. These mechanistic constraints are one of the sources of misrepresentation, if one wants to understand sensitivity in terms of representation (cf. § 5.5.2). As an illustration, consider the magnetotactic bacterium, which thrives in oxygen-poor water and perishes in oxygen-rich environments. Its circumstances are constrained by its goals in that its chances of survival and reproductive fitness are improved if it can avoid oxygen-rich water. Magnetotactic bacteria are sensitive to these circumstances. However, their sensitivity is constrained by the fact that they take magnetic field polarities as a proxy for oxygenation levels. In circumstances where local magnetic field properties do not serve as effective proxies for local oxygenation levels, the bacteria's mechanisms manage their behavior in ways that do not promote their goals. The state of affairs in which a bacterium is in an environment where this proxy relationship does not hold is a circumstance, since being in such a state of affairs has material consequences for the bacterium's well-being. However, the bacteria are not sensitive to this latter circumstance. One might say of such states of affairs that they lead to misrepresentation on the part of the bacteria. However, I do not wish to take any side here regarding entrenched debates about what if anything is the content of the bacterium's representations (e.g. between Millikan 1984; Dretske 1986). I am merely introducing, for the purposes of understanding cognitive systems, a minimal interpretation of *sensitivity* that does not purport to require representation in the philosopher's sense (but see § 6.3 below).

Much more can be said about the logical details of sensitivity, but in the interest of brevity I shall say just that sensitivities can be classified in a rough-and-ready way under two headings. On the one hand, perceptive and interoceptive sensitivities are those that concern states of affairs

including the organism's present circumstances—facts about the environment, the organism, or both. Thus, sensory inputs as well as pleasures, pains, and feelings like itchiness reflect sensitivities. If Jesse Prinz (2004) is right that emotions are perception-like states representing organism-environment relations, then emotions are sensitivities. Non-perceptive sensitivities include all other ways that the operations of cognitive mechanisms can be regulated according to circumstances. Mnemonic sensitivities contribute to the adaptive operation of mechanisms by regulating their operation according to past states. One might use the term "inferential sensitivity" to describe any sensitivity to future circumstances in light of present or past states, as well as any sensitivities that are mediated by relatively impressive mechanistic architecture. These classifications are not meant to be exclusive or exhaustive; I mention them merely to emphasize that a permissive conception of *SENSITIVITY* includes many ways that the operations of mechanisms can depend on the circumstances of an organism.

6.2.3 Behavior

I said rather little about behavior above (§ 5.5.3) because relatively little is said in the literature of the border wars about how *BEHAVIOR* should be interpreted. Adams and Garrison suggest that behavior is the causing of bodily movement (2013, 341; though surely behavior is the movement and cognition is the cause?), and cognition is the causing of bodily movement for reasons. The permissive conception I recommend here is similar in some respects, but confessedly unusual, and I will say a little about it here.

I suggest that on a permissive interpretation of *BEHAVIOR*, behavior be understood as the autonomous movement of the body or manipulation of the environment, and that behavior must be rationalizable by a goal of the organism. By "autonomous" I just mean that the behavior is caused by the operation of cognitive mechanisms. Of course, it is a necessary condition on cognitive mechanisms that they manage behavior, but since both MECHANISM and *BEHAVIOR* have other necessary conditions this mutual requirement is not circular. It is just that, on the permissive

interpretation of behavior I advocate, there are no cognitive mechanisms without behavior and no behavior without cognitive mechanisms. Movement of the body is the reconfiguration of body posture or movement through space. It does not include mere growth, but for critters like plants who grow in particular manners—more this direction than that, in this shape rather than another—such performances amount to reconfigurations of posture on the permissive interpretation I advocate because they position the plant to exploit the resources in its environment in new ways. I also include manipulation of the external environment by means other than bodily movement. Lima beans (*Phaseolus lunaatus*) combat predators (like caterpillars that damage their leaves) by releasing volatile chemicals and producing different nectar when under attack, so as to attract predators of their predators (Heil and Silva Bueno 2007). If telekinesis were possible, it would be a candidate for behavior even if it were accomplished without movement of the body. All of these behaviors are externally observable; immune responses like the production of antibodies are not bodily movements or manipulations of the environment external to the organism, are not behaviors.

6.2.3.1 Rationalization to goals Rationalization to goals can be understood roughly along the familiar lines of traditional belief-desire causal action theory (e.g. Davidson 1963), on which rationalization of an action ϕ is achieved by means of a primary reason. Primary reasons are ordered pairs consisting of a belief (that one can ψ by ϕ-ing) and a desire or pro-attitude (that it would be somehow good to ψ), and they rationalize an action just in case they bear the correct logical relationships and cause the action "in the right way." In the cognitive case, however, having a goal stands in for a desire or pro-attitude, and sensitivity to circumstances stands in for belief. In the context of describing intentional action, it is difficult to say what it is to cause an action "in the right way." In cognitive science, however, the central explanatory project is precisely to spell out how goals and sensitivity produce behavior, as mediated by cognitive mechanisms. Thus this rationalization requirement is not an ambitious naturalization of behavior—it is an analogue of practical inference, not an explanation or grounding of it. Furthermore, a behavior may be

rationalized in this sense even if it is not, all things considered, an ideally rational behavior. Attention to mechanistic constraints on rationalization is often attention to the way the processes fall short of perfect rationality.

Following Davidson, then, the rationalization requirement might be spelled out like this:

RATIONALIZATION:

A pattern of behavior *B* of an organism *o* is rationalized by a goal *g* of *o* if and only if

1. *B* is a means to *g*, given a pattern of circumstances *C*,
2. *o* is sensitive to *C*,
3. *o* possesses a mechanism that manages *B* according to its sensitivity to *C*.

This scheme refers to patterns of circumstances and behavior, rather than to individual circumstances and behaviors, because rationalization is to be understood contrastively. Some behavior is performed rather than another, and some circumstance obtains rather than another. So policies of behavior are rationalized relative to possible circumstances. However, there can be many means to the same end. RATIONALIZATION does not require a behavior to be an optimal means to the end given constraints by circumstances, sensitivities, or mechanistic constitution.

This schema is similar to the traditional Davidsonian model of intentional action in that we can reconstruct reasons (or their less discursive analogues) for behaviors, where reasons are characterized in terms of motivational and doxastic factors. However, unlike traditional views of intentional action (and unlike Adams and Garrison's view) RATIONALIZATION does not require the organism to possess beliefs and desires in the folk psychological sense, nor is it necessary that the organism be capable of representation in the rich way that human language permits. It is only necessary that the organism have goals and be sensitive to its circumstances regarding those goals.

Furthermore, the account stipulates that the motivational and doxastic factors be related to behavior not by global constraints of rationality, but by constraints implementable in mechanisms that the organism possesses. Thus, I distinguish here between *rationalization* and *rationality*. Rationality, as understood in a normative theory like rational choice theory, is a normative standard for action or inference achievable with unbounded computational resources and, often, contexts where the evidence base of the reasoner is not circumscribed. (Thus, for example, one might argue that, having constructed an elaborate model and crunched the numbers, a common decision strategy is, surprisingly, irrational after all.) The cognitive mechanisms of actual organisms do not have unbounded computational resources and often operate under unfavorable circumstances regarding the availability of evidence or information. Information is limited by an organism's sensitivities. Memory and other computational resources vary between organisms according to their mechanistic constitutions. Plants and microbes exhibit comparatively limited capacities for learning relative to animals, and for them major variations in behavioral policies occur between individuals or over generations (and regarding some matters, this is true for humans as well). So while many behavioral and biological systems may approximate rationality in some ways,[29] the tools of rational choice theory have limited application in rationalizing the behavioral policies of cognitive organisms. And this is to say nothing of Aristotelian conceptions of rationality as responsiveness to reasons as such (cf. Sellars 1981; Broome 2007), which are clearly too strong a requirement on a permissive interpretation of BEHAVIOR.

Rationalization is a much weaker requirement than ideal rationality. The trickiest requirement is that, in order to count as behavior, it must be possible to construe a performance as a means to a goal of the performing organism. Attractive candidates for cognitive behaviors can be rationalized this way. For example, the magnetotactic bacterium's motion is easily construed as a means to its survival, given that it must avoid oxygen-rich water. The phototropism of *Lavatera*

[29] The role of optimality assumptions in the biological sciences is controversial (Parker and Maynard Smith 1990; Green, Levy, and Bechtel 2014), and I take no position on it here. The assumption of optimality may still be a good methodological presumption, and not be a requirement on behavior per se.

cretica is a means for it to exploit natural resources more efficiently. The production of aromatic pollen is a means for lima beans to protect themselves from predators. The dances of honeybees are a means for the hive to exploit food resources more efficiently. Where gesturing helps the gesturer to learn, even without her knowledge, it might be construed as a behavior that helps her achieve her goal of mastering a task. When someone dances out of joy, it is expressive and fun. An involuntary muscle spasm, by contrast, is relatively difficult to construe as a means to some end. All of these behaviors can fail to achieve their ends—experimenters may move food sources so that honeybees search fruitlessly. Dancing may turn out not to be much fun. Nevertheless, the failure to achieve a goal does not defeat a judgment that a behavior was a means to achieving the goal.

There are a number of constraints built into RATIONALIZATION. Of course, construals of means-end relationships are constrained by which goals we attribute to organisms. Explaining why we attribute some goals and not others is no simple task, but we do it and those attributions are in general less contentious than judgments about whether a process is cognitive. Rationalization is also constrained in that cognitive behaviors must be managed according to an organism's sensitivity to its circumstances—different circumstances call for different means. But organisms can only manage their behavior according to their sensitivities to their circumstances, not according to their actual circumstances. Thus, the standard of judgment for whether a performance rationalizable is the informational perspective of the organism, not the perspective of the inquirer. An organism may be insensitive to circumstances to which we are sensitive. The magnetotactic bacterium, for example, is insensitive to the fact that magnetic field polarities may become decorrelated from oxygenation gradients, though in some conditions scientists are sensitive to that circumstance. We can still rationalize the behavior of the bacteria because their movement is constrained by the limitations of its sensitivity. Rationalization is also constrained by the mechanisms an organism possesses for managing its performances based on its sensitivities. Consider the counterfactual case of fruitless honeybee dances, in which dances reflect information about the location of food but that information is not used to manage the behavior of the hive (discussed above in § 5.5.3). In this imaginary case, the hive structure possesses no mechanism for

capitalizing on this information, although its component bees are sensitive to it. Now suppose a honeybee that witnessed a dance does proceed to the food source indicated. This is a manifestation of chance, not a behavior of the hive, because it was not managed by a mechanism.

I have described a permissive conception of cognition, on which each parameter remains construal-sensitive even after being assigned an interpretation. On this conception of cognition, construing something as an organism requires seeing it as a system with its own goals and mechanisms that enable it to pursue those goals. Construing its mechanisms as sensitive to circumstances requires seeing their operations as influenced by the organism's circumstances, and seeing a state of affairs as a circumstance requires construing it in terms of its potential to help or hinder the organism in its pursuit of its goals. Finally, construing a performance as a behavior requires seeing it as a means for the organism to pursue its goals. The overall picture is that thinking of a system as a cognitive system is just to think of it as a system with concerns, and mechanisms that shape its behavior in response to its information about its environment.

6.3 COGNITION AND REPRESENTATION

One of the most unusual features of SMH, on this assignment of permissive interpretations to its parameters but also in its bare state without interpretations, is that it does not explicitly invoke the notion of REPRESENTATION or even INFORMATION, at least in a robust or precise way. Each of the extant views on the "mark of the cognitive" invokes representations (§ 1.2.2.2), and many more informal statements about the nature of cognition in other literatures tend to do the same (e.g. Burge 2014). Furthermore, it is often alleged that embracing the concept of representation is the founding insight and sine qua non of cognitive science (§ 1.2 ;Ramsey 2007; Buckner 2015). So it might appear that either I have gotten COGNITION dreadfully wrong, or else I have gotten it right and cognitive science has turned its back on its founding insight.

This worry is misplaced. First of all, it is not clear that positing representations was ever essential to cognitive science. Another narrative, also found in textbooks and the whiggish histories of psychology, is that the cognitive revolution was driven by renewed interest not in representations per se but in *intermediate states*; i.e. complex states that processes that mediate stimuli and responses. This is precisely what SMH requires, even on the permissive interpretations I described above. The *SENSITIVITY* parameter is broad enough that, on any of the available interpretations I mentioned in Section 5.5.2 above, it is not restricted to the registration of sensory stimuli. There is no danger here of relapsing into the much-dreaded behaviorism of Watson and Skinner.

In fact, contemporary textbooks (e.g. Anderson 2010) are more likely to cite information-processing than representations as crucial to the cognitive revolution, and only Rowlands' "mark" deigns to mention information-processing explicitly. I decline to characterize cognition in terms of information, as well, because the notion is vague and open to interpretation and controversy regarding its role in cognitive science (§ 1.2.2.1). Thus, an ecumenical explication in terms of information faces objections similar to explications in terms of representation.

The most compelling reason to eschew representation talk in an ecumenical explication of cognition is that the nature of natural representation remains deeply controversial. Furthermore, it is nowadays a matter of open controversy whether representational posits are necessary or advisable features of cognitive models (van Gelder 1995; Chemero 2009). My own inclination is not toward anti-representationalism, but ecumenical goals demand that I not rule out such views a priori. The SENSITIVITY parameter in SMH preserves a place in my model under which all this controversy may fall. Disagreement among cognitive scientists about the nature of representation is so ubiquitous and deep that invoking REPRESENTATION to explicate COGNITION is illuminating the obscure by means of the equally obscure. The parameterized predicates in SMH are open to controversy, and some of them are admittedly neologistic, but I contend that they are less embattled and prone to confusion than REPRESENTATION. In fact, I am prepared to argue that if the sensitive management hypothesis is true, it offers resources for clarifying the concept of representation as it figures in cognitive science (§§ 6.3.2–6.3.3).

In the remainder of this section, I will discuss a number of more specific topics germane to representation: the existence of non-derived content, William Ramsey's job description challenge, and the point of talking in terms of "representations" at all. My remarks will be preliminary; it is beyond the scope of the present chapter to carefully exposit and evaluate the extant arguments on any of these topics. My present aim is merely to show that the permissive conception of cognition has promise as a resource for novel treatments of each of them.

6.3.1 Non-derived content

It is a common assumption in the literature on cognition (though also commonly contested) that cognition involves representations that bear something called, variously, original intentionality, intrinsic content, or *non-derived content* (Searle 1980; Haugeland 1981b; Fodor 1987; Dennett 1990; Adams and Aizawa 2001, 2008, 2010; Aizawa and Adams 2005; Clark 2005, 2008; Rowlands 2009; 2010. I will adopt the latter term here). Characterizations of non-derived content are usually ostensive. It is the sort of content that is possessed by mental states—e.g. beliefs, desires, possibly emotions. It is contrasted with derived intentionality, or derived content, which is the sort of content that is possessed by maps, writing, and sounds. Representations with derived content inherit their content from representations with non-derived content, usually with the help of some sort of conventional system (e.g. language, diagramming conventions) that provides a mapping to intrinsically meaningful contents. Or so the story goes.

Possession of non-derived content is a necessary condition for cognitive representations, according to Adams, Aizawa, and Garrison. Thus, according to them, cases like Herbert and Otto with his notebook cannot be cases of cognition. Herbert is a mere machine, incapable of possessing intrinsically meaningful representations. The representations realized in Otto's brain are representations of the right kind, but the representations in his notebook bear their content only derivatively. They have content only because of a conventional system that assigns them content

based on the psychological states of human beings that read English orthography. So the representations in the notebook cannot be parts of a cognitive state.

I am skeptical about this reasoning. Mostly, I am skeptical because I am unconvinced that there is sufficient reason to think that brain states are somehow capable of original intentionality whereas ink-marks and sound are not. Surely if anything has non-derived content it is cognitive or mental states per se, and not the brain states that realize them. One might endorse a form of physicalism on which properties of a functional system are also properties of their physical realizers, so that brain states can inherit non-derived content from the mental states they realize (notwithstanding that it is rather odd to say that an item can inherit non-derived properties), but it is not clear that such a view would rule out Otto's notebook while ruling in brains. However, I will not argue for my view here.[30] Rather, I will propose an alternative way of understanding what is involved in the attribution of non-derived contents. A story like the one above—on which derivative content is derived from intrinsically meaningful representations—is not the only game in town. Here is another conjecture, based on the resources of the sensitive management hypothesis: we are tempted to think of contents as non-derived just when we are tempted to think of them as contents borne by cognitive representations of organisms, in the sense of *ORGANISM* peculiar to the sensitive management hypothesis. There is at least anecdotal data consistent with this suggestion. For example, Fred Adams' unwillingness to attribute non-derived content to Herbert's representations correlates with his unwillingness to ascribe genuine goals to Herbert.

The case of Otto is more complicated, since Otto is the subject of his extended processes and Otto is by universal agreement a cognitive organism. Nevertheless, Adams and Aizawa evince a temptation to attribute to extended cognition's defenders the claim that Otto's notebook itself—rather than Otto or an Otto-notebook system—has cognitive states. For instance, they open a critical essay on the subject with the following joke and commentary:

[30] Clark argues—persuasively, in my judgment—for a view like this one in many places. (His most pointed arguments are found throughout his 2008. Perhaps the punchiest argument is the one that accompanies the hippo-world thought experiment, pp. 109–110.)

