UC San Diego
UC San Diego Electronic Theses and Dissertations

Title
Perceptual Context and the Nature of Neural Function

Permalink
https://escholarship.org/uc/item/2x01v8k1

Author
Burnston, Daniel Charles

Publication Date
2015

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Perceptual Context and the Nature of Neural Function

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Philosophy and Cognitive Sciences

by

Daniel Charles Burnston

Committee in charge:

Professor William Bechtel, Chair
Professor Thomas Albright
Professor Jonathan Cohen
Professor Rick Grush
Professor John Serences

2015

Copyright

Daniel Charles Burnston, 2015

All rights reserved.

The Dissertation of Daniel Charles Burnston is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2015

DEDICATION

I would like to thank my committee for a series of extremely helpful and influential discussions during the development of this project, and William Bechtel for his constant guidance and thorough comments on previous drafts.

I would like to thank my family for their continuous and unwavering support during my time in graduate school.

TABLE OF CONTENTS

LIST OF FIGURES

VITA

2007	Bachelor of Arts, Rutgers University
2009	Master of Arts, Georgia State University
2015	Doctor of Philosophy, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Perceptual Context and the Nature of Neural Function

by

Daniel Charles Burnston

Doctor of Philosophy in Philosophy and Cognitive Sciences

University of California, San Diego, 2015

Professor William Bechtel, Chair

Functional localization is a primary goal of neuroscience, but there is a long history of debate about whether it is possible. That is, can one decompose the brain into parts, whose concerted functioning explains psychological states and behavior? These debates continue to the present day. In my dissertation, I criticize "absolutism,"

a meta-theoretic assumption about functional explanation that has shaped dialectic throughout the history of debates about localization.

Absolutism is the view that each neural area performs a univocal function, and performs that function regardless of the contexts in which it operates. Functional accounts in neuroscience have traditionally been absolutist. This was true of phrenology's attempts to isolate personality traits in parts of the brain, and remains true in current discussions about whether, for instance, a particular area of the brain is the "face-recognition" area. More deeply, absolutism has been seen as virtually definitive of localization—those denying absolutist function ascriptions have tended to deny localization altogether.

I argue, based on a detailed case study of perceptual area MT, that absolutism should be rejected. MT is traditionally construed as the "motion detector" of the visual system, but recent evidence strongly suggests that MT performs different functions—including, most compellingly representing depth—depending on the perceptual information available. I suggest that this evidence is best read as showing that MT is *context-sensitive* in its functioning. That is, it performs different functions depending on the context. Context-sensitivity is often considered to be incompatible with giving powerful theories, and several theorists, motivated by this concern, have attempted to save absolutism by modifying the kinds of functions posited. I argue that these attempts fail, and that, appropriately understood, contextualist theories can be seen as providing functional explanations that meet the epistemic goals of functional decomposition.

Chapter 1. Functional Decomposition in Neuroscience

1.1. Introduction.

Functional decomposition is a primary goal in the life sciences. According to a variety of influential views (Bechtel & Richardson, 1993; Cummins, 1975; Simon, 1962), a fundamental way of understanding a biological system is to divide it up into parts, describe what the parts do, and use those descriptions to understand the functioning of the whole system. This project has both metaphysical and epistemic aspects. On the one hand, its success depends on there being real divisions in the world, which are captured by our functional explanations. On the other hand, the functional divisions we posit must be epistemically useful—they must meet desiderata on good theories. I will discuss functional decomposition in a close analysis of a particularly vexed case, that of functional localization in the brain. For much of its history, neuroscience has been concerned with localization as a means of doing functional decomposition. One of its primary goals is to describe the parts of the brain in functional terms, characterize their interactions, and explain mental phenomena in virtue of those descriptions. However, for as long as the goal of functional localization has been pursued, there has been debate about whether it is in fact possible. These worries, as I will show, have focused on both the metaphysical and epistemic sides of the decomposition question. Opponents of localization have doubted, first, whether the brain does in fact contain meaningful functional divisions, and therefore have doubted whether the project of localization in neuroscience has genuine epistemic value.

I will assume that functional decomposition in general is a desirable epistemic goal, in the following minimal sense: *if* there are meaningful functional differences between parts of a biological system, we are epistemically better off exploring them, and theorizing about them, than we are in not doing so. This of course leaves open the possibility that opponents of localization are right. If there are no functional divisions, then we are not in fact better off in looking for them. It also leaves open the question of what the right kinds of theories are for doing functional localization. This is the primary question that I will address.

Debates about localization in neuroscience, both historical and in the present day, have traded on a meta-theoretic assumption about what successful functional decomposition amounts to. I call this assumption "absolutism."

- *Absolutism*: for any neural area A, there is some univocal description D, such that D explains A's activity in any context in which A functions.

Absolutism is tied to a certain view of what explanation consists in, which might be called atomism or essentialism. The idea is that for each (in this case) functional kind, what a successful explanation *is* is an explanation that gives a complete, simple account of that kind. In neuroscience, this means giving a single function ascription for each functional part of the brain.

I will argue that absolutism is a flawed view of functional explanation, and should be abandoned. In its place, I argue for its antithesis, a *contextualist* approach to localization. Absolutism is an inherently anti-contextualist position. Due to the essentialism inherent in absolutism, coming up with an explanation of a functional kind or property just consists in describing the one thing it does essentially. As such, that description will not change depending on the context in which the system is being studied. Contextualism argues, alternatively, that the functional properties of a part of the brain are in fact variable depending on contexts external to that part. Put simply, a part of the brain can do different things in different contexts. Contextualism sounds rather anodyne put so simply, but the historical prevalence of absolutism bespeaks a deep concern that context and localization are incompatible. I will argue that this worry is misplaced, and that contextualism is the appropriate meta-theoretic approach to neural function. Contextualism has been voiced, in very peremptory form, by theorists both in neuroscience (McIntosh, 2004) and philosophy (Klein, 2012), and I will discuss this development below. No one has yet, however, given a thorough exposition of the contextualist view and how it grounds decomposition. That is my purpose here.

My burden will be to show, first, the inadequacy of absolutism. Absolutism, in the case study I undertake, fails to be an empirically adequate description of function. That is, the brain does not appear to be the kind of system amenable to absolutist analysis (at least if the results from the case study generalize). My second burden will be to show that this failure is *not* because the brain has no meaningful functional divisions, but instead because absolutism is itself a flawed meta-theoretic assumption. Doing so will involve arguing that the contextualist alternative is itself epistemically desirable—that we should be satisfied with a contextualist approach to localization. Absolutism, especially in its modern form, is often motivated by views about epistemic adequacy. It is assumed that theories that are more general, more universally predictive, and more systematic are better, and that embracing context into functional explanation undermines these theoretical aims to a significant enough degree to make localization unfeasible or unfulfilling. I will argue that appropriately construed, contextualist theories can produce good functional theories. Contextualism does not posit universal or essential function ascriptions, but it *does*, on my view, provide powerful theories that both generalize across instances and support systematic science.

Contextualism takes a different view of both the metaphysical and epistemic aspects of functional decomposition. An absolutist assumes a certain form of stable organization in the world—any functional properties are atomistic, and if there are no atomistic properties, there is therefore no functional differentiation (e.g., in distinct parts of the brain). Wimsatt (2007), however (amongst others), has proposed an alternative view of organization, theory, and their relationship. He argues that the world is not organized according to univocal, constant divisions, such that we can give simple theories that capture constant aspects of organization. His view is partially based on the notion that the behavior of parts of the world (including biological systems) is heavily dependent on conditions outside of the system of study. But if the world is in fact organized in this way, then it is unlikely that we will find theories that apply universally. Wimsatt argues for a heuristic approach to theories, on which theories, and the types of reasoning that produce them, are not universally applicable, but instead have defined limits of applicability. I will draw on this perspective in several instances throughout, as a way of showing that abandoning absolutism does not produce useless theories of localization, but instead ones whose theoretical value is in part based on their heuristic roles. In the next section, I characterize some key notions in more detail, and explain how these notions are addressed by my primary case study—the functional decomposition of the visual system.

1.2. Functional Localization as a Goal in Neuroscience.

I will adopt the following general characterization of the notion of 'function':

- *Function of X*: what X contributes to the behavior of the system of which it is a part.