> Question: Why did the pencil think that 2 + 2 = 4?
>
> Clark's answer: Because it was coupled to the mathematician.
>
> That about sums up what's wrong with Clark's extended mind hypothesis. (2010, 67)

Of course, Clark's claim is not that a prop like the mathematician's pencil or Otto's notebook thinks anything. His claim is that in some circumstances we can attribute cognitive processes and representational contents *to organisms* like Otto in virtue of interactions between the organism and its props, rather than solely in virtue of processes in the organism's brain. Adams and Aizawa disavow this misinterpretation. And indeed, outside of the joke they take care to characterize the extended cognition hypothesis correctly. However, Clark argues that careful attention to their arguments and examples reveals that they are committed to the misinterpretation in crucial but non-obvious ways, concluding that

> Adams and Aizawa are committed to the usefulness of pressing a question that, to us, looks pretty clearly to be among the very reddest of possible herrings. That is the question whether Otto's notebook (to put the matter bluntly) is "cognitive." (Clark 2010a, 90).

It is beyond the scope of this discussion to evaluate Clark's arguments carefully. My present purpose is merely to motivate the suggestion that ascriptions of non-derived content can be accounted for as a construal-sensitive phenomenon, consistent with the permissive conception of cognition as the sensitive management of organismal behavior.

6.3.2 The job description challenge

Another common view about the concept of REPRESENTATION as it is used in the cognitive sciences is that it is quite commonly over-applied. For example, Robert Cummins contrasts

representation (in a relatively robust philosophical sense) with "indication" (which is a relation that bears something like Grice's or Dretske's "natural meaning"; Cummins and Poirier 2004), but many ascriptions of REPRESENTATION in cognitive science, especially in the controversial cases I have been considering, fall more naturally under the heading of indication than representation. William Ramsey provides a book-length articulation of this worry:

> New accounts share with conventional computational theories the basic idea that inner structures in some way serve to stand for, designate, or mean something else. The commitment to inner representations has remained, despite the rejection of the symbol-based habitat in which the notion of representation originally flourished.
>
> …[Nowadays] we have accounts that are characterized as "representational," but where the structures and states called representations are actually doing something else. This has led to some important misconceptions about the status of representationalism, the nature of cognitive science and the direction in which it is headed. (2007, 3)

Ramsey describes what he calls the "job description challenge": to give an account of the unique properties of representations in virtue of which appealing to them serves a special explanatory role, and to give a model or models of how that role is fulfilled by some natural entities that have those properties. Ramsey invokes a conception of representation that has significant affinities with the philosophical notions that animated the classical computational theory of cognition (q.v. § 1.2.2), and concludes that while some uses of REPRESENTATION in cognitive science satisfy this conception, many do not. In particular, Ramsey argues that there is a common use of the term "representation" among cognitive scientists on which it means something more akin to Cummins' "indication," and that this is a deplorable misuse of language. Again, the careful description and evaluation of Ramsey's arguments are beyond the scope of the present discussion, but I will sketch a counter-narrative to Ramsey based on the permissive conception of SMH.

Ramsey works through examples of several conceptions of representation. I will focus on what he calls the "receptor notion." According to Ramsey, it is sufficient for being a representation on this conception that a neural or other structure is regularly and reliably activated by a distal condition (119). Thus, for example, according to the receptor notion a magnetosome in a bacteria represents the local oxygenation gradient (or the direction of a local magnetic field, or whatever) because its states are regularly and reliably activated by corresponding states of the local oxygenation gradient (or magnetic field), and no other conditions need obtain. Alternately, the triggering mechanism that causes the jaws of Venus flytrap to close in the presence of insects represents the presence of insects because it is regularly and reliably activated by the presence of insects. However, Ramsey observes, this is an absurdly weak notion of REPRESENTATION. By the same criteria, smoke represents the presence of fire, and the firing pin in a gun represents the squeezing of the trigger, and the rusting of iron represents the presence of water and oxygen. Furthermore, such a notion of representation is well-known to face the disjunction problem (Fodor 1987): why should a bacterium's magnetosome represent oxygenation rather than the local magnetic field, if both cause the same state in the magnetosome? Why should Venus flytraps represent insects rather than contact with a sufficiently large object? Why does frog visual cortex represent flies rather than wiggly, dark spots? Why does our concept GOAT represent goats rather than goats and also unusual-looking sheep?

Ramsey's criticisms of the receptor notion as such are quite decisive. However, I question whether his characterization of the receptor notion is a charitable understanding of what happens in cognitive science. Cognitive scientists do not, after all, claim that the concept GOAT sometimes represents sheep (at least for competent judges of goathood), or that rusting iron represents oxygen (there is room for disagreement about the representanda of sub-organismal mechansims like magnetosomes, frog visual cortex, and Venus flytrap triggers). I suspect that in cases where cognitive scientists appear to invoke a receptor notion, they in fact invoke a richer notion of REPRESENTATION that is not subject to the same worries. Namely, they invoke the permissive notion of *SENSITIVITY* outlined above (§ 6.2.2).

Ramsey's sufficient condition for receptor-representation, that a state be tokened regularly and reliably by a distal condition, is sufficient for the permissive interpretation of *SENSITIVITY* only together with further conditions on judgment. In particular, a sensitive state must also be construed as a state of a mechanism that belongs to an entity we construe as an organism, and the distal condition must be a circumstance for that organism. Now, I acknowledge that this more demanding condition is not illuminating in the case of human linguistic concepts like GOAT, but I contend that it provides a more satisfying understanding of the other cases.

First, take the biological cases of the frog, the flytrap, and the bacterium. As biological organisms, we construe these systems as possessed of Darwinian goals and mechanisms for pursuing those goals. The goals also structure our understanding of states of affairs as circumstances for these organisms. The attribution of representations to these systems occurs in the context of these construals, and the ascription of content to these representations tends to invoke the organisms' circumstances as we understand them in the context of such construals. For example, we are inclined to judge that flytrap triggering mechanisms represent insects because the presence of insects, rather than bits of putty or plastic, is a circumstance on which the flytrap can act in order to promote its goals. Given the limitations of the flytrap's mechanistic constitution, however, it cannot distinguish the presence of insects from the presence of various other objects. The use of the same language to describe both putative contents—the insects and various indigestible objects—may well be a bit of sloppiness worth reforming. However, I think the conceptual practices of the scientists are not as crude as the practice articulated in Ramsey's "receptor notion."

Regarding the non-biological cases of smoke, rust, and the gun's firing pin, the permissive interpretation of *SENSITIVITY* does rule them out except in the context of unusual construals. We are not inclined to construe smoke as an organism, possessed of goals and mechanisms for pursuing those goals, where the presence of fire is a circumstance that bears on its pursuit of its goals. The same goes, mutatis mutandis, for hunks of iron and guns. Nevertheless, I confess that it is possible to construe at least some of these items as cognitive organisms with sufficient effort. For example,

people sometimes construe weapons as having individual personalities, with goals of their own apart from the goals of their wielders. In feudal Japan, for example, some swords were said to be particularly bloodthirsty, and to have an effect on the dispositions of their bearers (a relatively familiar example is found in the weapon of Sasaki Kojirō in Yoshikawa Eiji's novel *Musashi*, 1971). Were we to construe a gun in this way, we might find a way to think of it such that its firing pin exhibited the right sort of sensitivity to the gun's circumstances.[31] As I understand the Japanese folk conception, though, swords supposedly exhibited their influence through some supernatural influence rather than through their mechanistic constitution. They are thus not proper objects of cognitive science, which is concerned with the explanation of capacities by appeal to natural mechanisms. However, if (contrary to fact) weapons really did influence their bearers, and if their influence were explainable through natural mechanisms rather than supernatural influence, I suspect that there might be some parties in the border wars arguing that they are cognitive systems. I am not sure how to evaluate a counterfactual statement like this, but if it is true it is consistent with the permissive conception of cognition I defend in this chapter, and I hereby bite the relevant bullet (no pun intended).

At any rate, I contend that the permissive, construal-based conception of cognition described in this chapter is more adequate than Ramsey's receptor notion as a model of the pattern of judgments exhibited by cognitive scientists. Perhaps Ramsey is still right that it is disappointing to use the word "representation" to refer to such diverse phenomena as natural language, conventional signs and maps, and instances of organismal sensitivity as exhibited in a Venus flytrap. It is certainly true that Ramsey's favored conception of representation (i.e. "S-representation")

[31] Though I am hard-pressed to work out a satisfying such construal, unless the gun's goal is to obediently function as designed. I suspect, though, that as a matter of fact we are more likely to attribute goals to most artifacts when they resist their designed functions. Buggy computers, e.g., are said to have minds of their own or work against their owners. Even brains are said to have minds of their own when they malfunction in ways inconvenient or debilitating for the organism that possesses them. I confess, though, that I am not sure what to make of the fact that these construals are easier or more difficult to make in some circumstances. I am mindful of the fact that I have not provided a formal semantics for construals. However, the articulation of a semantics for modal language was one of the greatest philosophical achievements of the last hundred years—this is not the context to offer a comparable achievement for construals.

differs in some dramatic ways from any conception appropriate for flytraps. However, the ubiquity of representation talk is so well-established in scientific practice that I suspect it will not change on the basis of objections like Ramsey's. More importantly, though, I contend that the ascription of content is based on a richer conception of representation than Ramsey's receptor notion. This contention leaves open the question: What use are representation-ascriptions to cognitive science?

6.3.3 The use of representation talk

A more adequate treatment of this question would situate the practice of using representation talk in the context of a story about the uses of any kind of scientific talk. Yet again, a detailed treatment of the topic is beyond the scope of this discussion. However, after a brief digression, I will offer a schematic suggestion based on a standard view in the philosophy of action.

There is a well-known class of objections to computational functionalism, called triviality objections, alleging that any sufficiently complex physical system can be assigned any functional structure (Putnam 1988; Searle 1990; Godfrey-Smith 2009b). These objections allege that, according to the computational functionalist, so long as there is some isomorphism between the state-transitions of a functional system and the state-transitions of a physical system, the physical system implements the functional system. But if there are no constraints on how states of physical systems are to be individuated or how state-transitions occur, then any physical system of sufficient complexity—say, a bucket of water in the sun, or the wall in John Searle's office—implements every functional system—an addition machine, a word-processing program, the Google search algorithm, a human mind, &c. I am not concerned at present with computational functionalism, but there is a similar worry about cognitive science: on what basis do we categorize parts and activities of brains as parts and activities of cognitive mechanisms? And why one mechanistic structure rather than another?

Here are two caricatures of how we go about studying mechanisms. In one caricature, we examine the causal structure of the world, cataloguing relationships and forms of causal

organization. We then infer from various special cases of causal structure that some forms of organization produce systems with distinctive capacities. We can say that these structures explain those capacities. When we discover actual instances of those structures, we call them mechanisms and we call the capacities they explain phenomena. This is a caricature that makes the triviality objections seem compelling. Why should we select one way of construing the causal organization of a system rather than another? Why construe John Searle's brain, for example, as an implementation of a human mind and not an implementation of a word processor? On the second caricature, we move in the opposite direction. We begin by selecting a phenomenon, relatively stable and general feature of the world that we wish to explain. We then examine the causal structure of the world, taking note of such features as contribute to an explanation of the phenomenon. In some cases, we discover forms of organization (at some level of abstraction and generality) that explain the phenomenon. In such cases we call the instances of that structure a mechanism.

Both of these caricatures are simplistic, and in practice we do a little of both, adjusting in turn our understanding of the phenomenon, of the causal organization of the world, of the boundaries of the mechanistic system, and of the generality and abstraction of the categories with which we characterize all of these. But the second caricature is somewhat more recognizable: we often begin with a phenomenon in search of an explanation. For instance, Carl Craver's main exemplar of mechanistic explanation, the Hodgkin-Huxley model of the action potential, involves a known phenomenon in need of a mechanistic explanation (Craver 2007). It is plausible that scientific inquiry is structured by prior conceptions of explanandum phenomena, however subject to revision, and that the construal of a causal structure as a system, or as having mechanistic organization, is constrained by our conception of the explanandum.

Now consider how this generalization applies to cognitive science. Cognition is, so I argue, the sensitive management of organismal behavior. So the explanation of cognitive capacities per se requires the specification of how a system is organized so as to exhibit capacities to manage its behavior in a manner sensitive to its circumstances. The construal of a system as an organism,

possessed of goals and circumstances, is a presupposition of the project of explaining its cognitive capacities. This picture does not clarify why we choose to construe some systems as cognitive rather than others, e.g. why we take the cognitive stance toward animals, and perhaps to plants, but not to guns and swords. However, the picture does suggest a use of representation talk: if the explananda of cognitive science are capacities for sensitively managing organismal behavior, then representation talk provides scientists a vocabulary for expressing the causal influences of circumstances on the operations of mechanisms. In short, assumptions about quasi-intensional relations are part of the presuppositions of cognitive scientific inquiry, and representation talk is an easy vocabulary for describing non-extensional relations.

Consider Davidson's discussion of intentional action: it is under the guise of a particular description that one performs an intentional action. Davidson turns on the light switch in order to turn on the light, but not in order to alert the prowler outside his house (whose presence is unknown to Davidson) that he is home, though he also does the latter. This feature of intentional actions is sometimes called description-relativity; Davidson calls it "quasi-intensional character" (Davidson 1963, 5). The quasi-intensional specification of an intentional action is derived from the intensional specification of a primary reason (q.v. § 6.2.3.1). Many features of cognition also have a quasi-intensional character; cognitive behavior, sensitivity, and the attribution of functions to parts of mechanisms all have privileged characterizations, and thus a quasi-intensional character. The privileged characterizations are those that participate in a complex of such descriptions that are assertable in the context of construing a system as a cognitive system, i.e. in terms of the sensitive management hypothesis. For example, the magnetotactic bacterium moves this way in order to avoid oxygenated water, not in order to follow a magnetic field, though it also does the latter. Activity in FFA is for recognizing faces, not consuming glucose, though it also does the latter. The ink in Otto's notebook is for preserving a legible record of Otto's commitments, not absorbing more light than the surrounding paper, though it also does the latter.

The traditional conception of representation is of a contentful state, where the content can be expressed as an intensional description. Representational language is not necessary for

characterizing the quasi-intensional character of cognitive mechanisms, but it was always available to cognitive scientists and they use it inter alia to express privileged descriptions of cognitive systems consistent with construals in terms of the permissive conception of cognition I describe here.

6.4 COUNTENANCING CONTINUITY

The permissive conception of cognition suggests a dramatic continuity between cognitive and non-cognitive phenomena. This implication bears some discussion—discussion that will largely apply to the sensitive management hypothesis per se as well its permissive version. I will begin by discussion the sense in which the sensitive management hypothesis is ambitious or unambitious relative to extant "marks of the cognitive" (§ 6.4.1). I will also discuss the apparent liberality of the permissive conception (§ 6.4.2), and what can be said about the place of computational inquiry in cognitive science (§ 6.4.3). I will end this section, and the chapter, with some remarks on the continuity between the sensitive management hypothesis and historical conceptions of the mind (§ 6.4.4).

6.4.1 Ecumenical explication and ambition

I set out in this chapter to provide an account of COGNITION as a construal-sensitive concept. This aim is rather different than the aims that animate other "marks of the cognitive." Consequently, the permissive conception of cognition is both more and less ambitious than other "marks." It is relatively ambitious in that it purports to provide an adequate criterion by which one can reckon whether something is a bit of cognition or not. The four extant "marks" do not purport to be so complete. Adams, Aizawa, and Garrison offer only necessary conditions on cognition; Rowlands offers only sufficient conditions. Buckner offers neither necessary nor sufficient conditions.

However, in another way the sensitive management hypothesis is far less ambitious than the extant "marks." The kind of adequacy aimed for by SMH is ecumenical extensional adequacy—adequacy to the judgments of contemporary cognitive scientists. The adequacy aimed for by each of the extant marks is *speculative*, adequacy to the true nature of cognition. They are, in effect, proto-theories of cognition, and they are confessedly incomplete because cognitive science has yet to discover the true nature of its object. Or, to put the contrast another way, I have aimed to give an account of COGNITION, the scientific concept, whereas other "marks of the cognitive" have aimed to be accounts of cognition, the thing itself.

It might be objected that I have given up too quickly on the possibility of articulating a speculative "mark of the cognitive." Nothing I have said rules out the possibility that we will someday discover a unified and univocal explanation of cognitive phenomena in terms of an underlying causal or functional process, and if that happens then it will have been shown that there is a "genuine" speculative criterion that is more robust and precise than the sensitive management hypothesis. Perhaps it will even be a set of individually necessary and jointly sufficient conditions for exemplifying the natural kind cognition. And when we are in possession of such a criterion, we will be able to adjudicate the disputes of the border wars, demarcating the genuinely cognitive phenomena from those that merely resemble them in non-essential respects. A related worry is that the sensitive management hypothesis is bound eventually for obsolescence. The judgments of cognitive scientists about which phenomena are cognitive will almost certainly continue to change, and in all likelihood will come to converge on greater agreement than they exhibit now. Thus, my Pyrrhonian suspension of judgment about particular research traditions will turn out to be unjustified.

My reply is modest. The permissive conception, as a bit of logical anthropology, is only supposed to serve as a model of the conceptual practices of the present. I am wary of placing bets on how a scientific concept will develop in light of evidence we do not yet have or decisions we have not yet made (§§ 1.4.2, 2.4,). Through the use of parameterization I have tried to articulate a criterion that will survive some diachronic variation in the norms for ascribing COGNITION

without trying to predict precisely how those norms will change (§ 4.3.1). Nevertheless, I do hold that the anthropological method holds significant promise for yielding knowledge—as good, anyway, as there is to be had—of the nature of cognition, insofar as there is a nature there to be known (§ 1.3.1).