This description of function parallels that provided by Cummins (1975). On "Cummins-style" functional decomposition, attributing functions is inherently a question of *using* a function ascription for a particular part in order to explain something—the broader behavior of the system of interest. A few points to note about this characterization, and the others that follow in this chapter. First, it is intended as a characterization, rather than a definition. I am interested, primarily, not in the appropriate *analysis* of the concept of a function as such, but instead in explanation in neuroscience. That is, I am interested in what kinds of function ascriptions can localize functions to parts of the brain in a way that meets the explanatory aims and norms of neuroscience. Hence, I will sidestep much of the philosophical literature which debates the appropriate analysis of the concept of 'function'.[1]

---

[1] Most philosophical discussion about function has centered on the role that evolution has in specifying function ascriptions (Ariew, Cummins, & Perlman, 2002). Evolutionary theories of function face a range of challenges about how to specify functions. Often, the debate is about whether they can specify *univocal* functions for biological traits. Unfortunately, since I will be arguing against univocal

Second, the characterization is admittedly and intentionally vague. The reason for this is that, since the explanatory purposes for which particular functions are employed are important, the notion of a “contribution”—and the type of functions that are ascribed—must be filled in with details from the particular scientific projects in which functions are being posited. The functions ascribed must be appropriate for the physical systems being studied, and it is up for grabs both how to best characterize functions for a specific part of a specific system *and* how to describe the overall system. For instance, in this project I will be looking closely at visual neuroscience. The system being studied is the *visual system* of the brain. As such, the kinds of functions posited for parts of the visual system are those that best help describe how its parts contribute to overall visual function. I will largely assume the standard notion that the visual cortex is involved in visual perception (see Bechtel, 2008 for the scientific developments leading to this view, and chapter 5 for further discussion). The focus of my discussion will then be on what *kinds* of function ascriptions are appropriate for functional localization. Precisely the debate I am concerned with is how best to fill in the characterization of the types of contributions made by particular

---

functional ascriptions, these debates have relatively little relevance. Moreover, they are not specifically focused on how function ascriptions are developed in such fields as systems neuroscience, which are not primarily concerned with evolutionary history. Cummins’ (1975) theory of functions is more amenable to the kinds of projects I am discussing. Cummins proposes that a functional description of system in psychology consists of a box-and-arrow description of the functional roles played by different parts or processes within a system. Moreover, he opens the possibility that the form of these explanations can be sensitive to “analytical context” (p. 762). However, he does discuss this role for context in detail. This is the kind of question I will be taking up. “New mechanists” in philosophy of science such as Craver (2007) note that Cummins’ view of functions must be amended to include a view of causal interaction in order to provide a full mechanistic understanding of the brain. The relationships between my analysis of the MFH view and mechanistic views of explanation (Bechtel & Abrahamsen, 2005; Machamer, Darden, & Craver, 2000a) are too complex to delve into in detail. I will draw on these views as appropriate, but addressing the ultimate upshot of contextualist views of function for mechanistic explanation will need to be put off until a future time.

parts or areas of the brain to overall brain functioning. I will consider several different versions of absolutism before proposing a contextualist alternative. I will focus on explaining visual function, and specifically the contribution of MT to visual function, in the hopes that the discussion in regards to this specific brain area will prove constructive for debates regarding other areas, participating in other (i.e., non-visual) aspects of brain functioning (see section 5.7).

Absolutism is an atomistic claim about function, namely that all of the functionally defined areas of the brain should have one, univocal description that defines their function. That is, each area makes *one* contribution to overall brain function, and makes that same contribution in all contexts in which it functions. This supports the fact, of course, that there are multiple ways of describing and grouping areas. Several distinct areas are "visual" areas, for instance, but on absolutism these are further individuated and specified by playing precise, unique roles within perception.[2]

Absolutism is implicitly at work in many historical debates about neural function. The idea is that if a univocal description that distinguishes each area from all others can be found, then localization of function is possible. If not, then localization is impossible. If localization is impossible, then one is a "holist" about

---

[2] What makes a part of the brain a distinct area is another question that often requires a fair amount of scientific practice to answer. In the case of the visual system, areas are individuated, normally, by a combination of morphological and functional properties. Area MT was initially individuated anatomically—as having particular patterns of neural connections to other parts, and being morphologically distinct from other areas. I will discuss the initial reasons for demarcating MT in the next chapter. Then, in both chapters 3 and 5 I will defend the idea that 'MT' is still a useful functional category despite evidence supporting contextual variation in its function. That is, the evidence should not be taken as reason to abandon the project of finding the function of area MT, in favor of attempting to describe the function of some other anatomical division of the brain.

brain function. Holists posit that each part of the brain is “equipotential”—that all parts of the brain contribute equally to any given mental process. As I will show below, a variety of historical figures have argued from the purported failure of absolutist localization to holism, and this tendency continues to the present day.