6.4.2 "Liberals are ruining cognitive science"

It would be natural to press the worry that, even considering my ecumenical aims, the permissive conception of cognition is excessively liberal. For example, although I cannot manage to construe a gun as a cognitive system, an automated thermostat can be construed so as to satisfy SMH. It has the goal of keeping the ambient temperature in a particular range, and mechanisms for mediating between its sensitivity to that goal and behavior, i.e. activating and deactivating climate control hardware. Similarly, the view that cells of multicellular organisms have cognitive processes is an unpopular one, though such cells could be construed so as to satisfy SMH. So long as these construals are possible, then the permissive interpretation of SMH is committed to the consequence that there is no deep, important, conceptual difference between the systems cognitive scientists in fact study and thermostats.

My reply is two-pronged; the first prong is concessive, but the second prong rejects the worry, with some qualifications. The concessive prong is that there is certainly some justice in this liberality worry. It is not so serious as to constitute an objection to the permissive conception, I think, for reasons I will describe below. There are some abstract similarities between most things, and the permissive conception, perhaps surprisingly, picks out some abstract similarities that arguably hold between human brains and thermostats. I do bite the bullet here, but I would argue that insofar as this is a problem for the permissive conception it is an outstanding problem for cognitive science, as it is not clear what principled criterion could include all the cognitive phenomena but rule out the possibility of cognitive thermostats. More conservative philosophers suggest criteria such as non-derived content, or mereological principles ruling out the composition

of cognitive systems by components that are themselves cognitive systems (e.g. Block 1981), but appeals to non-derived content are no more principled than appeals to conservative intuitions about a speculative criterion for cognition. They simply legislate away complications that cognitive science grapples with. For example, these criteria rule out cases like extended cognition and group cognition, which remain subjects of controversy.

That being said, the permissive conception does provide better resources for ruling out cognitive thermostats than Dennet's intentional stance provides for ruling out thermostats with minds. Many liberal cases are excluded precisely because of our prior commitments about what counts as an organism, genuinely possessed of its own goals, on whatever basis justifies those prior commitments. The truth, though, is that the permissive conception is Janus-faced. It restricts "genuinely" cognitive systems to those that we construe as "genuinely" managing their behavior according to their sensitivity to their circumstances. Yet it also acknowledges the dimensions of continuity between the more plausible cases and the more marginal cases. I contend that this ambivalence is merely honesty about the complexity we find in the natural world.

6.4.3 On the contributions of the computational sciences

This ambivalence is also crucial for explaining a little-remarked feature of cognitive scientific practice. Most discussions about the problem of cognition in the present century focus on cases and evidence from psychology (and sometimes behavioral biology), but cognitive science includes other component disciplines, for example computer science and artificial intelligence (AI). Disputes about Herbert and other autonomous robots notwithstanding, most models in computer science and AI research are not taken to be cognitive systems, themselves. For example, the construction and evaluation of neural network models, Bayesian models, search algorithms, &c. are not taken to be in vivo experiments on cognitive systems. They are generally not taken to be detailed models of actual cognitive systems. Even Herbert was first defended as an existence proof

of a decentralized control architecture, not as a mad scientist exercise in artificial cognition. So in what way do these inquiries contribute to the study of cognition?

The problem of cognition is inter alia about which investigative practices count as contributions to cognitive science (§ 2.2). Both defenses and criticisms of controversial perspectives draw on empirical results and claims about theory-choice in science (e.g. Adams and Aizawa 2008; Clark 2008; Clark and Chalmers 1998; Rupert 2004; Rowlands 2010). Robert Rupert, for example, suggests that if extended cognition were real we would have to discard most of our extant findings since we would have discovered that we had such mistaken views on the nature of cognition (2004). Rowlands suggests that cognitive scientists should not be permitted to study phenomena that are not genuinely cognitive (2010). Those who endorse such a conception of the relation between the problem of cognition and the practice of cognitive scientific inquiry owe us a story about why topics in computer science, AI, and robotics are topics in cognitive science, whereas may other relevant topics in computation and algorithmic problem-solving do not belong to cognitive science. The proponent of the sensitive management hypothesis, by contrast, has a straightforward answer: computational and robot models are parts of cognitive science precisely when they model processes or mechanisms that, were they in a position to sensitively manage the behavior of an organism, would be cognitive processes or mechanisms. Thus, some contributions of computer science, AI, and robotics are straightforward contributions to cognitive science despite the fact that in general they do not involve tests on cognitive systems. And these contributions are possible precisely because of the continuity and qualified dimensions of similarity between cognitive and non-cognitive phenomena. Thus, the apparent liberality of the permissive conception is a strength, not a liability, of the permissive conception of cognition.

After all, I began Chapter 2 with a puzzle: Cognitive science is the study of a motley of capacities by a motley of disciplines. Why did we ever find it useful to think of all this inquiry as part of a common enterprise? I suggest that it is because of these continuities, the surprising extent to which some parts of this motley can inform other parts. The common stance that permits cognitive science to be so broadly interdisciplinary brings these disciplines together, and the border

wars have been cognitive science's struggle with the recognition that this stance is more flexible than was anticipated in the mid-20th century.

6.4.4 A broader view of the history

I have been concerned in this dissertation to restrict my attention to cognitive science, rather than conceptions of the mind and mentality more generally. However, the picture of cognition that is provided by the sensitive management hypothesis has significant affinities with other conceptions of mentality outside the context of cognitive science. Perhaps most apparent from the foregoing discussion is the extent to which the sensitive management hypothesis adapts resources in the belief-desire model of intentional action. I had occasion to remark on these affinities above, particularly in my discussion of the rationalization requirement on behavior (§ 6.2.3.1), and in my discussion of the quasi-intensional character of cognitive descriptions (§ 6.3.3). I do not mean to suggest at all that the sensitive management hypothesis provides an adequate response to long-standing philosophical puzzles about the mind-body problem in its most recalcitrant forms. This is certainly not the case—I have acknowledged frankly that the permissive conception of cognition offers little succor to those who are puzzled by human linguistic and conceptual performances, and I would add that the sensitive management view generally shares with cognitive science a paucity of resources to describe or understand phenomenal consciousness. However, I take it as an attraction of the sensitive management view that it might encompass conceptions of mentality that predate cognitive science. After all, the scientific concept of COGNITION is descended from more traditional conceptions of the mind. Therefore, if I can be forgiven for some simplified historical generalizations, I will remark briefly on the historical continuity between the sensitive management view and its predecessors.

Many conceptions of cognition identify it as part of a picture that Susan Hurley (1998) refers to as the "classical sandwich"—something that happens in between perception and action (or in between their more naturalized relatives, e.g. sensation and behavior). This is also the picture

received from computational functionalist accounts of cognition. In the functionalist version, cognition is cast as a computational process that mediates sensory inputs and behavioral outputs. However, older conceptions of cognition also presuppose a folk notion of LIFE. On Aristotle's account of the perceptive soul, for example, it is a subset of the living organisms that have capacities to perceive and act (Aristotle; see also Dewey 1958, Ch. 7). It may have been the power of the computer metaphor, as much as or more than any relic of behaviorist austerity, that diminished the strength of association between cognition and life. Surely, by the mid-20th century the failures of vitalism were overwhelming, and the promise offered by the computer metaphor of marrying normativity and naturalism were exciting. No one could blame computational functionalists for trying to explain cognitive performances in terms of physical structure, without mediation by means of biological categories.

However, the biology and cognitive science of recent decades have conspired to blur the line between cognitive science and biology. In particular, many cognitive performances are revealed to be less rationalistic and rule-based than was dreamt of by early AI researchers (see e.g. Turing 1950, 452f). Failures of globally integrative sensitivity are commonplace (q.v. § 5.5.2.2), and a significant proportion of research effort even in human psychology is concerned with failures of rationality. At the same time, the adaptive character of plant and microbe performances has been revealed to be much more impressive than was previously thought. These trends have no doubt contributed to the popularity of anti-classical perspectives that emphasize the continuity between human cognition, animal cognition, and the adaptive performances of non-animal organisms.

This is to say that the sensitive management hypothesis conforms to a long tradition of thinking about cognition (or minds) as devices for acting in response to sensitivity, and re-appropriates a pre-behaviorist tradition of thinking of minds as internally related to life. The resources of construal-based explication permit me to describe a conception of COGNITION that is organic, but lacks vitalist metaphysical commitments.

Another feature of the permissive conception of cognition is that it always admits of characterizing cognitive systems in terms of a thin sort of belief-desire psychology in which

sensitivity to circumstances replaces belief, goals replace desire, and the operations of mechanisms replace rationality. This is not to say that the sensitive management hypothesis (or cognitive science) is a pale imitation of folk psychology, however. The modification of folk psychological concepts into these particular scientific successors is motivated by decades of concerted theoretical efforts by cognitive scientists. And, of course, the most significant contributions of cognitive science to our understanding is not expressed in the broad commitments articulated by SMH, but in the development of models capable of describing the variety of sensitivities, behaviors, and mechanisms that belong to the various sectarian extensions of the sensitive management hypothesis.

6.5 SUMMARY

In this chapter I described the most permissive conception of cognition that I am willing to countenance. This permissive conception is one that does not eliminate construal-sensitivity from the parameters of the sensitive management hypothesis; the criterion it articulates does not determine a precise sectarian extension, but rather a stance: a criterion for which ways of thinking count as thinking of a system as a cognitive system, rather than a criterion for which bits of the world are cognitive systems. I argued that the permissive conception has the potential to illuminate a number of further facts about cognitive scientific practice. In particular, it has the potential to illuminate certain features of representation talk—including disputes over the existence of non-derived content, Ramsey's job description challenge, and the ubiquity of representation ascriptions by cognitive scientists—and the continuities it presupposes between the cognitive and the non-cognitive explain how computer science, AI, and robotics contribute to the cognitive scientific enterprise.

However, although the permissive conception purports to be a story about how to think of something as cognitive, it does suggest a story about what cognition is. It is not something magical,

or something with a true nature that we might one day discover (like the chemical composition of water). It's not even something all that specific. But it is an object that can be described with the analytical resources I described in the early part of this dissertation. On this story, the value of thinking about the world with cognition talk is not that there clearly is such a thing as cognition, at least in the same sense that the realist believes there are fundamental particles. Rather, the value of using cognition talk lies in the expressive potential of such language to articulate certain kinds of abstract fact. I suspect that this is the most honest way to think about the subject matter of cognitive science.

7.0 THE COGNITIVE SCIENTIFIC IMAGE

7.1 OVERVIEW

I set out to describe a Lewisian foundation for cognitive science, one that can support the patterns of conceptual practice we discover in cognitive scientific discourse by swaying along with them. The practices I set out to support with this foundation were the diverse norms for ascribing the concept COGNITION—the category that denotes cognitive science's object of inquiry. In this concluding chapter, I will review the foregoing discussion, redescribing my problem, the problem of cognition, and four benefits that might be secured by a solution. I will also describe the analytical techniques I developed to treat that problem and the solution I defended (§ 7.2). I will then reflect on the extent to which I succeeded in securing the four benefits I articulated (§§ 7.3–7.4) and make my final remarks (§ 7.5).

7.2 LEWISIAN FOUNDATIONS

Cognitive science has been beset for the last thirty years by what I called the cognition border wars—dramatic foundational disputes about the nature and extension of cognition, and the best ways to model cognitive processes (§§ 1.2.2, 2.2, 5.3). Scientific enterprises can generally stand just fine without philosophical accounts of their objects of inquiry, as biology continues on apace with no satisfying account of the concept LIFE. But the disputes of the border wars have resulted in far more dramatic variability among cognitive scientists regarding their norms of use for COGNITION

than is normally discovered among scientists regarding their conceptions of their disciplines. As a result, it has become a common refrain among some cognitive scientists that their field is in a "scandalous" state.

In Chapter 2, I argued that more is at stake than embarrassment. A successful explication of COGNITION has the potential to serve four kinds of benefit (Table 2.1): First, (I) it can settle the open questions of the border wars, and some of these questions bear on the decisions of publishers, funding organizations, and other means of producing or disseminating cognitive scientific research. Second, (II) an explication of cognition might make clear what we have learned about minds by doing cognitive science. I argued that the border wars are evidence of ongoing conceptual development in cognitive science, and while the ultimate course of development is not yet known to us it would be valuable and interesting to have resources for understanding the presuppositions that underpin contemporary cognitive scientific inquiry. Third, (III) a compact explication of cognition (or a compact slogan of a more complicated explication) would enable cognitive scientists to describe their work succinctly and effectively to students, other researchers, and the public. Finally, (IV) a clear expression of the scientific conception of cognition would serve as a more secure resource for appeal in philosophical discussions that turn on the contemporary science of the mind.

However, any attempt to construct an explication that serves these functions gracefully is troubled by the fact of widespread expert disagreement in the border wars and the fact that COGNITION is generally acknowledged to be at least somewhat polysemous. Thus, before introducing my favored explication I set out to expand the conceptual analyst's traditional toolkit. In Chapter 1 I made explicit the anthropological commitments that inform my project, and in subsequent chapters I described three variations on the traditional scheme for conceptual explication (q.v. § 3.2). In Chapter 2 I described ecumenical extensional adequacy, an alternative to the traditional conception of extensional adequacy. Instead of reproducing the genuine extension of a concept, or a more precise extension for a concept, an ecumenically adequate explication reproduces the patterns of agreement and disagreement in a discourse. In Chapter 3 and Chapter 4 I described analytical methods for constructing ecumenically adequate explications. Construal-

based conceptual explication (§ 3.3.2) achieves variable extensions by employing construal-sensitive terms in the explicans, the extensions of which vary according to the mental states of the construer. Parameterized explication (§ 4.2) achieves varying "sectarian" extensions (q.v. § 4.2.1) by employing parameterized predicates in the explicans, which can be assigned various interpretations corresponding to the background commitments of different judges. These three variations on traditional conceptual explication permit the construction of Lewisian conceptual foundations, and I put each of them to work in my explication of COGNITION.

In Chapter 5 I articulated my favored explication, the sensitive management hypothesis (SMH). This hypothesis holds, in a slogan, that cognition is the sensitive management of organismal behavior. The sensitive management hypothesis employs three parameters—*BELONGING TO AN ORGANISM*, *SENSITIVITY TO CIRCUMSTANCES*, and *BEHAVIOR*—which together subsume the diversity of background commitments that correspond to differing judgments regarding the extension of COGNITION. I defended this explication by producing a roughly representative collection of phenomena that are either paradigmatically cognitive, controversially cognitive, or generally agreed to be non-cognitive. I then argued that on various available interpretations of the three parameters, the explication includes the paradigmatically cognitive phenomena in every sectarian extension, excludes the agreed-upon non-cognitive phenomena from every sectarian extension, and selectively includes or excludes the controversially cognitive phenomena depending on which interpretations are assigned to its parameters. Thus, the sensitive management hypothesis is ecumenically extensionally adequate. In Chapter 6 I expanded on the most permissive interpretations of parameters for SMH, discussing their coherence. I argued that on this permissive conception, SMH articulates what it is to take a cognitive stance toward a system. I briefly discussed the potential of the cognitive stance to offer novel treatments of a few topics regarding the concept of NATURAL REPRESENTATION.

7.3 CONCEPTUAL DEVELOPMENT IN COGNITIVE SCIENCE

Now, before concluding, I will return to the four benefits (I–IV) I described in Chapter 2, in order to take stock of the extent to which the sensitive management hypothesis succeeds and fails to secure them.

7.3.1 Making explicit what had been implicit

I argued above that the scientific concept of COGNITION is developing—changing in response to empirical and formal evidence gathered by cognitive scientists. One of the benefits of characterizing that concept is an epistemological benefit (II), of articulating what it is that we have learned about cognitive science's object of inquiry, such that we have shifted the presuppositions of our inquiry. In the 1970s, before the border wars, it was a common view that cognition is something like the manipulation of semantically-rich symbols according to formal rules (Turing 1950; Fodor 1975; Newell and Simon 1976). I've referred to this view as the classical computational theory of cognition (CCTC; § 1.2.1). This view is still held by some (e.g. perhaps by Anderson 1993; Ramsey 2007), but the border wars have shattered that consensus, and it is no longer a dominant view informing the great majority of cognitive scientific research.

The sensitive management hypothesis provides an alternative articulation of the core presuppositions of cognitive science: that cognition is the sensitive management of organismal behavior. According to SMH, cognition is not essentially the manipulation of symbols but the operation of mechanisms that mediate an organism's sensitivity to its circumstances and its management of its behavioral performances. This conception of cognition countenances a great deal of continuity between the cognitive and the non-cognitive, and a different vision than classical views of the kind of metaphysical naturalism that is embodied by cognitive scientific inquiry (§ 6.4). In particular, SMH expresses a notion of COGNITION in which computational metaphysics is just one of many ways we might understand the normative dimensions of cognition. The

displacement of older conceptions of cognition—expressed by CCTC—by newer conceptions—expressed by SMH—reflects a recognition of the diversity of ways that mechanisms can mediate sensitivity and behavior, and the diversity of representational techniques employed by cognitive scientists in constructing models. I have not provided anything like a detailed survey of this diversity. However, as a parameterized explication, the sensitive management hypothesis can support the development of a catalogue of these techniques and their presuppositions through continued refinement of the available interpretations of its parameters.

7.3.2 Explication as a resource for appeal

I argued that an explication that achieves this epistemological benefit is also well-poised to secure two other benefits: a public relations benefit and a metaphilosophical benefit (§ 2.5). Regarding the public relations benefit (III), I contend that the sensitive management hypothesis, in providing a characterization of cognition that is faithful to the diversity of presuppositions and practices in the cognitive sciences, is a superior resource for communicating about cognitive science to non-experts. In particular, its slogan form (or an equivalent phrase) offers a compact expression of the nature of cognition according to an ecumenical view of contemporary cognitive science. The expression "cognition is the sensitive management of organismal behavior" is, in my judgment, no more cryptic than alternative slogans like "cognition is information-processing" or "cognition is what the brain does," and has the additional benefit that it is not so embarrassingly susceptible to liberal counterexamples (cf. § 1.2.2.1).

Furthermore, the parameterized structure of the explication offers a variety of paths for expanding that expression into a discussion of variation and controversy between scientists. I sketched these paths in my own discussion of the parameters (§ 5.5), first providing a picture of what the parameter requires in general (e.g. "circumstances are…") and proceeding to organize a survey of controversial views as potential interpretations for the parameter (e.g. "one interpretation of *SENSITIVITY* is that it requires…"). My discussion was of course rather academic in style, but it

would be straightforward to develop stylistic variations on this expository strategy for the purposes of communicating to the public, to students of cognitive science, or to academic researchers in other fields.

This public relations benefit extends to philosophy (IV). There are many philosophical controversies that turn, to varying extents, on the nature of cognition or of minds. In many cases (though not always), arguments concerning these topics depend on presuppositions about the conception of minds that we inherit from science. I discussed some of the sensitive management hypothesis' potential to contribute to disputes about natural representation in the biological and behavioral sciences (§ 6.3). However, there are many other philosophical topics where results from cognitive science prove relevant. Much discussion of the mind-body problem throughout the 20th century has drawn on scientific results or authority (e.g. Place 1956; Smart 1959; Putnam 1967b; Sellars 1981). Appeals to the results of cognitive science are also common in the literature on multiple realizability, scientific abstraction, scientific reductionism, and the nature of computation and representation. The sensitive management hypothesis offers a more secure and less partisan resource for appeal in arguments regarding these literatures than sectarian characterizations of cognition. Furthermore, the interpretation of parameters can be tailored to more specific dialectic contexts. And SMH certainly offers a more illuminating and less naïve picture of scientific practice than the most common resource for appeal in philosophy of cognitive science—computational functionalism (see Appendix). It is my hope that the sensitive management hypothesis proves to be fruitful in philosophical discussions beyond that concerning the "mark of the cognitive."

7.4 CONCEPTUAL DISAGREEMENT IN COGNITIVE SCIENCE

An ecumenical explication like the sensitive management hypothesis is less clearly poised to provide clarity about (I) the proper boundaries of cognitive science. It is ecumenically adequate precisely because it does not adjudicate the disputes of the border wars, so how could it serve to

adjudicate those disputes? I suggested earlier that an ecumenical approach offers indirect promise for settling boundaries (§ 2.5). The sensitive management hypothesis holds this promise, I think in virtue of the fact that parameterization provides three kinds of resource for dealing with controversy (§ 4.4).

First, a parameterized explication like SMH identifies the background commitments that contribute both to agreements and to disagreements regarding the proper norms of use for COGNITION. While each of the parameters of the sensitive management hypothesis admits of many more specific interpretations, my discussion of each parameter incorporated some necessary conditions. Organisms possess goals and mechanisms for pursuing those goals Sensitivity is characterized in terms of quasi-intensional organismal circumstances, and has robust effects on the operation of cognitive mechanisms, at least within certain background conditions. Behavior must be rationalizable in relation to organismal goals. I contend that these requirements describe universal norms of use for the concept COGNITION among cognitive scientists, and the recognition of these norms might sometimes inspire more conceptual sympathy between disputants.

Second, by making clear which commitments regarding cognition are shared and which are not, SMH provides a map for Quinean retreats. In sketching the common terrain upon which conceptual disagreements are played out, and encouraging partisans to make explicit their divergent commitments within a common framework, the sensitive management hypothesis also facilitates a process of what Brandom calls "elucidative rationality" (Brandom 2000, 56f). Making implicit commitments into explicit ones lays them out so as to be more open to fair and charitable criticism. And third, the sensitive management hypothesis provides a framework within which a "divide and conquer" strategy toward increased agreement may be conducted. Discussions about representation (subsumed, along with some related disputes, under the heading of *SENSITIVITY* in SMH) have received the lion's share of attention in the extant literature, while implicit tensions concerning the proper interpretation of *ORGANISM* and *BEHAVIOR* have been largely permitted to grow like mold in a dark, forgotten closet. I hold out hope that the sensitive management hypothesis may provide a salubrious measure of light and fresh air.

However, as an alternative to increased agreement in conceptions of cognition, it is also possible that cognitive scientists will (or should) embrace pluralism about the possible forms of cognition. Perhaps the cognitive stance (Ch. 6) is an adequate conception of cognition. Or perhaps the sensitive management hypothesis, with its open parameters, is sufficiently specific for cognitive scientific inquiry to continue operating productively, just as biologists operate productively despite having no definition of life and recalcitrant disputes over how to understand the details of natural selection. If these possibilities bear out, then there may be no need to adjudicate the disputes of the border wars definitively.

7.5 A COGNITIVE SCIENTIFIC IMAGE

One of the challenges of characterizing cognitive scientific practice is that it inhabits a crowded institutional neighborhood. First of all, cognitive science is an interdiscipline encompassing contributions from many traditional disciplines—particularly psychology and computer science, of course, as well as philosophy and linguistics, and to some extent biology and anthropology. So many phenomena, and of course many models, are shared between cognitive science and parts of its component disciplines. But cognitive science also competes for institutional space with other interdisciplines—most notably neuroscience, AI, and robotics. It is perhaps a cause for amusement that in their textbooks and other introductory or didactic materials, each of these interdisciplines (including cognitive science) claims to include each of the others as components. However, this state of affairs raises a question: what is it, really, that is distinctive about cognitive science? Why is it *a thing*, among so many similar things? I will not draw out this conclusion with a comparison between cognitive science and these other fields. However, I do think that I have provided the beginning of an answer to this question: the cognitive stance.

I remarked in my discussion of the cognitive stance upon its affinity with historical conceptions of the mind (§ 6.4.4), and upon its quasi-intensional character—the existence of

preferred descriptions for many things when construed in terms of the cognitive stance (§ 6.3.3). Organisms are causal systems construed as possessing goals and mechanisms. Circumstances are states of affairs construed in terms of organismal goals. Behaviors are events construed as means to organismal goals, and construed in contrast to other such events. I also suggested that many ascriptions of representational content in cognitive science are in fact inchoate ways of making explicit the quasi-intensional character of mechanisms, and that even the identification of biological and behavioral mechanisms in general requires scientists to construe causal systems in terms of explanandum phenomena. When all of these quasi-intensional descriptions are considered together, they form a picture—painted in scientific language, but depicting the world from the point of view of a cognitive system. This stance allows us to articulate, in scientific parlance, the form of life (q.v. Wittgenstein 1953; Thompson 2008) of particular organisms. I suspect that this potential plays a role in driving the popularity and endurance among cognitive scientists of J. J. Gibson's ecological approach—and in particular his theory of affordances—and the popularity of appeals to Jakob von Uexküll's notion of the *Umwelt* (at least among those with a taste for phenomenology; Gibson 1979; Uexküll 1957).

Wilfrid Sellars famously claimed that in modernity, we are left with two pictures of the place of humanity in the world: the manifest image, populated by persons and reasons, and the scientific image, populated by particles and laws (Sellars 1962). Myself, I think the scientific image is not as austere as Sellars suggests, and that it is a rather fractured pictured of the world: like looking at a reflection in a broken mirror. And in the cognitive scientific images of the world, we can see the diversity of forms of life inhabited by various organisms, and we can see the diversity of ways of describing and explaining the ways that organisms inhabit the world.

APPENDIX

RETHINKING FUNCTIONALISM AS A CENTRAL DOCTRINE IN COGNITIVE SCIENCE

A.1 FUNCTIONALISM AS A CENTRAL DOCTRINE

Functionalism is the doctrine that mental or cognitive states are functional states, whose identity conditions are articulable in terms of functional descriptions. Functional descriptions are abstracted descriptions that specify a state's characteristic inputs, outputs, and relations to other intermediate states (or similarly, their characteristic causes and effects). The canonical defense of functionalism as a framework for characterizing mental states was given by Hilary Putnam (Putnam 1967a, 1967b), but since then functionalism and its variations have become a central doctrine in the philosophy of cognitive science. Functionalism has been defended as a framework for understanding cognitive scientific practice (Fodor 1968; Chalmers 2011), and has been an important target of criticism of cognitive science (e.g. Searle 1980; Buechner 2011). It now comes in many versions (Levin 2013; Maley and Piccinini MS). Functionalism has been an embattled doctrine for decades (Block and Fodor 1972; Block 1980; Godfrey-Smith 2009b; Shagrir 2005), and the canonical argument for functionalism—the argument from multiple realizability—has been subjected to a variety of criticisms (e.g. Bechtel and Mundale 1999; Batitsky 1998). Functionalism is less central to much contemporary discussion than it once was (Chemero and Silberstein 2008), but it nevertheless retains the notoriety of an orthodoxy in philosophy of mind (Buechner 2011;

Jackson 1998) and in contemporary philosophy of cognitive science (Clark 2008; Sprevak 2009; Chalmers 2011; Eliasmith 2002) despite the fact that it is rarely defended from its many objections. I do not intend to review the viability of the extant objections in the present paper, in part because they seem to have been largely ineffective at moving the discussion forward. This is all, importantly, to say nothing of other views that happen, unhappily, to sometimes be called "functionalism" in biology or in pre-behaviorist psychology (e.g. James 1890; Cummins 1975; Sober 1985; Chemero 2009). The discussion in this appendix applies only to Putnam's machine functionalism and derivative views.

There are at least two reasons for functionalism's lingering popularity. The first is largely sociological: it is the fate of many "received views," such as the belief-desire model of intentional action or the deductive-nomological model of explanation, to remain central to a literature despite decades of convincing criticism so long as there is no successor sufficiently dominant as to attract a critical mass of discussion. The new mechanist model of explanation (Machamer, Darden, and Craver 2000; Bechtel and Abrahamsen 2005; Craver 2007) has recently achieved this status in the philosophies of the biological sciences, supplanting the deductive-nomological model and other law-subsumption models (q.v. § A.3) as a received view of explanation in those sciences. This is not to say that new mechanism is uncontroversial, only that in those literatures it has replaced other models of explanation as a primary object of interpretation and criticism. Functionalism, however, has largely retained its "received view" status.

The second reason for functionalism's lingering popularity is that it serves as a resource for appeal in so many different areas. My present focus is on functionalism's status in philosophy of cognitive science, so I will ignore the role it plays in the literatures on a priori philosophy of mind, philosophy of language, and naturalistic metaphysics generally. In the philosophy of cognitive science, functionalism still seems to serve as a ready resource for appeal regarding three topics: the ascription of cognitive kinds, explanation in cognitive science, and generalization in cognitive science. First, it is sometimes thought that cognitive scientific kinds—like memory, belief, object-recognition, speech perception, and so on—will be ascribed and individuated on the basis of

functional descriptions. So if a state of a system satisfies a functional description of memory, then the state can be said to be a memory state. If it turns out that there are distinct functional roles played by memory-like states, then it will be the case that are different kinds of memory (i.e. short-term memory, semantic memory, episodic memory). Second, it is sometimes thought that the primary explanatory task of cognitive science is to produce such functional descriptions. Functional descriptions can serve as lawlike statements describing the regularities of cognitive structure and processing. If the articulation of universal laws of nature is the most central explanatory activity of theoretical physics, then perhaps the articulation of functional descriptions is the correlative activity in cognitive science. Finally, since functional descriptions serve as criteria of ascription for cognitive kinds and are main constituents of cognitive explanations, they are sometimes also thought to be the basis of generalization. Cognitive kinds generalize insofar as their functional descriptions can be applied across states and processes and systems. An explanation of one system will generalize to other systems insofar as the functional descriptions can be successfully applied to those other systems. This is a species of the "covering law-subsumption" picture of scientific generalization. When I say that functionalism serves as a "central doctrine," I mean that functional descriptions serve as a prominent resource for appeal in all three of these topics. Since functionalism as a central doctrine purports to give a unified treatment to these three different topics, and in a way that has affinities with popular (functionalist or nomological) approaches to topics in philosophy of mind, philosophy of language, philosophy of science, metaphysics, and other philosophical subfields, there is considerable pressure to continue using functionalism as a resource for appeal in philosophy of cognitive science, objections notwithstanding.

Nevertheless, it would be salutary to finally abandon functionalism as a central doctrine. In this chapter I aim to motivate two conclusions. First, functional descriptions cannot simultaneously perform the three roles of articulating conditions of ascription for kinds, providing a framework for explanation, and providing a framework for generalization. Second, functional descriptions cannot perform any of these roles well. In arguing for this second conclusion, I will appeal to

differences between cognitive scientific practice and the functionalist's picture thereof. In order to make the discussion more concrete, I will begin by describing a recent dispute in philosophy of cognitive science in which functionalism serves as a resource for appeal by proponents of each side: controversy over the hypothesis of extended cognition. Consideration of the dispute will serve as a kind of philosophical case study informing arguments throughout this appendix. Drawing on Mark Sprevak's contribution to the dispute, I will argue that explanation and generalization require different kinds of treatment, and that both explanation and generalization require different treatment than the ascription of cognitive kinds. However, Sprevak's discussion makes little contact with scientific practice, so the conclusions I draw from his discussion will require further refinement. Next I will describe functionalism's affinity with covering-law views of explanation, an affinity which undergirds its ambition to play each of its three roles. Since the covering-law view of explanation is a poor account for understanding the practices of the biological sciences, functionalism's ability to play any of its three roles is undercut. Bearing in mind the tenacity of received views, however, my aim in this appendix is not simply to poke more holes in the sinking ship of functionalism. I will conclude with reflections on the resources that mechanism offers for replacing functionalism, particularly with respect to practices of generalization. I will stop short of endorsing mechanism as a new "central doctrine." While the new mechanists have resources for addressing each of the three topics treated by functionalism, the picture we get from the mechanists is not as simple as the picture offered by functionalism. Nevertheless this is progress, since it is natural to accept that as our understanding of cognitive scientific practice becomes more sophisticated, simple but flawed conceptions of that practice will give way to more nuanced views.

A.2 FUNCTIONALISM AND EXTENDED COGNITION

Andy Clark and David Chalmers (1998) contend that cognition (like meaning) ain't all in the head. They argue that in certain cases the use of external props in some activities—a computer processor

while playing some video games, one's notebook while carrying out one's plans for the day, perhaps one's partner while remembering past events—is such that those props should be considered parts of one's own cognitive economy, similarly to parts of one's brain. This claim is sometimes called the hypothesis of extended cognition (HEC). Their best-known example concerns Inga and Otto. Inga is a healthy woman who one day decides to see an exhibit at MoMA in New York City. She has a desire to see the exhibit, a belief about where MoMA is and how to get there, and she acts accordingly. Otto is an older gentleman with a poor memory who uses a notebook to help him remember facts and plans.[32] Otto similarly decides to see the exhibit, then uses his notebook to navigate to MoMA. In their argument, Clark and Chalmers appeal to what has become known as the "parity principle," which states that

> If, as we confront some task, a part of the world functions as a process which, *were it done in the head*, we would have no hesitation in recognizing as part of a cognitive process, then that part of the world *is* (so we claim) part of the cognitive process. (Clark and Chalmers 1998, 8)

One way to interpret this principle is as a corollary of functionalism: cognitive states are individuated by their functional relations (to inputs, outputs, and each other), and it is immaterial whether their realizers are located inside the brain or outside the body.[33] Thus, activities should count as cognitive processes if those body-external processes exhibit the same functional relationships (to inputs, outputs, and cognitive states) as other processes that we already happily consider cognitive processes. Of course, understood this way the parity principle only justifies a

[32] It is worth observing that the Otto example is originally presented not as an illustration of extended cognition, but as an illustration of the extended mind, which is a distinct claim subject to qualifications that HEC is not. However, although this distinction is important for charitably evaluating Clark and Chalmers' arguments, it is rarely observed in the critical literature (even by Clark himself). Since I am not evaluating HEC here, I will ignore the distinction for ease of exposition.

[33] This claim accepted by most of Clark and Chalmers' critics (Adams and Aizawa 2001; Rupert 2004; Sprevak 2009), and is therefore not quite satisfactory as a charitable interpretation of the parity principle, since the parity principle is supposed to imply HEC.

commitment to extended cognition if the functional relations are specified so that body-external activities and props do satisfy those specifications, and many cognitive and psychological processes can be specified in a variety of ways. Fred Adams and Ken Aizawa (2001) argue that Otto's use of his notebook cannot count as memory because it must be described via inter alia relations to perceptual and motor intermediaries (he flips through is notebook, reads it, &c.) whereas canonical examples of internal memory are not related to perceptual and motor activities in this way.