It should be made clear that both absolutist localization and holism are extreme positions, and that not all theorists or practicing neuroscientists can be placed clearly in one camp or the other. Indeed, both localizationist and holist views tend to come qualified in certain ways (I will discuss some examples below). This does not mean that absolutism has not been, historically, the guiding principle for functional localization. The case study from perceptual neuroscience that I will pursue in detail shows absolutism at work in the modern analysis of a specific neural system, and this analysis is often taken as a textbook case. The long-standing dialectic between localizationists and holists, on my view, has made relatively little progress, and one reason for this is the lack of a well-articulated middle ground between absolutist localization and holism. I believe, however, that neuroscientists in several subfields have begun to look for such a middle ground, and that contextualism provides the appropriate way of conceptualizing it.

Traditionally, absolutist function ascriptions have been based on one of two conceptions of localization, which are closely related. The first is the idea that each area should be responsible for one particular kind of task. For instance, Broca’s area is traditionally construed as the language-production area, whereas Wernicke’s is traditionally the language-comprehension area (but see Bergeron, 2007 for an argument that this was not Wernicke's view). A famous, more modern example is that of the “Fusiform Face Area,” posited to be functionally specified for the recognition of faces (Kanwisher, McDermott, & Chun, 1997). Absolutist localization predicts that each area should be responsible for one task. The second conception is based on what causes an area to become active, the idea being that if A’s function is to process information of type X, then only information regarding X should activate A. A’s function is thus specified by discovering a privileged informational relationship between A’s activity and some aspect of the world, and these relations are investigated both by imaging and by electrophysiological recordings.

The task and information conceptions are closely related, and are often pursued together in investigation. If A is specified by task as performing a language comprehension task, then it should be activated primarily by perception of lexical stimuli. If A is a perceptual area, then it should contribute to perceptual judgment tasks involving the information that it carries. I will call these combined conceptions “traditional absolutism” or TA. It is TA that has largely shaped traditional discussion of function. However, the emergence of new versions of absolutism, discussed in chapter 4, shows how deeply entrenched absolutism is as a general conception of what functional localization amounts to. I will argue that neither TA nor the new forms of absolutism that have recently been developed are successful approaches to localization.

My strategy will be to analyze one of the most highly cited success cases of TA-style function attributions—the areas of the visual system. Visual neuroscience’s particular implementation of TA, the “modular functional hierarchy” (MFH) theory of visual system function, is one of the most influential explanatory frameworks in the history of neuroscience, and is often cited in textbooks and other fields as an exemplar success case for functional explanation (Bechtel, 2001). The MFH theory posits that each area in the visual system performs the function of representing a specific type of perceptual information, a particular type of *feature*. That is, each area contributes a specific type of feature representation to overall visual processing, and eventually to perception. Features are macroscopic properties of objects. Objects exhibit features when they instantiate particular properties, perhaps transiently. As with the characterization of function above, what is an explanatorily relevant feature depends on the explanatory purposes for which one is positing features. In studying the visual system, this has had a very interesting consequence. The particular features that have historically been used to describe the functions of parts of the visual system correspond strongly with the intuitive perceptual properties that we attribute to objects—attributives like shape, color, and motion. The absolutist description of the visual system is based on the idea that these different features have dedicated areas in charge of representing them. As I will discuss in the next chapter, the story is more complex than this initial gloss. On the standard absolutist reading of the visual system some parts of the visual cortex are in charge of representing very simple features—such as edges, wavelength, and displacement—which then contribute to later processes that represent the full-blooded features of color, motion, etc. Then these feature representations are later used for the representation of objects of specific types.