Robert Rupert argues for the same conclusion because Otto's use of his notebook fails to satisfy the most fruitful functional description of human memory. Cognitive psychologists have documented many features of human memory—for example susceptibility to interference effects, generation effects, and conformity to the Rescorla-Wagner law (see Rupert 2004, 413–419). Interference effects may serve as an example. Say that experimental participants are asked to memorize a list of given names and corresponding surnames. If the participants are then told that some of these people have changed their surnames, and given a second list of given names and surnames, participants make more mistakes when remembering the second list than they do remembering a list of totally new given names and surnames (Anderson 2000, 239–243; Bower 2000). Otto's notebook-assisted "memory," however, does not exhibit these interference effects like this. If he were to write down the second list of modified name pairs, he would "remember" it just as well as any list. Otto's notebook therefore does not exhibit the same functional relations as Inga's biological memory, since it regulates Otto's behavior differently. Since Otto's notebook-assisted memory has a different functional description than "brainbound" human memory, our functionalism-based interpretation of the parity principle does not license the attribution to him of "extended" memory processes.

Mark Sprevak (2009) calls these objections the RAA (for Rupert, Adams, and Aizawa) objections. Notice that the appeal to functionalism in the RAA objections is supposed to work because of functionalism's status as a central doctrine. The RAA objections appeal to the role that functional descriptions appear to play in explanations of cognitive capacities and to the fact that such descriptions do not generalize to "extended" cases in order to support claims about how to

properly ascribe cognitive kinds. That is, claims about the ascription of kinds according to functional descriptions are supported by appeal to intuitions about the generalizability of descriptions based on scientific explanations. Sprevak's discussion of HEC and the RAA objections gives us resources for distinguishing the role that functional descriptions play in contexts of explanation and generalization. He suggests that "All varieties of functionalism contain a parameter that controls how finely or coarsely functional roles should be specified (how much should be abstracted and ignored)" (Sprevak 2009, 510). Fine-grained functional descriptions, which are relatively detailed descriptions of the causes, effects, and operation of cognitive processes, serve many paradigmatic explanatory purposes in cognitive science. Thus, explanations of human memory incorporate fine-grained properties such as interference effects, and so on. Sprevak observes that the RAA objections trade on *fine-grained* differences between Otto's use of his notebook and canonical examples of memory. However, fine-grained functional descriptions are inappropriate criteria for ascribing cognitive kinds like MEMORY. Imagine scientists soon discover a human who does not exhibit interference effects and so on. Sven Walter observes

> The proper response on behalf of practicing cognitive scientists would obviously be to say "What an amazing and unusual kind of memory!" and not "What an amazing and unusual performance that far exceeds our own mnemonic capacities—and that without any memory at all!" It is thus plausible to assume that cognitive processes need not be individuated so finely that the differences between the Otto-cum-notebook system and Inga's biomemory... inevitably lead to substantially different cognitive kinds. (Walter 2010, 291)

After all, the exhibition of interference effects is not a requirement on mnemonic capacities per se, but an empirical fact about human memory performance. Insofar as we are tempted to treat MEMORY as a kind of capacity exhibited by non-human systems, it is inappropriate to generalize from the fact that humans exhibit interference effects to the conclusion that any genuine memory

exhibits similar effects. In fact, it is inappropriate even to take the exhibition of such effects as requirements on human memory, for Walter is right that in his vignette, the correct response is not to deny that the unusual subject has memory, or to claim that she has inhuman memory, but to say that she has an unusual kind of human memory. Thus, while it may be fruitful for cognitive scientists to develop fine-grained functional descriptions of cognitive kinds, it is inappropriate to use the most fine-grained functional descriptions as criteria for ascribing cognitive capacities such as memory.

Sprevak describes a related objection to the use of fine-grained functional descriptions as criteria for ascribing cognitive kinds. He suggests that such descriptions conflict with the common intuition that there could be Martians who have cognitive processes but whose cognitive architecture is quite different from ours.[34] Such Martians, unlike us, may not exhibit interference or generation effects, and may even store information by manipulating ink-marks on paper inside of their brains, and retrieve it by reading the marks back with photosensitive organs. The "Martian intuition" is that this is an alien form of memory, but that it is memory nonetheless. In order to accommodate generalization to intuitively possible Martian cognition, functional descriptions of cognitive states must be suitably coarse-grained. Since such possible Martians have memory, and their memory may have the same fine-grained functional description as Otto's actual use of his notebook, we must ascribe kinds such as MEMORY according to suitably coarse-grained descriptions. Therefore, the parity principle demands that we consider Otto's a case of extended cognition, and functionalism implies HEC. This is a defense for HEC that Rupert (2004) anticipates and that Clark (2008, 98–99) appeals to. However, Sprevak argues that coarse-grained, Martian-countenancing functional descriptions are also unacceptable criteria for ascribing cognitive kinds. Since we can imagine many far-fetched Martian cognitive architectures, the parity principle licenses an unacceptably radical form of HEC. For example, we might imagine ink-mark-manipulating Martians like those above, but which are born with factual information encoded in

[34] Sprevak draws the device of "Martian intuitions" about conceivable but non-actual cognitive architectures from Clark's own defenses of extended cognition (2005, 2008).

their brains "innately." Imagine further that such Martians may not examine or become aware of this information for years, but that the information can be retrieved with effort. Such a memory process has a coarse-grained functional parity with the human process of looking up information in a library or archive. Such possibilities seem to license radical cases of extended cognition: that contents of volumes in a library are beliefs of any person in the library, or that being in possession of a graphing calculator gives one a knowledge of integral calculus (Sprevak 2009, 517–518). These consequences, Sprevak argues, are absurd, and justify a reductio of radical HEC and, since it entails radical HEC, of functionalism. In fact, Sprevak argues there is no setting of the grain parameter on which acceptable functional descriptions of cognitive kinds can be formulated, since there are always possible Martians corresponding to cases of tool use that it would be absurd to consider purely cognitive processes. Relatively fine-grained descriptions are chauvinistic (i.e. human-specific), coarse-grained descriptions are absurd, and descriptions with intermediate levels of grain are both chauvinistic and absurd.

I am sympathetic to Sprevak's conclusion, but his argument depends on considerations that are mostly independent from features of scientific practice. Before I discuss the connection between functionalism and practice I would like to draw attention to three consequences of the above discussion. First, the contexts of explanation for specific kinds of system (e.g. humans) and generalization of kinds across different kinds of system (e.g. humans and Martians) often require different functional descriptions—relatively fine-grained and coarse-grained descriptions, respectively. (See Shea, MS for a discussion of the distinction between species-specific and generic functional descriptions.) So a functional description cannot generally serve as a suitable resource for appeal regarding both the explanation of a system and generalization across kinds of system. Second, fine-grained functional descriptions used in explanations are not apt to be conditions of ascription for cognitive kinds, since they are often chauvinistic. Walter's vignette about the proper response to an unexpected case also supports this claim. So functional descriptions that might support explanatory projects are often unsuitable for settling criteria of ascription. Finally, coarse-grained functional descriptions suitable for generalization across different kinds of system are also

not suitable for conditions of ascription, since this would have absurd consequences about libraries and graphing calculators. Thus, if functional descriptions can fulfill any of functionalism's three roles as a central doctrine, the same functional descriptions cannot fulfill any of the others. In the following sections I will draw on recent philosophical results about explanatory practices in the biological and behavioral sciences that will require me to refine these conclusions (see especially § A.3.3).

A.3 LAYING DOWN THE LAW

The dispute over HEC is not seen by its partisans as a purely philosophical discussion, but as a battle for the soul of cognitive science. If HEC is true, it is claimed, it has dramatic consequences for the way cognitive scientists conduct their research. Hence, both defenses and criticisms of HEC draw on empirical results and claims about theory-choice in science (e.g. Adams and Aizawa 2008; Clark 2008; Clark and Chalmers 1998; Rupert 2004; Rowlands 2010). The fact that so many of the arguments concerning HEC trade on interpretations of functionalism reveals the common belief that functionalism provides a suitable framework for understanding cognitive scientific models. Traditional formulations of functionalism owe a great deal to the contemporaneous dominance of the covering-law conception of explanation and generalization (Hempel 1965; Hempel and Oppenheim 1948). Putnam's functionalism in its original form is primarily a view about how to individuate mental kinds, especially folk-psychological kinds like pain and preference. However, in the philosophy of cognitive science it has become a master doctrine with implications for the ascription of cognitive kinds, the form of cognitive scientific explanations, and the practice of generalization. It has remained an influential view because it purports to provide a link between these topics, and its role in adjudicating disputes about HEC depends on this putative link. However, functionalism's contribution to all three topics depends on its relation to nomological conceptions of explanation and generalization, which are untenable. In this section I will discuss

functionalism's affiliation with nomological views in connection with criteria of ascription for kinds and with explanation.

A.3.1 Functionalism, laws, and kind-ascription

Functionalism in its original form is primarily a view about how to individuate mental kinds, especially folk-psychological kinds like pain and preference (Putnam 1967b). Putnam's multiple realizability argument works essentially by demonstrating that functionalism is appropriately *general*, and in particular that functionalism is more general than type physicalism. A type-physicalist account of pain (Place 1956; Smart 1959), like the simplistic conjecture that pain is the activation of C fibers, denies without motivation that animals that lack C fibers have pain-states. Putnam argues that functionalism achieves the appropriate level of generality by proposing abstracted descriptions that omit physiological and other details. Sprevak's grain parameter makes this maneuver more explicit by proposing a continuum of descriptions that are increasingly abstract, in the sense of omitting detail, and therefore increasingly general.[35] Imagine a toy functional description of pain: pain is caused by tissue damage, and causes stress, increased metabolic activity, evasion of the damaging stimulus, and avoidance of the damaging stimulus in the future. This description denotes processes in a variety of complex organisms, including primates and cephalopods (whose afferent nerve fibers are not classed into A, B and C groupings). Now elements can be added to this description to make it more fine-grained, and to denote processes in progressively more restricted classes of organisms. For example, if we specify that pain also tends to cause excited vocalization, we have a more detailed and therefore more fine-grained functional description of pain. However, creatures like cephalopoda do not vocalize, and will therefore not satisfy this more fine-grained description of pain. Thus, the addition of detail to a

[35] Sprevak acknowledges that this "grain" continuum is a simplification of a multidimensional space of variation—specifically in which kinds of details are abstracted from—but argues that it makes no difference to his argument (Sprevak 2009, 510n12).

description can be a means to achieving greater specificity, and the omission of detail a means to greater generality.

Functionalism settles the ascription of cognitive kinds by means of their role in lawlike statements, perhaps on the model of ENERGY in physics—the meaning of ENERGY can be exhaustively described, it is sometimes claimed, by citing its role in various physical laws (cf. Lewis 1972, 254). Conditions of ascription for cognitive kinds may be expressed by means of the Ramseyification method described by David Lewis (1972), with these lawlike statements in place of platitudes.[36] If we return to our example of memory, the Ramsey sentence for memory might state that there is something such that it is described by the Rescorla-Wagner law, and is subject to interference effects, and is subject to generation effects, and so on, and that thing is memory. We may imagine an ideal description, *M*, that is structured as a Ramsey sentence and includes conjoined clauses that articulate all the well-confirmed effects and features of mechanisms of human memory.[37] Generalization for functionalism is also understood on the covering-law model, so that an item is understood as belonging to a kind if its behavior is subsumed under a covering law. So a system counts as a memory system just in case it is described by *M*. Functionalism cannot serve this purpose, however, because Ramsey sentences like *M* are inappropriate criteria of ascription for cognitive kinds. Sprevak's argument, described in the previous section, implies that any functional description articulating criteria of ascription for cognitive kinds is either chauvinistic, absurd, or both. If the lawlike statements of cognitive science, such as the Rescorla-

[36] The relation between Ramsey sentences and lawlike statements is straightforward. Suppose we have laws or lawlike statements L_1, L_2 … L_n that mention a functional kind *F*. These statements may be, as in the case of memory, expressions of the Rescorla-Wagner law, of interference effects, &c. We may construct an appropriate Ramsey sentence in two steps. First, replace any reference to *F*s in L_1 … L_n with a logical variable (e.g. x) yielding open sentences $L_1'(x)$ … $L_n'(x)$. Second, conjoin the open sentences and existentially bind the variable, e.g. $\exists x\,(L_1'(x)\ \&\ \ldots\ \&\ L_n'(x))$. The resulting sentence is a Ramsey sentence constructed from the laws or lawlike statements in which the definiend appears. Notice that on this scheme, lawlike statements need not be universally quantified (as paradigmatic laws are, but plausible psychological or cognitive laws may not be). The procedure here is essentially the one described by Frank Jackson (1998), with empirical generalizations or lawlike statements in place of platitudes, and omitting his a priori method for revising Ramsey sentences.

[37] It may be interesting to note that appeals to functionalism in cognitive science almost never involve articulating full functional descriptions in the form of Ramsey sentences. Rather, they appeal to elements of functional descriptions that are diagnostic for a restricted class of cases under discussion. I follow this custom without further comment.

Wagner law, serve as elements of criteria of ascription for cognitive kinds, the RAA objections would have been trenchant objections to HEC. However, Walter's vignette pumps our intuitions that cognitive scientists do not in fact treat their own conclusions in this way. And since a description like *M* is plainly chauvinistic (as shown by Sprevak's grain argument), it is inappropriate to use *M* as a basis to reject the generalization of kinds to non-human systems. If functionalism does provide a framework for characterizing cognitive kinds based on results in cognitive science, it does not do so in the way Putnam imagined. That is, it does not provide non-chauvinistic criteria of ascription, since descriptions constructed from cognitive "laws" and effects surely are chauvinistic in a way that makes them unfit for Putnam's project. In fact, it is also inappropriate to use such descriptions as criteria that rule out new forms of human memory (as shown by Walter's vignette). Upon finding a plausible instance of human memory that is not described by *M*, the proper response is to investigate an unusual case, not to deny that the case is a case of memory because it differs from previously-described cases of memory. And if it is good practice to regard instances of memory that do not conform to *M* as unusual cases, there must be some criterion other than *M* (perhaps not precisely describable) for identifying instances of memory—hence, *M* does not serve as a criterion of ascription even for a plainly chauvinistic kind like HUMAN MEMORY. In order to more soberly evaluate the potential that functional descriptions have for providing criteria of ascription for cognitive kinds, I will follow Robert Cummins in distinguishing between various categories of cognitive kinds below in Section A.3.2.

A.3.2 Functionalism, laws, and explanation

It might be thought that even if functionalism fails as a strategy for describing cognitive kinds, functional descriptions can function as lawlike statements that contribute to explanations in cognitive science, according to some covering-law view of explanation. On covering-law views, successful scientific explanations are sound arguments in which laws of nature and descriptions of start conditions or circumstances are premises, and the description of the explanandum

phenomenon is the conclusion. The view I am considering here is a functionalist covering-law view, according to which the role of laws in the covering-law view is played, in cognitive science, by functional descriptions. There are a number of well-known objections to the covering-law view. I shall very briefly review three: that expectability on the basis of such a deduction is not sufficient to explain, that it is not necessary to explain, and that it relies on the questionable notion of a law.[38] First, such deductions are not sufficient to explain. To take an example made famous by Wesley Salmon, one can deduce the length of the shadow cast by a flagpole if one knows the height of the flagpole and the angle of the sunlight. However, one can also deduce the height of the flagpole or the angle of the sun in the sky given the other two pieces of information even though it is not the case that the length of the shadow causes or explains either of these things (Salmon 1989, 47). Second, deductions of the sort specified by the covering-law model are not necessary to explain. Many biological phenomena appear to be, so far as we can explain, probabilistic. In another of Salmon's examples, untreated syphilis results in paresis 20% of the time. So Mr. Jones' paresis is explained by his having syphilis and his not being treated. However, the covering-law model does not account for this explanation because one could not deduce that he would have paresis on the basis of those premises. Nor is it even the case that the deductive-nomological model can be easily softened. For instance, given the facts about Mr. Jones it was not more likely than not that Mr. Jones would develop paresis (Salmon 1984). Third, even if revised versions of a covering-law view could overcome such general objections they would require an antecedently plausible account of scientific laws. There are familiar controversies about how to distinguish laws from accidental universal generalizations (Hempel and Oppenheim 1948; Dretske 1977; Lewis 1983). Nevertheless, I am happy enough to bracket these familiar objections to covering-law views in general. I have a special objection to the functionalist covering-law view in cognitive science: that they fail to properly distinguish the explananda of cognitive science from the explanantia.

[38] These and other objections are reviewed in a different form quickly but with more care in Craver (2007), pp. 34–40.

Table A.1: Cummins' explanatory distinctions.

Cummins' explanatory distinctions	
A. psychological capacities	*primary explananda*
B. psychological laws and effects	*secondary explananda*
C. psychological mechanisms	*explanantia*

To begin, it appears that however scientific explanation works in the physical sciences, explanation in the biological and behavioral sciences does not normally involve appeal to laws at all (Cummins 2000; Bechtel and Abrahamsen 2005; Woodward 2000, 2001). Rather, according to Cummins, the generalizations called "laws" in psychology describe patterns of phenomena that are to be explained, not generalizations that explain phenomena. The Rescorla-Wagner law, for example, describes changes in patterns of learned associations, not the causal processes that produce those patterns. It provides a quantitative framework for predicting memory performance, but does not explain how the systems it describes exhibit a capacity to learn. Similarly, interference and generation effects on memory do not explain how cognitive systems exhibit mnemonic capacities; they describe surprising features of memory performance.