On the MFH version of absolutism, the functioning of the whole visual system, and eventually perception, are to be explained via relations between distinct representations occurring in different, functionally specified areas. Despite the theory's widespread acceptance, evidence has begun to emerge that areas of the visual system are not adequately described as representing one and only one kind of information. Recognition is growing that TA-type function ascriptions may not be adequate even for their most supposedly successful cases. Specifically, I analyze the history of investigation into visual area MT. MT is the traditional "linear motion" detector, generally construed as specifically representing patterns of motion in the visual field. Continuing investigations in the last 20 years, however, have slowly uncovered more and more sophisticated MT responses to other perceptual features, including color and, most impressively, depth. Importantly, uncovering these responses has required perceptual neuroscientists to explicitly and consciously explore changes in perceptual context. I argue that these results put sufficient strain on the standard, motion-specific TA interpretation to merit searching for a new way of describing MT function. In doing so, I am following the field—it is now widely accepted that MT has a more diverse function than solely representing linear motion. The challenge is to describe the change in function ascription in a way that accounts for the variation in MT responses without inviting the swing towards equipotentiality and holism. I will argue that the best way to do so is to embrace an explicitly contextualist perspective on the notion of neural function.

I characterize an individual 'context' as follows:

- Context: an instantiation of a particular set of explanatorily relevant properties.

A context is, most fundamentally, a situation. Contextualism embraces the idea that the broader situation in which parts of the brain are functioning matters for the particular function ascriptions that we make. Thus, our theories/explanations of functions in particular parts of the brain should be structured in such a way that contexts are part of the function ascription. I will give a particular theory of function ascriptions in chapter 5 that is designed to capture this view. Unsurprisingly, at this point, I take the foregoing as a peremptory characterization to be filled in in particular cases. In the study of the visual system, for instance, contextualism and absolutism deeply differ in how they view intuitive perceptual features or attributives. While absolutism takes them as psychological primitives to be localized in the brain, contextualism views them as aspects or parameters of perceptual contexts, which vary in their values and in their relationships to one another, thus comprising particular contexts.

Contextualism also admits that given parts of the brain might represent different types of information in different contexts. Much of the development of functional understanding of MT, or so I will argue, has been predicated upon finding new ways to manipulate, modify, and recombine particular perceptual features, such that in distinct circumstances MT makes different contributions to visual function. Perception researchers are often interested in *types* of contexts. Some types of contexts that I will discuss include contexts in which motion in the scene is segmented by color, in which depth is present independently of motion, in which tilt is defined by relative binocular disparity, or in which the same feature is defined by relative velocity. Perceptual scientists study contexts by manipulating the stimulus they present to the visual system in particular investigations. Much of the work of perceptual science involves continually constructing and manipulating new types of stimuli, which model distinct types of context, looking for differences in context that make a difference in (e.g.) visual phenomena. So, a stimulus that only contains motion information models contexts in which only motion is occurring in the visual scene. When more aspects of the stimulus are added, they model contexts in which multiple parameters are present, in particular combinations. Absolutists about vision take this project to be geared towards showing that one and only one feature is represented in each area—i.e, that manipulations of other aspects of context will *not* make a difference in what functional contribution is made by MT, and at most will clarify the ways in which MT represents motion. But I suggest that the evidence speaks against this assumption. I will argue that the functions of perceptual areas have been shown to be sensitive to changes in context—that is, what a particular brain area contributes to vision in a particular situation depends on the type of context instantiated in that situation.

Using MT, and more generally the MFH theory, as a target for analysis has several advantages. First, since it is construed as an exemplar success case of TA, showing that TA is inadequate for MT at least suggests questioning the view at large. This is especially so since many current criticisms of TA are in more “cognitive” settings. As I discuss below, debates about localization have often focused on such broad functions as language, memory, and decision-making. There are significant challenges in understanding, not only whether such functions are localized, but how they should be operationalized in the first place (Klein, 2012; Uttal, 2001). As I discuss below, conceptions of the functions that are potentially localized have changed considerably in the modern history of neuroscience, and characterizing them adequately is still a glaring problem. While perceptual neuroscience is not free of these problems, perceptual stimuli have physical correlates in the world that can be manipulated in extremely fine detail. As such, the ways in which perceptual neuroscientists have learned to manipulate contexts provide examples of how context can be rigorously incorporated into experimental methodology. I will rely on this potential for rigor both in articulating the contextualist theory of function, and for resisting the slide towards equipotentiality. On my view, perceptual neuroscience should still be seen as an exemplar, but no longer as a parade case for absolutism—instead, the case study I advance shows how to begin to manage the transition from an absolutist analysis of function to a contextualist one.

Making this case will involve, in chapters 2 and 3, characterizing the commitments of the MFH theory as an absolutist view, and arguing against that view. But this philosophical work is being done to make explicit a development that the field itself has undergone, and continues to grapple with. An important aspect of the transition is showing how a contextualist view can maintain and build upon the genuine advances that did occur under the aegis of the MFH theory. Theory successions in science, it is widely held, should maintain the successes of previous theories, while both avoiding anomalies that plague them and incorporating a wider range of results. Arguing against the absolutist interpretation of the MFH view opens up a potential reading of its successes that is more confluent with a contextualist perspective—the contextualist view I advocate reads the MFH theory (and perhaps other particular absolutist accounts) as a powerful heuristic for decomposing the brain. It is a commitment of the heuristic view, however, that there be specific ways and situations in which the heuristic fails, and importantly, that these failures provide the impetus for further advances (Wimsatt, 2007). The contextualist theory I defend is specifically intended to provide conceptual resources for moving beyond absolutist-style functional theorizing. It is thus both a descriptive and a normative framework, as befitting a philosophical theory of explanation in the life sciences (Craver, 2007). For instance, inertial commitment to the MFH theory has often underestimated results that are out of keeping with it. The contextualist view shows how to view the MFH theory in relation to new findings, and eventually how to move beyond it. The result, I hope, will be a way of understanding how decomposition can succeed while embracing the complexity of the brain and its connections to the world.

In the remainder of this chapter, I first describe the desiderata that a successful theory of functional localization must meet (section 1.3). In sections 1.4 and 1.5, I show how absolutism has shaped dialectic about neural function in both historical (section 1.4) and modern (section 1.5) debates. In section 1.6, I discuss several theoretical movements that, in my view, are best seen as gestures towards a contextualist perspective. The remainder of the dissertation is structured as follows. In chapter 2, I lay out the MFH theory of visual system function in detail, and articulate its commitments. In chapter 3 I then discuss the results that I take to undermine the MFH theory's functional description of MT, and thus the application of TA to the visual cortex. Chapter 4 assesses a new kind of absolutism, computational absolutism, which has been proposed as a replacement for TA. I argue that this new version of absolutism fails as an advance beyond TA—that is, it fares no better at meeting the desiderata on localized function in an absolutist way. This failure motivates abandoning absolutism entirely. Chapter 5 articulates the contextualist alternative, and argues that it meets the desiderata, thus avoiding the collapse into holism.

1.3. Desiderata on a Successful Theory of Functional Localization.

When I discuss "theories of functional localization," I will mean one of two things. First, I will use the term to refer to meta-theories of how functional localization should work. Such theories describe the principles and methods by which particular brain areas should be functionally distinguished. Second, I will use the term to refer to the particular theories themselves—i.e., theories about the functions of particular brain areas and systems. Absolutism (and varieties thereof) can be viewed as a meta-theory about functional localization. It contends that particular functional theories should seek a univocal functional description for the particular area under study. As mentioned, much of the dialectic about functional localization as such has implicitly been about whether absolutism is a true meta-theory of how the brain can be divided. Similarly, many debates about function in particular areas or systems can be read as disagreement about whether some particular absolutist claim is true.

There are a few desiderata that any theory of localization must meet. The first, naturally, is decomposition. It is definitive of localizationist theories, as contrasted with equipotentiality theories, that they divide the brain into dissociable parts, each of which is defined functionally. A successful meta-theory must provide principles that successfully do so in particular cases. If decomposition fails, the result is holism, the negative claim that there are not meaningful functional divisions in system.[3] Equipotentiality is the positive version of the holism claim. In defending a theory of localization, then, one must give good reason to think that not only are the principles one uses to differentiate areas supported by the *current data*, but also that they are likely to still be sufficient when new data has been uncovered. Only in doing so convincingly can a theory of function hope to avoid holism and equipotentiality. The next two desiderata unpack this need further.