Cummins makes some distinctions that I propose to adopt, though the glosses I provide are my own. Cummins distinguishes between (A) psychological or cognitive capacities, (B) laws and effects in psychology and cognitive science, and (C) mechanisms or the structure of cognitive systems (Table A.1). Cognitive capacities are the primary explananda of cognitive science: the capacities of systems to exhibit intelligent behavior, and certain supporting capacities such as those for perception, memory, inference, and so on. It is the main explanatory task of cognitive science

to explain how organisms exhibit these capacities. Effects are patterns exhibited by cognitive systems that are incidental to the manifestation of cognitive capacities *per se*, but that are nevertheless characteristic of the way that certain systems manifest cognitive capacities. For example, the McGurk effect (McGurk and MacDonald 1976) is a pattern exhibited in human phonemic perception when visual and auditory information conflict. The "laws" of cognitive science typically describe effects. The articulation of such laws and effects is one of the main activities cognitive scientific inquiry. However, Cummins claims that the cataloguing of such "effects" does not explain the manifestation of cognitive capacities (i.e. the primary explananda). Rather, effects provide additional constraints on successful explanations of cognitive capacities, and thus they are "secondary explananda." Finally, cognitive mechanisms are the constitutions and organizations of particular systems in virtue of which they exhibit their cognitive capacities and effects. Explanation in cognitive science requires the description of mechanisms—whether by analytical explanations, as Cummins argues,[39] or by the articulation of mechanism models, as argued by the new mechanists. I find it felicitous to refer to the products of cognitive science—be they phenomenal laws, descriptions of effects, "theories," or characterizations of mechanisms—as models, and shall follow this convention for the remainder of this appendix.[40]

The functional descriptions appealed to in the RAA objections are constituted by the description of capacities, effects, and mechanisms, mixed together. In particular, Rupert appeals not only to the capacity to remember, but the contribution of memory capacities to other capacities

[39] Cummins uses the phrase "functional analysis" to describe his favored form of explanation, but the homonymy of "functional" in "functional analysis" and "functional description" is misleading. Cummins' functional analysis involves explanations of the capacity of a system by appeal to sub-capacities. Although his 2000 article preceded the vogue of new mechanism, his view is broadly consistent with it (but cf. Craver 2001). His defense of functional analysis as an explanatory framework appeals to explanations that describe the "structure" of the system being explained, and he uses the word "mechanism" to refer to the true structure of the system. Functional analysis does not involve appeal to functional descriptions in the sense under discussion. Indeed, one of the main conclusions of Cummins 2000 is that explanation in psychology does not proceed by appeal to laws.

[40] It might be tempting to assume that the distinction between capacities, effects and mechanisms maps neatly onto David Marr's (1982) distinction between the computational, algorithmic, and implementational levels. In particular, it is tempting to imagine that capacities and effects are descriptions at the computational level, and that mechanisms will be described at the algorithmic or implementational levels. However, this is a problematic mapping. There can, for example, be neurological effects, viz. fMRI localization effects in moral judgment described by Joshua Greene and colleagues (2004).

such as conversation. Rupert also appeals to effects such as interference effects, generation effects, and the Rescorla-Wagner law. Adams and Aizawa also appeal to parts of mechanisms, such as the role that vision and motor control play in reading. Since our ideal functional description *M* of memory includes all well-confirmed effects and parts of mechanisms, it contains a mixture of explananda and explanantia. Nomological conceptions of explanation in which functional descriptions like *M*, or the empirical statements that are used to construct them, serve as lawlike statements are poorly suited to describe explanation in cognitive science because they do not distinguish between capacities, effects, and mechanisms. That is, they do not distinguish between the explananda and the explanantia of cognitive science. For example, Rupert's appeals to interference effects are indeed appeals to empirical discoveries, but they are discoveries of secondary explananda, not of explanantia. Compact descriptions of such effects are scientific achievements, but they are not parts of explanations of mnemonic capacities in humans or any other creatures. They are part of the explanandum phenomemon of memory in humans. They thus cannot be used in explanations of the capacity for memory, on pain of circularity. Only descriptions of mechanisms (whether by analytical description, mechanist description, or some other format) can explain the exhibition of explanandum capacities and effects. That is, an explanation of human memory must specify how we—our brains or cognitive architecture—are constituted so as to exhibit our mnemonic capacities, and also various effects.

A functional description that denoted only elements of mechanisms, and not explanandum capacities or effects, would avoid this objection. But such restricted descriptions are not generally appealed to by functionalists, and have not been described in the discussion of HEC. Furthermore, the Ramsey-Lewis method for producing functional descriptions involves Ramseyifying an entire theory. If we construct functional descriptions from a principled selection of cognitive scientific results rather than our fullest understanding thereof, we depart from the Ramsey-Lewis method in manner that is significant and (so far as I know) never remarked upon. The functionalist may reply further that these objections can be answered by reconstructing a version of functionalism that does not presuppose nomological conceptions of science. However, the need for such

reconstruction only corroborates my claim that functionalism developed in close association with nomological views.

A.3.3 Functionalism, models, and generalization

I suggested above that the products of cognitive science are models—primarily models of effects, and models of mechanisms. The functionalist may reply that the proper view of generalization in cognitive science is indeed a covering-*model* view, on which models generalize in place of laws. That is, the systems that have psychological states are just the systems that can be subsumed under the models that cognitive scientists have produced to characterize those states. There are stronger and weaker ways to understand this claim. On the stronger claim, the covering-model view provides criteria of ascription for cognitive states in general by invoking the empirical findings of cognitive scientists, just as the functionalism of the RAA objections did. Thus, a model on which memory exhibits interference effects articulates a restriction on which systems have mnemonic capacities, or humanlike mnemonic capacities. To wit, only those systems whose behavior is describable, at least in certain circumstances, by the interference model should be considered to have memory, or humanlike memory. This view suffers from the same flaw that the RAA objections do. Many of the models or claims in question were only constructed in order to describe the behavior of certain systems—usually humans, or humans in certain contexts (e.g. without external memory aids). Walter's vignette is an intuition pump that we should not use such findings as criteria of ascription for cognitive capacities, generally. The weaker way to understand the claim above is that the coverage of models somehow provides criteria of appropriate ascription for the model. Thus, a model on which memory exhibits interference effects articulates a restriction on which systems exhibit memory with interference effects. Any system appropriately described by such a model can be characterized as a system that exhibits interference effects, and no system exhibits interference effects unless its behavior is described by the model (or perhaps a competing model of interference effects). Proper evaluation of the weak covering model claim depends on how

one determines which systems are properly described by a model, but I will not object to the claim here.

However, acceptance of the weaker claim does imply that I must revise a claim I made above: empirical descriptions developed by cognitive scientists can serve as criteria of ascription for cognitive kinds. Nevertheless, two caveats should be observed for this revised claim. First, cognitive models can provide criteria of ascription for effects or for mechanisms. The weak covering-model claim does not imply that such models provide criteria of ascription for cognitive capacities, like memory or perception in general, since cognitive scientists do not develop models of capacities in general. They develop models of the mechanisms by which restricted classes of system (e.g. humans) manifest those capacities, and models that describe the effects present in the performances of those restricted classes of systems. Thus, the appeals to empirical findings made in the RAA objections are still inappropriate as objections to the liberal ascription of cognitive capacities such as memory. Interpreted as objections to the ascription of cognitive effects or mechanisms, the RAA objections show too little: proponents of extended cognition expect extended cognitive processes to have distinguishable mechanisms and effects (Clark 2008). The second caveat is that acceptance of the weak covering-model claim does not vindicate appeals to functionalism as a central doctrine. I argued above that functional descriptions do not do fundamental explanatory work in cognitive science, and that even models do not give us criteria of ascription for cognitive capacities. Furthermore, the practice of generalization in cognitive science, and the biological and behavioral sciences generally, is more complicated than portrayed by the weak covering law claim. Different systems with the same cognitive capacities may implement different mechanisms, and exhibit different effects. Furthermore, it is sometimes more fruitful to characterize the dimensions of similarity and difference between mechanisms, rather than classifying them as simply identical or distinct in relevant respects. These facts complicate generalization in cognitive science, so that it is often not possible to predict, in advance of investigating many kinds of system under many conditions, whether a model will generalize. I will discuss these complications below, in connection with new mechanist accounts of generalization.

A.4 NEW MECHANISM AND COGNITIVE SCIENCE

In recent years the covering-law view has been supplanted by the new mechanist view of explanation, at least in the biological sciences. The new mechanists hold that many scientific explanations, including a preponderance of explanations in the biological sciences, are achieved by specifying models of mechanisms. Let us suppose that the primary explananda of cognitive science are intelligent behaviors or cognitive capacities. Intelligent behavior is behavior that is sensitive to the circumstances of an organism and that can be rationalized by its relation to a goal of the organism; cognitive capacities are capacities that are exhibited in intelligent behavior. A cognitive mechanism, then, is a structure of component entities and component operations that are organized such that they produce intelligent behavior (§ 5.4 above).[41]

The mechanist view of explanation, unlike the functionalist covering-law view, can distinguish between the explananda and explanantia of cognitive science. Mechanisms are entities and operations organized such that they exhibit an explanandum phenomenon, and cognitive mechanism models are models of systems that produce behaviors, or exhibit capacities or effects (possibly or actually—see Craver 2007, 112–113). My appeal to Cummins' distinctions between the explananda and explanantia (Table A.1) is not question-begging here. Although I (like Cummins and many cognitive scientists) use the word "mechanism" to denote the working structure of cognitive systems, the distinctions describe facts about cognitive scientific practice and are not defended as consequences of a mechanist view of explanation. The mechanist view imposes restrictions on the structure of mechanisms—that they are composed of entities and operations that are structured in describable ways—that need not be accepted by everyone who uses the word "mechanism." Indeed, Cummins articulated these distinctions before the new mechanist view was first popularly articulated (Machamer, Darden, and Craver 2000), and does not himself argue for

[41] The extension of the mechanist model to cognitive science requires the suppression of certain assumptions developed for biological contexts (especially certain assumptions of Craver 2007). See (Weiskopf 2011) and (Chirimuuta 2014) for criticism. Not all mechanists make these assumptions (cf. Machamer, Darden, and Craver 2000; Bechtel 2008).

the mechanist view. Furthermore, my present contention is not that new mechanism is the only view of explanation that is consistent with Cummins' distinctions, only that it is one. Cummins' account of analytical explanation is putatively an alternative (but see Craver 2001 for a discussion of analytical explanation and new mechanism).

So mechanism offers a more faithful and less problematic framework for understanding explanatory practices in cognitive science. However, new mechanism is not a new central doctrine. Mechanism does not purport to provide criteria of ascription for cognitive capacities, and therefore would not have been an appropriate resource for appeal, as functionalism seemed to be, for objections like the RAA objections. What about generalization? The covering-law view of explanation has a companion view of generalization: that an explanation of one target system or phenomenon generalizes insofar as other target systems or phenomena can be subsumed under the same laws. Thus, Putnam's functionalism sought to identify the realm of mental phenomena with laws that were more general than physiological laws, i.e. with descriptions that denote more kinds of target system than physiological descriptions. According to this picture, the description of characteristic causes and effects constitutes the articulation of psychological laws, and the systems that have psychological states just are the systems that can be subsumed under the laws describing the causes and effects of those states.

I offer no a priori objection to the covering-law view of generalization. However, William Bechtel and Adele Abrahamsen (2005) have argued that this picture is not faithful to the practices of the biological and behavioral sciences. Mechanism, on the other hand, does offer fruitful avenues for enrichment (Darden 2002; Bechtel and Abrahamsen 2005; Craver 2007, 66–70). William Bechtel and Adele Abrahamsen suggest that mechanistic explanations are generalizable not because the target systems are *identical* in the relevant respects, but because they are *similar*:

> The need to invoke similarity relations to generalize mechanistic explanations seems to be a limitation of the mechanistic account. But in fact it may be the mechanistic account that provides a better characterization of how explanations are generalized

> in many sciences. Laws are generalized by being universally quantified and their domain of applicability is specified by the conditions in their antecedents. On this account, no instance better exemplifies the law than any other. But in actual investigations of mechanisms, scientists often focus on a specific exemplar when first developing their accounts. (Bechtel and Abrahamsen 2005, 438)

The claim that generalization is based on similarity is less satisfying than the picture of subsumption under a covering law. Bechtel and Abrahamsen's discussion does little to constrain the practice of licit generalization, and their observation that scientists "seem to have an intuitive sense" of how to generalize is distinctly unsatisfying (438). However, given the lack of universal or exceptionless laws in the biological sciences, a more complicated conception of generalization is needed than that provided by the covering-law view.

The need to be more specific about "similarity"-based generalization is not a drawback of the mechanist framework, but a demand for further research by philosophers of science. The mechanist framework offers richer resources than functionalism for constrained similarity-comparisons. In comparing two mechanisms, one can appeal to similarities and differences between the sets of entities, of operations, or in their organization. For example, consider normalization models of neural activity (Carandini and Heeger 1994; Chirimuuta 2014), which describe the activation of neurons as a combination of linear excitation from a "driving input" and lateral inhibition according to a normalization equation. Heeger's normalization model was originally developed to describe neural activity in visual cortex (where the "driving input" is primarily from the lateral geniculate nucleus). Early data was collected from cat visual cortex and generalized to human visual cortex (Heeger 1992), but there is evidence that the normalization model generalizes to many other systems, including invertebrate olfaction, the retina, visual area MT, auditory cortex, and other cortical areas (Carandini and Heeger 2012). Thus, normalization models of neural activity follow the pattern mentioned by Bechtel and Abrahamsen whereby models are often developed for some exemplar case (in this case, feline visual cortex, specifically

area V1) and the potential for generalizing the model is tested in a variety of cases. In these cases, the model must be reinstantiated with different entities and organizations, e.g. neurons or groups of neurons in different anatomical regions, and the "driving input" and lateral components must be identified in the new context. It may also be the case that the neural mechanisms that give rise to normalization phenomena are distinct in different cases (Carandini and Heeger 2012, 58–60). There are still open questions about how many neural systems are well-described by normalization models, and in precisely what ways the entities and organizational structures of those systems are similar and different. Importantly, the result of such comparisons is not a judgment that mechanisms described by different models are simply the same or different, but that they are similar in certain respects and dissimilar in others. The subsumption[42] of several systems under the renormalization equation is the beginning of a story about how normalization models generalize, not a stable endpoint. Generalization of a model to a new system requires empirical examination of the mechanism of the new system, confirmation that the model applies insofar as it does, and discovery of the ways in which it must be revised to be extended. Frequently, a model may apply but with modifications (e.g. different values for the free parameters of the normalization equation), with the consequence that insights are gained both for the new and for the original target systems. For example, the two visual streams hypothesis (Milner and Goodale 2006) was developed for primate visual systems, primarily with data from humans and macaques, but comparisons of primates and other organisms such as frogs enrich the model (see e.g. Goodale and Humphrey 1998, 183–185) and provide a framework around which similar models can be developed for most vertebrates (Jeannerod and Jacob 2005, 301). This is mechanistic generalization by, if you like, functional similarity, but not functional identity in Putnam's sense.

It might be possible to provide similarity-based generalizations of functionally-individuated kinds, but such a strategy is not pursued by those who appeal to functionalism in

[42] "Subsumption" may be too strong and suggestive a word in the case of the normalization equation, for the equation contains several free parameters. Since new instantiations of the model may vary along those parameters, the normalization equation does not have the empirical strength of a paradigmatic law.

order to settle other questions in philosophy of cognitive science. For example, the strategy is not pursued by Adams and Aizawa, Rupert, or Sprevak in their criticism of HEC, who instead seek categorical descriptions of mental or cognitive processes. Clark and Chalmers appeal to a relatively abstract specification of memory to argue that Otto's notebook functions as a part of his memory. The RAA objections appeal to relatively detailed specifications of memory to argue that the notebook does not so function. An ecumenically-minded theorist might suggest that alternative specifications—with different degrees and kinds of abstractness—delineate various dimensions of similarity and difference between paradigmatic memory and Otto's notebook-augmented memory. However, if such a strategy were adopted by functionalists of the future, it would mean giving up the claim—essential to functionalism of the sort envisioned by Putnam and by contributors to the HEC dispute—that cognitive capacities can be ascribed on the basis of applying functional descriptions. Functionalism, in its long tenure as a central doctrine, was attractive precisely because it provided a simple and compelling picture of cognitive scientific discourse, about the way to specify kindhood, about the relation between abstraction and the practice of generalization, and about the structure of cognitive scientific models. New mechanism offers a less simple but more plausible picture of cognitive scientific discourse.

A.5 SUMMARY

My intention in this appendix was to show that the discussion of HEC reveals assumptions of functionalism that are inappropriate for thinking clearly about cognitive science. I claimed that often, as in the discussion of HEC, functionalism serves as a "central doctrine" in philosophy of cognitive science, a resource for appeal regarding criteria of ascription for cognitive kinds, for explanatory practice, and for generalization. I argued, repurposing arguments by Sprevak and Walter, that functional descriptions could not simultaneously serve as resources for appeal in more than one of those roles. I then described the way that appeals to functionalism as a central doctrine

rely on nomological conceptions of explanation, and argued that this reliance undercuts functionalism's potential to serve as an effective resource for appeal regarding any of its roles. This is because functionalism incorporates a problematically simplified picture of scientific practice. It is notable that whereas the mechanist model of explanation (and close relatives) have gained popularity in philosophical work about most of the biological sciences—including cellular biology and neuroscience—they have been the object of less attention in philosophy of cognitive science (but see Bechtel 2007; Weiskopf 2011; Stinson forthcoming). Perhaps this is a consequence of the fact that analytic philosophy of cognitive science came into its own some decades before the philosophy of biology or the philosophy of neuroscience, and has both a more prominent role in the theoretical foundations of its science and more "inertia" deriving from the extent of its literature. However, it is also possible that the stubborn popularity of functionalism has played a role. This is especially plausible given that, as I have claimed, functionalism is best suited to the covering-law models of explanation that compete with mechanism.