A theory of localization should be empirically adequate. Both meta-theories and specific theories should either explain—or explain away—the set of known data about how brain areas relate to the world and to each other. Theories of localization often have particular data sets that strongly support them, and others that must be explained away (or worse, ignored). Which data get explained, versus explained

---

[3] More specifically, it is the claim that there are not functional divisions within the system that can explain what we want to explain. Traditional vitalists in biology were holists—they believed that *life* could not be explained in terms of mechanistic interactions between biological parts. This is different from believing that there are no parts. Generally, a holism claim is relative to a proposed domain of explanation. In the neuroscience case, it is that there are no meaningful functional divisions relevant for explaining psychological processes. This is compatible, for instance with a large degree of anatomical differentiation, which few in the history of neuroscience have seriously denied.

away, reflects the meta-theoretical commitments of particular views. I will illustrate in detail in chapter 2 how a commitment to TA has shaped the development of the MFH theory, and how data surrounding it is interpreted. In embracing TA, the MFH theory is committed to analyzing the visual system in terms functional parts, each of which is individuated by represented a specific type of information. This commitment, in turn, determines which data can easily be explained by the MFH view, and which must be explained away. I will discuss several strategies that MFH theorists have developed for explaining away seemingly-contradictory data. If data incompatible with a TA-style attribution cannot be explained away, then the data constitute an anomaly for the particular TA theory.

Next, there are the dual theoretical virtues of generalization and projectability. Generalization is simply the notion that our theories should explain more than one instance. It is standardly accepted that good theories in biology generalize about the systems they study (Griesemer, 2005; Levy & Bechtel, 2013; Machamer, Darden, & Craver, 2000), and there are good epistemic reasons for wanting theories that do so. No theory in neuroscience would be acceptable if it described only one single situation—our understanding of the world would not be significantly increased by theories with such a restricted scope. Relatedly, projectability is the idea that coming up with a successful explanation in one instance or context should tell us something about what to expect in other instances or contexts. The notion of a theory “telling us” something about what to expect, however, is deeply ambiguous. Absolutists presuppose that the correct theory of functional localization for a particular brain area should explain or predict how that area will function in all contexts. At first glance, this seems an extremely strong demand on projectability, but given certain background assumptions it is reasonable that it should seem attractive. The worry about projectability spans the metaphysical and epistemic concerns about decomposition. On the one hand, if a system is genuinely decomposable, then we should be able to posit theories whose generalizations capture the dissociations that actually exist. If we cannot, then either we have not described those divisions accurately, or the system is not dissociable after all. The worry that inspires absolutism is that, once we allow function ascriptions to vary with context, we will never know that we have found the right description of the divisions. A new change in context can always show our current divisions to be inaccurate. At the outside, our theories might never extend beyond very specific sets of contexts, and thus we would not in fact have a very good understanding of how the system is really divided (if indeed it is).

However, I will argue throughout that projectability across all contexts is an unreasonable demand for theories in neuroscience. First, it is in tension with the empirical data I will discuss in chapter 3, which seems to show strong contextual variation in function for area MT. Second, it does not provide either good descriptive or good normative principles for fields that are in the process of trying to further understand MT function (see chapter 4). Moreover, I will argue in chapter 5 that, appropriately understood, contextualism can be seen as supporting both generalizability and projectability. Doing so requires abandoning the idea that projectability occurs across all contexts—on contextualism, we should not expect a current function ascription to tell us what will happen whenever an area functions. What we should expect, however, in keeping with Wimsatt's view of theories, is that we should be able to both (i) define the limits of projectability for a certain function ascription, and (ii) that the function ascription provides us with the resources to tell whether a particular context falls under it or not. In chapter 5, I argue that a contextualist perspective on localization can ground this non-universal understanding of projectability. Giving up on absolute projectability is a significant epistemic concession, but the rewards, I argue, are plentiful: contextualism rates strongly both in terms of its ability to provide empirical adequate decompositions of the brain, and in terms of its ability to provide descriptive and normative guidelines for inquiry in neuroscience. This combination of advantages both allows us to see how the brain really is decomposable—at least on a contextualist interpretation of the evidence—and to embrace the heuristic aspects of theories.

In the remainder of this chapter, I briefly discuss historical and modern debates about function to show how absolutism has often been implicitly at work. The goal here is not historical thoroughness, but instead to give enough of the flavor of the debates to show that absolutism has indeed been a background assumption of those arguing for and against localization. Fuller historical accounts are available elsewhere (Bechtel & Richardson, 1993; Phillips, Zeki, & Barlow, 1984; Tizard, 1959; Zola-Morgan, 1995).