Finally, I described new mechanism and argued that it does not suffer from functionalism's drawbacks with respect to explanation. I also suggested, following a suggestion by Bechtel and Abrahamsen, that where the functionalist framework hides the complexity in cognitive scientists' practice of generalization, the mechanist literature provides a more fruitful framework for exploring that complexity. I have not argued that mechanism settles whether HEC is true or false. However, if disagreements about HEC are to be a battle for the soul of cognitive science, I contend that the proper battleground is over what kinds of mechanisms are cognitive ones, not over functionalist descriptions of mental states (cf. Walter 2010). Of course, I would suggest that the sensitive management hypothesis (Ch. 5) provides us with resources for thinking about this question.

The mechanist framework does not provide us with resources for determining the identity conditions of cognitive phenomena like belief and memory, as the functionalist framework does. However, cognitive scientists do not take conformity to their models as a criterion of exhibiting a phenomenon. For example, psychologists do not claim that exhibiting interference effects is a

necessary condition on memory. That a system does not exhibit interference effects implies that memory models that do exhibit such effects must be modified in order to be generalized to that target system, not that the target system lacks genuine memory. It is therefore peculiarly contentious for philosophers to appeal to these models in order to settle the identity conditions for cognitive phenomena under the guise of being good naturalists. The contentious nature of this form of argument is, no doubt, obscured by the common belief that functionalism is an orthodoxy of cognitive science. I have provided some reasons to believe that functionalism is not an articulation of the tacit commitments of cognitive scientists, and therefore propose that we look elsewhere, perhaps to mechanism, for articulate expressions of scientific commitments. I do not doubt that functional descriptions—that is, description in terms of relations to inputs, outputs, and other intermediate states—may have some legitimate uses in cognitive science. But the belief in the utility of such descriptions is not sufficient for maintaining an interest in functionalism. If these functional descriptions do not provide criteria of ascription, do not form the basis of explanations, and do not determine how explanations generalize, then we may keep our functional descriptions but must abandon functionalism.

BIBLIOGRAPHY

Adams, Fred, and Ken Aizawa. 2001. "The Bounds of Cognition." *Philosophical Psychology* 14: 43–64.

———. 2008. *The Bounds of Cognition*. Malden, MA: Blackwell.

———. 2010. "Defending the Bounds of Cognition." In Richard Menary (ed.), *The Extended Mind*. Cambridge, MA: Bradford, pp. 67–80.

Adams, Fred, and Rebecca Garrison. 2013. "The Mark of the Cognitive." *Minds & Machines* 23: 339–352.

Aizawa, Ken. forthcoming. "Cognition and Behavior." *Synthese*.

Aizawa, Ken, and Fred Adams. 2005. "Defending Non-Derived Content." *Philosophical Psychology* 18: 661–669.

Anderson, John R. 1993. *Rules of the Mind*. Hillsdale, NJ: Lawrence Erlbaum.

———. 2000. *Learning and Memory*. 2nd Ed. New York: Wiley.

———. 2010. *Cognitive Psychology and Its Implications*. Seventh Ed. New York: Worth.

Anderson, Michael L. 2014. *After Phrenology: Neural Reuse and the Interactive Brain*. Cambridge, MA: MIT Press.

Aristotle. "On the Soul." In Jonathan Barnes (ed.), *The Complete Works of Aristotle*. Vol. 1. Princeton: Princeton University Press, 1984, pp. 641–692.

Atkinson, Richard C., and Richard M. Shiffrin. 1968. "Human Memory: A Proposed System and Its Control Processes." In Kenneth Wartenbee Spence and Janet T. Spence (eds.), *The Psychology of Learning and Motivation: Advances in Research and Theory*. Vol. 2. New York: Academic Press, pp. 90–197.

———. 1971. "The Control of Short-Term Memory." *Scientific American* 225: 82–90.

Bacon, Francis. 1620. *Novum Organum*. New York: P.F. Collier, 1902.

Barton, Jason J.S., Daniel Z. Press, Julian P. Keenan, and Margaret O'Connor. 2002. "Lesions of the Fusiform Face Area Impair Perception of Facial Configuration in Prosopagnosia." *Neurology* 58: 71–78.

Barwise, Jon. 1981. "Scenes and Other Situations." *The Journal of Philosophy* 78: 369–397.

Barwise, Jon, and John Perry. 1983. *Situations and Attitudes*. Cambridge, MA: MIT Press.

Batitsky, Vadim. 1998. "A Formal Rebuttal of the Central Argument for Functionalism." *Erkenntnis* 49: 201–220.

Bechtel, William. 2007. *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. New York: Psychology Press.

———. 2008. "Mechanisms in Cognitive Psychology: What Are the Operations?" *Philosophy of Science* 75: 983–994.

Bechtel, William, and Adele Abrahamsen. 2005. "Explanation: A Mechanist Alternative." *Studies in History and Philosophy of Biological and Biomedical Sciences* 36: 421–441.

Bechtel, William, and Jennifer Mundale. 1999. "Multiple Realizability Revisited: Linking Cognitive and Neural States." *Philosophy of Science* 66: 175–207.

Block, Ned. 1980. "Troubles with Functionalism." In Ned Block (ed.), *Readings in Philosophy of Psychology*. Cambridge, MA: Harvard University Press, pp. 171–184.

———. 1981. "Psychologism and Behaviorism." *Philosophical Review* 90: 5–43.

———. 1995. "On a Confusion About the Function of Consciousness." *Behavioral and Brain Sciences* 18: 227–287.

Block, Ned, and Jerry A. Fodor. 1972. "What Psychological States Are Not." *Philosophical Review* 81: 159–181.

Bower, Gordon. 2000. "A Brief History of Memory Research." In Endel Tulving and Fergus I.M. Craik (eds.), *The Oxford Handbook of Memory*. New York: Oxford, pp. 3–32.

Boyd, Richard N. 1991. "Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds." *Philosophical Studies* 61: 127–148.

———. 1999. "Kinds, Complexity, and Multiple Realization." *Philosophical Studies* 95 (67–98).

Brandom, Robert B. 1994. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press.

———. 2000. *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, MA: Harvard University Press.

Brooks, Rodney. 1991. "Intelligence without Representation." *Artificial Intelligence* 47: 139–159.

Brooks, Rodney, Jonathan Connell, and Peter Ning. 1988. "Herbert: A Second Generation Mobile Robot." *A.I. Memos* 1016: 0–10.

Broome, John. 2007. "Does Rationality Consist in Responding Correctly to Reasons." *Journal of Moral Philosophy* 4: 349–374.

Buchsbaum, Daphna, Sophie Bridgers, Deena Skolnick Weisberg, and Alison Gopnik. 2012. "The Power of Possibility: Causal Learning, Counterfactual Reasoning, and Pretend Play." *Philosophical Transations of the Royal Society B* 367: 2202–2212.

Buckner, Cameron. 2015. "A Property Cluster Theory of Cognition." *Philosophical Psychology* 28: 307–336.

Buechner, Jeff. 2011. "Not Even Computing Machines Can Follow Rules: Kripke's Critique of Functionalism." In Alan Berger (ed.), *Saul Kripke*. Cambridge: Cambridge University Press, pp. 343–367.

Burge, Tyler. 2014. "Pereception: Where Mind Begins." Paper presented at the 40th Annual Meeting of the Society for Philosophy and Psychology, University of British Columbia, Vancouver, British Columbia, 18–21 June.

Burgess, Alexis, and David Plunkett. 2013a. "Conceptual Ethics I." *Philosophy Compass* 8: 1091–1101.

———. 2013b. "Conceptual Ethics Ii." *Philosophy Compass* 8: 1102–1110.

Calvo Garzón, Francisco. 2007. "The Quest for Cognition in Plant Neurobiology." *Plant Signaling & Behavior* 2: 208–211.

———. 2011. "Plants: Adaptive Behavior, Root-Brains, and Minimal Cognition." *Adaptive Behavior* 19: 155–171.

Carandini, Matteo, and David J. Heeger. 1994. "Summation and Division by Neurons in Primate Visual Cortex." *Science* 264: 1333–1336.

———. 2012. "Normalization as a Canonical Neural Computation." *Nature Reviews Neuroscience* 13: 51–62.

Carnap, Rudolf. 1947. *Meaning and Necessity*. Chicago: University of Chicago Press.

Chalmers, David. 1995. "Facing up to the Problem of Consciousness." *Journal of Consciousness Studies* 2: 200–219.

Chalmers, David J. 2011. "A Computational Foundation for the Study of Cognition." *Journal of Cognitive Science* 12: 323–357.

Chang, Hasok. 2008. *Inventing Temperature: Measurement and Scientific Progress*. New York: Oxford University Press.

Chemero, Anthony. 2009. *Radical Embodied Cognitive Science*. Cambridge, MA: MIT Press.

Chemero, Anthony, and Michael Silberstein. 2008. "After the Philosophy of Mind: Replacing Scholasticism with Science." *Journal of Philosophy* 75: 1–27.

Chirimuuta, Mazviita. 2014. "Minimal Models and Canonical Neural Computations: The Distinctness of Computational Explanation in Neuroscience." *Synthese* 191: 127–153.

Churchland, Paul M. 1981. "Eliminative Materialism and the Propositional Attitudes." *Philosophy of Science* 78: 67–90.

Clark, Andy. 1997. *Being There: Putting Brain, Body and World Together Again*. Cambridge, MA: MIT Press.

———. 2005. "Intrinsic Content, Active Memory, and the Extended Mind." *Analysis* 65: 1–11.

———. 2008. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: Oxford University Press.

———. 2010a. "Coupling, Constitution, and the Cognitive Kind: A Reply to Adams and Aizawa." In Richard Menary (ed.), *The Extended Mind*. Cambridge, MA: Bradford, pp. 81–100.

———. 2010b. "Memento's Revenge: The Extended Mind, Extended." In Richard Menary (ed.), *The Extended Mind*. Cambridge, MA: Bradford, pp. 43–66.

Clark, Andy, and David Chalmers. 1998. "The Extended Mind." *Analysis* 58: 7–19.

Clarke, Ellen. 2013. "The Multiple Realizability of Biological Individuals." *Journal of Philosophy* 110: 413–435.

Cleland, Carol E. 2012. "Life without Definitions." *Synthese* 185: 125–144.

Craver, Carl F. 2001. "Role Functions, Mechanisms, and Hierarchy." *Philosophy of Science* 68: 53–74.

———. 2007. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Oxford University Press.

Cummins, Robert. 1975. "Functional Analysis." *The Journal of Philosophy* 72: 741–765.

———. 2000. "'How Does It Work?' Vs. 'What Are the Laws?': Two Conceptions of Psychological Explanation." In Frank C. Keil and Robert A. Wilson (eds.), *Explanation and Cognition*. Cambridge, MA: MIT Press, pp. 117–144.

Cummins, Robert, and Pierre Poirier. 2004. "Representation and Indication." In Hugh Clapin, Phillip Staines and Peter Slezak (eds.), *Representation in Mind: New Approaches to Mental Representation*. Amsterdam: Elsevier, pp. 21–40.

Cushman, Fiery. 2013. "Action, Outcome, and Value: A Dual-System Framework for Morality." *Personality and Social Psychology Review* 17: 273–292.

Darden, Lindley. 2002. "Strategies for Discovering Mechanisms: Schema Instantiation, Modular Subassembly, Forward/Backward Chaining." *Philosophy of Science* 69: S354–S365.

Darwin, Charles. 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray. Reprint, Ed. Ernst Mayr (1964). A facsimile of the First Edition. Cambridge, MA: Harvard University Press.

Davidson, Donald. 1963. "Actions, Reasons, and Causes." *The Journal of Philosophy* 60: 685–700. Reprinted in Davidson, 2001, *Essays on Actions and Events* (Oxford: Oxford University Press), pp. 3–19.

Dawkins, Richard. 1976. *The Selfish Gene*. Oxford: Oxford University Press.

Dennett, Daniel C. 1984. *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge: Bradford.

———. 1987. *The Intentional Stance*. Cambridge, MA: MIT Press.

———. 1990. "The Myth of Original Intentionality." In K.A. Mohyeldin Said, W.H. Newton-Smith, R. Viale and K.V. Wilkes (eds.), *Modelling the Mind*. Oxford: Oxford University Press, pp. 43–62.

Devitt, Michael. 2012. "Semantic Epistemology: Response to Machery." *Theoria* 74: 229–233.

Dewey, John. 1958. *Experience and Nature*. Mineola, NY: Dover.

Dretske, Fred. 1986. "Misrepresentation." In Radu Bogdan (ed.), *Belief: Form, Content and Function*. New York: Oxford University Press, pp. 17–36.

Dretske, Fred I. 1977. "Laws of Nature." *Philosophy of Science* 44: 248–268.

Eliasmith, Chris. 2002. "The Myth of the Turing Machine: The Failure of Functionalism and Related Theses." *Journal of Experimental and Theoretical Artificial Intelligence* 14: 1–8.

Figdor, Carrie. 2014. "Verbs and Minds." In Mark Sprevak and Jesper Kallestrup (eds.), *New Waves in Philosophy of Mind*. New York: Palgrave Macmillan.

Flikschuh, Katrin. 2014. "The Idea of Philosophical Fieldwork: Global Justice, Moral Ignorance, and Intellectual Attitudes." *Journal of Political Philosophy* 22: 1–26.

Fodor, Jerry A. 1968. "The Appeal to Tacit Knowledge in Psychological Explanations." *The Journal of Philosophy* 65: 627–640.

———. 1975. *The Language of Thought*. New York: Crowell.

———. 1985. *Modularity of Mind*. Cambridge, MA: MIT Press.

———. 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT/Bradford.

Fodor, Jerry A., and Zenon W. Pylyshyn. 1988. "Connectionism and Cognitive Architecture: A Critical Analysis." *Cognition* 28: 3–71.

Gallagher, Shaun. 2005. *How the Body Shapes the Mind*. Oxford: Clarendon.

Gettier, Edmund. 1963. "Is Justified True Belief Knowledge?" *Analysis* 23: 121–123.

Gibbs, Raymond W., Jr. 2005. *Embodiment and Cognitive Science*. Cambridge: Cambridge University Press.

Gibson, James J. 1979. *The Ecological Approach to Visual Perception*. Hillsdale, NJ: Lawrence Erlbaum.

Gigerenzer, Gerd, Peter M. Todd, and the ABC Research Group. 1999. *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.

Godfrey-Smith, Peter. 2002. "Environmental Complexity and the Evolution of Cognition." In Robert J. Sternberg and James C. Kaufman (eds.), *The Evolution of Intelligence*. Mahwah, NJ: Lawrence Erlbaum, pp. 233–249.

———. 2009a. *Darwinian Populations and Natural Selection*. Oxford: Oxford University Press.

———. 2009b. "Triviality Arguments against Functionalism." *Philosophical Studies* 145: 273–295.

———. 2012a. "Darwinian Individuals." In Frédéric Bouchard and Philippe Huneman (eds.), *From Groups to Individuals: Perspectives on Biological Associations and Emerging Individuality*. Cambridge, MA: MIT Press.

———. 2012b. "Darwinism and Cultural Change." *Philosophical Transations of the Royal Society B* 367: 2160–2170.

Goldin-Meadow, Susan, and Susan M. Wagner. 2005. "How Our Hands Help Us Learn." *Trends in Cognitive Sciences* 9: 234–241.

Goodale, Melvyn A., and Keith G. Humphrey. 1998. "The Objects of Action and Perception." *Cognition* 67: 181–207.

Gopnik, Alison, and Henry M. Wellman. 2012. "Reconstructing Constructivism: Causal Models, Bayesian Learning Mechanisms, and the Theory Theory." *Psychological Bulletin* 138: 1085–1108.

Graham, George, and Terry Horgan. 1994. "Southern Fundamentalism and the End of Philosophy." *Philosophical Issues* 5: 219–247.

Green, Sara, Arnon Levy, and William Bechtel. 2014. "Design Sans Adaptation." *European Journal for Philosophy of Science*.

Greene, Joshua D., Leigh E. Nystrom, Andrew D. Engell, John M. Darley, and Jonathan D. Cohen. 2004. "The Neural Bases of Cognitive Conflict and Control in Moral Judgment." *Neuron* 44: 389–400.

Griffiths, Paul E., Edouard Machery, and Stefan Linquist. 2009. "The Vernacular Concept of Innateness." *Mind and Language* 24: 605–630.

Griffiths, Paul E., and Karola Stotz. 2006. "Genes in the Postgenomic Era." *Theoretical Medicine and Bioethics* 27: 499–521.

Hallin, Daniel C. 1986. *The Uncensored War: The Media and Vietnam*. New York: Oxford University Press.

Haslanger, Sally. 2000. "Gender and Race: (What) Are They? (What) Do We Want Them to Be?" *Noûs* 34: 31–55.

Haugeland, John. 1981a. "Representational Genera." In Stephen Stich and David E. Rumelhart (eds.), *Philosophy and Connectionist Theory*. Hillsdale, NJ: Lawrence Erlbaum.

———. 1981b. "Semantic Engines: An Introduction to Mind Design." In John Haugeland (ed.), *Mind Design: Philosophy, Psychology, Artificial Intelligence*. Cambridge, MA: MIT Press.

Haxby, James V., M. Ida Gobbini, Maura L. Furey, Alumit Ishai, Jennifer L. Schouten, and Pietro Pietrini. 2001. "Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex." *Science* 293: 2425–2430.

Heeger, David J. 1992. "Normalization of Cell Responses in the Cat Striate Cortex." *Visual Neuroscience* 9: 181–197.

Heil, Martin, and Juan Carlos Silva Bueno. 2007. "Within-Plant Signaling by Volatiles Leads to Induction and Priming of an Indirect Plant Defense in Nature." *Proceedings of the National Academy of Sciences* 104: 5467–5472.

Hempel, Carl G. 1965. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.

Hempel, Carl G., and Paul Oppenheim. 1948. "Studies in the Logic of Explanation." *Philosophy of Science* 15: 135–175.

Hurley, Susan L. 1998. *Consciousness in Action*. Cambridge, MA: Harvard University Press.

Hutchins, Edwin. 1995. *Cognition in the Wild*. Cambridge, MA: MIT Press.

Irvine, Elizabeth. 2012. *Consciousness as a Scientific Concept*. Dordrecht: Springer.

Jackson, Frank. 1998. *From Metaphysics to Ethics*. Oxford: Oxford University Press.

James, William. 1890. *The Principles of Psychology*. New York: Henry Holt and Company.

Jastrow, Joseph. 1899. "The Mind's Eye." *Popular Science Monthly* 54: 299–312.

Jeannerod, Marc. 1984. "The Timing of Natural Prehension Movements." *Journal of Motor Behavior* 16: 235–254.

Jeannerod, Marc, and Benjamin Biguer. 1982. "Visuomotor Mechanisms in Reaching within Extrapersonal Space." In David J. Ingle, Melvyn A. Goodale and Richard J.W. Mansfield (eds.), *Analysis of Visual Behavior*. Cambridge, MA: MIT Press, pp. 387–409.

Jeannerod, Marc, and Pierre Jacob. 2005. "Visual Cognition: A New Look at the Two-Visual Systems Model." *Neuropsychologia* 43: 301–312.

Johnson, Elizabeth K., and Peter W. Jusczyk. 2001. "Word Segmentation by 8-Month Olds: When Speech Cues Count More Than Statistics." *Journal of Memory and Language* 44: 548–567.

Kant, Immanuel. 1787. *Critique of Pure Reason*. Eds. Paul Guyer and Allen W. Wood. Cambridge: Cambridge University Press, 1998.

Kanwisher, Nancy. 2010. "Functional Specificity in the Human Brain: A Window into the Functional Architecture of the Mind." *Proceedings of the National Academy of Sciences* 107: 11,163–11,170.

Kaplan, David. 1989. "Demonstratives." In Joseph Almog, John Perry and Howard Wettstein (eds.), *Themes from Kaplan*. Oxford: Oxford University Press, pp. 481–563.

Keijzer, Fred. 2013. "The *Sphex* Story: How the Cognitive Sciences Kept Repeating an Old and Questionable Anecdote." *Philosophical Psychology* 26: 502–519.

Kelso, J.A. Scott. 1995. *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Cambridge, MA: MIT Press.

Kratzer, Angelika. 2016. "Situations in Natural Language Semantics." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Spring 2016 Ed. http://plato.stanford.edu/archives/spr2016/entries/situations-semantics/.

Laird, John E., Allen Newell, and Paul S. Rosenbloom. 1987. "SOAR: An Architecture for General Intelligence." *Artificial Intelligence* 33 (1–64).

Levin, Janet. 2013. "Functionalism." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Fall 2013 Ed. http://plato.stanford.edu/archives/fall2013/entries/functionalism/.

Levy, Arnon. 2013. "Three Kinds of New Mechanism." *Biology and Philosophy* 28: 99–114.

Levy, Arnon, and William Bechtel. 2013. "Abstraction and the Organization of Mechanisms." *Philosophy of Science* 80: 241–261.

Lewis, David K. 1972. "Psychophysical and Theoretical Identifications." *Australasian Journal of Philosophy* 50: 249–258.

———. 1973. *Counterfactuals*. Cambridge, MA: Harvard University Press.

———. 1983. "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61: 343–377.

Lewontin, Richard C. 1970. "The Units of Selection." *Annual Review of Ecology and Systematics* 1: 1–18.

Li, Bihui, and J.S. Hancox. MS. "Defining the Terms of Our Struggle: On the Role of the Concept of Woman in Political Action." Draft of 13 January, 2016.

Lloyd, Elisabeth. 2012. "Units and Levels of Selection." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Winter 2012 Ed. http://plato.stanford.edu/archives/win2012/entries/selection-units/.

Lyon, Pamela. 2006. "The Biogenic Approach to Cognition." *Cognitive Processing* 7: 11–29.

MacFarlane, John. 2014. *Assessment Sensitivity: Relative Truth and Its Applications*. Oxford: Clarendon.

Machamer, Peter K., Lindley Darden, and Carl F. Craver. 2000. "Thinking About Mechanisms." *Philosophy of Science* 67: 1–25.

Machery, Edouard. 2004. "Semantics, Cross-Cultural Style." *Cognition* 92: B1–B12.

———. 2009. *Doing without Concepts*. Oxford: Oxford University Press.

———. 2012. "Why I Stopped Worrying About the Definition of Life... And Why You Should as Well." *Synthese* 185: 145–164.

———. MS. *Philosophy within Its Proper Bounds*. Book manuscript. Draft of 1 February, 2016.

Mackie, J.L. 1977. *Ethics: Inventing Right and Wrong*. Harmondsworth, UK: Penguin.

Maley, Corey, and Gualtiero Piccinini. MS. "Get the Latest Upgrade: Functionalism 6.3.1."

Marr, David. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Cambridge, MA: MIT Press, 2010.

Maturana, Humberto Romesin, and Francisco Varela. 1980. *Autopoiesis and Cognition: The Realization of the Living*. Boston: D. Reidel.

McGurk, Harry, and John MacDonald. 1976. "Hearing Lips and Seeing Voices." *Nature* 264: 756–748.

Millikan, Ruth Garrett. 1984. *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge, MA: MIT Press.

———. 1995. "Pushmi-Pullyu Representations." *Philosophical Perspectives* 9: 185–200.

Milner, A. David, and Melvyn A. Goodale. 2006. *The Visual Brain in Action*. 2nd Ed. Oxford: Oxford University Press.

Modlmeier, Andreas P., and Susanne Foitzik. 2011. "Productivity Increases with Variation in Aggression among Group Members in *Temnothorax* Ants." *Behavioral Ecology* 22: 1026–1032.

Munz, Tania. 2005. "The Bee Battles: Karl Von Frisch, Adrian Wenner and the Honey Bee Dance Language Controversy." *Journal of the History of Biology* 38: 535–570.

Murphy, Dominic. 2006. *Psychiatry in the Scientific Image*. Cambridge, MA: MIT Press.

Myers, David G. 2004. *Psychology*. Seventh Ed. New York: Worth.

Newell, Allen, and Herbert Simon. 1976. "Computer Science as Empirical Inquiry: Symbols and Search." *Communications of the Association for Computing Machinery* 19: 113–126.

Noë, Alva. 2006. *Action in Perception*. Cambridge, MA: MIT Press.

Nozick, Robert. 1981. *Philosophical Explanations*. Cambridge, MA: Harvard University Press.

O'Keefe, John, and John Dostrovsky. 1971. "The Hippocampus as a Spatial Map: Preliminary Evidence from Unit Activity in the Freely-Moving Rat." *Brain Research* 34: 171–175.

Ohlsson, Stellan. 1993. "Abstract Schemas." *Educational Psychologist* 28: 51–66.

Orlandi, Nico. 2014. *The Innocent Eye: Why Vision Is Not a Cognitive Process*. New York: Oxford University Press.

Parker, Geoff A., and John Maynard Smith. 1990. "Optimality Theory in Evolutionary Biology." *Nature* 348: 27–33.

Piccinini, Gualtiero. 2015. *Physical Computation: A Mechanist Account*. Oxford: Oxford University Press.

Place, U.T. 1956. "Is Consciousness a Brain Process?" *British Journal of Psychology* 47: 44–50.

Prinz, Jesse. 2004. *Gut Reactions: A Perceptual Theory of Emotions*. Oxford: Oxford University Press.

Putnam, Hilary. 1967a. "The Mental Life of Some Machines." In Hector-Neri Castañeda (ed.), *Intentionality, Minds, and Perception*. Detroit: Wayne State University Press. Reprinted in Putnam, 1975, *Mind, Language and Reality* (Cambridge: Cambridge University Press), pp. 408–428.

———. 1967b. "Psychological Predicates." In William H. Capitan and Daniel Davy Merrill (eds.), *Art, Mind, and Religion*. Pittsburgh: University of Pittsburgh Press, pp. 37–48. Reprinted in Putnam, 1975, *Mind, Language and Reality* (Cambridge: Cambridge University Press), pp. 429–440, as "The Nature of Mental States".

———. 1988. *Representation and Reality*. Cambridge, MA: MIT Press.

Quine, Willard Van Orman. 1951. "Two Dogmas of Empiricism." *Philosophical Review* 60: 20–43. Reprinted in *From a Logical Point of View* (Cambridge, MA: Harvard University Press, 1953; 2nd, rev. ed. 1961), pp. 20–46.

———. 1960. *Word and Object*. Cambridge, MA: MIT Press.

Ramsey, William M. 1992. "Prototypes and Conceptual Analysis." *Topoi* 11: 59–70.

———. 2007. *Representation Reconsidered*. Cambridge: Cambridge University Press.

———. forthcoming. "Must Cognition Be Representational?" *Synthese*.

Robbins, Philip, and Murat Aydede, eds. 2008. *The Cambridge Handbook of Situated Cognition*. Cambridge: Cambridge University Press.

Roberts, Robert C. 1988. "What Emotion Is: A Sketch." *Philosophical Review* 97: 183–209.

Rosch, Eleanor, and Carolyn B. Mervis. 1975. "Family Resemblance: Studies in the Internal Structure of Categories." *Cognitive Psychology* 7: 573–605.

Ross, Lauren N. 2015. "Dynamical Models and Explanation in Neuroscience." *Philosophy of Science* 82: 32–54.

Rowlands, Mark. 2009. "Extended Cognition and the Mark of the Cognitive." *Philosophical Psychology* 22: 1–19.

———. 2010. *The New Science of the Mind: From Extended Mind to Embodied Phenomenology*. Cambridge, MA: MIT Press.

Rumelhart, David E. 1989. "The Architecture of Mind: A Connectionist Approach." In Michael I. Posner (ed.), *Foundations of Cognitive Science*. Cambridge, MA: MIT Press, pp. 133–159.

Rupert, Robert. 2004. "Challenges to the Hypothesis of Extended Cognition." *Journal of Philosophy* 51: 389–428.

———. 2009. *Cognitive Systems and the Extended Mind*. Oxford: Oxford University Press.

———. 2013. "Memory, Natural Kinds, and Cognitive Extension; or, Martians Don't Remember, and Cognitive Science Is Not About Cognition." *Review of Philosophy and Psychology* 4: 25–47.

Saffran, Jenny R., Richard N. Aslin, and Elissa L. Newport. 1996. "Statistical Learning by 8-Month-Old Infants." *Science* 274: 1926–1928.

Salmon, Wesley C. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

———. 1989. *Four Decades of Scientific Explanation*. Minneapolis: University of Minnesota Press.

Schank, Roger. 1975. "Using Knowledge to Understand." In Bonnie L. Nash-Webber and Roger C. Schank (eds.), *Proceedings of the 1975 Workshop on Theoretical Issues in Natural Language Processing*. Stroudsburg, PA: Association for Computational Linguistics, pp. 117–121.

Searle, John R. 1980. "Minds, Brains and Programs." *Behavioral and Brain Sciences* 3: 417–457.

———. 1990. "Is the Brain a Digital Computer?" *Proceedings and Addresses of the American Philosophical Association* 64: 21–37.

Sellars, Wilfrid. 1962. "Philosophy and the Scientific Image of Man." In Robert G. Colodny (ed.), *Frontiers of Science and Philosophy*. Pittsburgh: Pittsburgh University Press, pp. 35–78. Reprinted in Kevin Scharp and Robert B. Brandom (eds.), *In the Space of Reasons: Selected Essays of Wilfrid Sellars* (Cambridge, MA: Harvard University Press, 2007), pp. 369–408.

———. 1981. "Mental Events." *Philosophical Studies* 39: 325–345.

Shagrir, Oron. 2005. "The Rise and Fall of Computational Functionalism." In Yemima Ben-Menahem (ed.), *Contemporary Philosophy in Focus: Hilary Putnam*. Cambridge: Cambridge University Press, pp. 220–250.

Shannon, Claude. 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27: 379–423, 623–656.

Shapiro, Lawrence A. 2013. "Dynamics and Cognition." *Minds & Machines* 23: 353–375.

Shea, Nicholas. 2013. "Naturalising Representational Content." *Philosophy Compass* 8: 496–509.

———. MS. "Functionalist Interrelations between (Extended) Cognitive States."

Smart, J.J.C. 1959. "Sensations and Brain Processes." *Philosophical Review* 68: 141–156.

Sober, Elliott. 1985. "Panglossian Functionalism and the Philosophy of Mind." *Synthese* 64: 165–193.

Sprevak, Mark. 2009. "Extended Cognition and Functionalism." *The Journal of Philosophy* 106: 503–527.

Stewart, John. 1996. "Cognition = Life: Implications for Higher-Level Cognition." *Behavioural Processes* 35: 311–326.

Stich, Stephen. 1983. *From Folk Psychology to Cognitive Science: The Case against Belief*. Cambridge, MA: Bradford.

Stinson, Catherine. forthcoming. "Mechanisms in Psychology: Ripping Nature at Its Seams." *Synthese*: 1–30.

Sullivan, Jacqueline A. 2009. "On the Multiplicity of Experimental Protocols." *Synthese* 167: 511–539.

Thelen, Esther, and Linda B. Smith. 1994. *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press.

Thompson, Evan. 2010. *Mind in Life: Biology, Phenomenology, and the Sciences of the Mind*. Cambridge, MA: Harvard University Press.

Thompson, Michael. 2008. *Life and Action: Elementary Structures of Practice and Practical Thought*. Cambridge, MA: Harvard University Press.

Trewavas, Anthony. 2003. "Aspects of Plant Intelligence." *Annals of Botany* 92 (1–20).

Tulving, Endel. 1972. "Episodic and Semantic Memory." In Endel Tulving and Wayne Donaldson (eds.), *Organization of Memory*. New York: Academic Press, pp. 381–403.

Turing, Alan M. 1950. "Computing Machinery and Intelligence." *Mind* 59: 433–460.

Uexküll, Jakob von. 1957. "A Stroll through the Worlds of Animals and Men: A Picture Book of Invisible Worlds." In Claire H. Schiller (ed.), *Instinctive Behavior: The Development of a Modern Concept*. New York: International Universities Prss, pp. 5–80.

van Duijn, Marc, Fred Keijzer, and Daan Franken. 2006. "Principles of Minimal Cognition: Casting Cognition as Sensorimotor Coordination." *Adaptive Behavior* 14: 157–170.

van Fraassen, Bas C. 1980. *The Scientific Image*. Oxford: Clarendon.

van Gelder, Tim. 1995. "What Might Cognition Be, If Not Computation?" *Journal of Philosophy* 92: 345–381.

———. 1998. "The Dynamical Hypothesis in Cognitive Science." *Behavioral and Brain Sciences* 21: 615–665.

Varela, Francisco, Evan Thompson, and Eleanor Rosch. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

von Helmholtz, Hermann. 1867. *Handbuch Der Physiologischen Optik*. Leipzig: Voss.

Walter, Sven. 2010. "Cognitive Extension: The Parity Argument, Functionalism, and the Mark of the Cognitive." *Synthese* 177: 285–300.

Webb, Barbara. 1994. "Robotic Experiments in Cricket Phonotaxis." In David Cliff, Phil Husbands, Jean-Arcady Meyer and Stewart W. Wilson (eds.), *From Animals to Animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behavior*. Cambridge, MA: MIT Press, pp. 45–54.

Weisberg, Michael. 2007. "Who Is a Modeler?" *British Journal for the Philosophy of Science* 58: 207–233.

Weiskopf, Daniel A. 2011. "Models and Mechanisms in Psychological Explanation." *Synthese* 183: 313–338.

———. 2012. "The Human Stain: Concepts, Anthropic Kinds, and Realism." Paper presented at the 38th Annual Meeting of the Society for Philosophy and Psychology, the University of Colorado at Boulder, Boulder, Colorado, June 21–23, 2012.

Wilson, Mark. 1982. "Predicate Meets Property." *Philosophical Review* 91: 594–589.

———. 2006. *Wandering Significance: An Essay on Conceptual Behavior*. Oxford: Clarendon.

———. MS. *Physics Avoidance and Other Essays*.

Wilson, Robert A. 2010. "Review of Robert D. Rupert's Cognitive Systems and the Extended Mind." *Notre Dame Philosophical Reviews* March 7, 2010: https://ndpr.nd.edu/news/24309-cognitive-systems-and-the-extended-mind/.

Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. Trans. G.E.M. Anscombe. Eds. G.E.M. Anscombe and Rush Rhees. 3rd Ed. Oxford: Blackwell, 2001.

Woodward, James. 2000. "Explanation and Invariance in the Special Sciences." *British Journal for the Philosophy of Science* 51: 197–254.

———. 2001. “Law and Explanation in Biology: Invariance Is the Kind of Stability That Matters.” *Philosophy of Science* 68: 1–20.

———. 2012. “Interactions between Philosophical Theories and Psychological Research.” *Philosophy of Science* 79: 961–972.

Yoshikawa Eiji. 1971. *Musashi*. Trans. Charles S. Terry. Tōkyō: Kodansha, 1995.