1.4. The Historical Importance of Absolutism.

There are, very roughly, three kinds of evidence that have historically been relevant for functional localization. The first two are what Bechtel and Richardson (1993) have called “inhibition” and “excitation” experiments (see also Craver, 2007).[4] Given a phenomenon P and a system S suspected of being responsible for P, interrupting or removing S should cause surcease of P. These are inhibition studies. In excitation studies, the reverse pattern holds; artificially activating S should facilitate P, by making it more likely to occur, or by making it occur more strongly. In neuroscience, lesions studies (either natural or artificial) and electrical stimulation of particular parts of the brain have been the primary sources of inhibition and excitation experiments, respectively. The third kind of evidence is a correlative measure of activity, wherein one measures the activity of S in a range of conditions, and reasons that if S is responsible for P, then its greatest activation will occur along with P. Both modern imaging techniques, which measure the degree of blood flow in an area of the brain as an indicator of its neural activity, and electrophysiology, a direct measure of electrical activity in a part of the brain, are forms of this kind of evidence.

Each kind of evidence has contributed to discussions of functional localization, but there is a standard set of inferences, shaped by absolutism, which occurs in each case. As befitting absolutism, theorists in favor of localization have historically

---

[4] As Bechtel and Richardson (1993) discuss, the localizations that were often attempted in traditional debates were “direct localizations,” in which a single locus is sought for a broad phenomenon such as language or perception. As Bechtel and Richardson note, often debates about direct localization give way to more detailed decompositions, in which the phenomenon is explained in terms of parts performing more specific operations, none of which corresponds one-to-one with the broader phenomenon. Localization projects in neuroscience have broadly followed this pattern. One no longer seeks to localize “perception” as such, but aspects of perception, and sub-aspects of those aspects (e.g., motion, and different kinds of motion). Interestingly, the commitment to absolutism has continued unabated through this transition, as the rest of this chapter will show.

proposed that one of the types of evidence shows *univocal* functioning for some part(s) of the brain. That is, it is one and only one psychological phenomenon that is implicated in excitation, inhibition, or correlative measurement of a brain area. Correspondingly, those against localization have generally proposed evidence that no such univocal functions can be found in particular brain areas. They purport to show either that (i) many or all brain areas contribute to each P, or (ii) that each individual area contributes to more than one P. Historical developments surrounding localization, as well as those taking place currently, can be laid out roughly along these lines.

Modern debate about localized function began with phrenology. Gall (1825), and later Spurzheim (1829) argued that psychological traits were due to specific parts of the brain, and that each trait had a particular locus in the brain. The style of explanation was to find a set of traits that collectively would exhaustively explain the behavior of individuals, and then to localize each trait to a part of the brain. Indivduals' proclivities for behaviors corresponding to the traits (in one particularly colorful example, their proclivity to murder) were explained by the size of their brain loci for the traits in question. The phrenologists preferred method, craniology, rapidly came under criticism from many sources (Zola-Morgan, 1995). More important for our purposes are the conceptual and empirical considerations levied against phrenology. Flourens' (1846) criticisms were based equally on a philosophical conception of the mind, and on empirical considerations stemming from his systematic application of lesion studies. Flourens took the unity of the mind to be manifest in conscious experience, arguing that any division of the mind into components or parts would fundamentally misdescribe it. He tied the possibility of subjectivity, willed action, and religious experience to the purported fact of an undivided mind. “The unity of the understanding, the unity of the *me*, is a fact of the conscious sense, and the conscious sense is more powerful than all the philosophies together” (Flourens, 1846, p. 40). As such, he took any seemingly dissociable psychological processes to in fact reflect “modes” (p. 41) or manifestations of the unified mind. But if there were no reasonable distinctions in parts of the mind, Flourens reasoned, then there would not be corresponding divisions in the brain.

Flourens did not, in fact, think the *whole* brain was undivided, but he recognized only four parts, only one of which (contra the phrenologists, who localized traits in the entire brain) he took to be responsible for “intelligence”—the acting of the understanding.[5] This he located in the “hemispheres” (cortex), and he steadfastly denied any possible decomposition of the hemispheres. The empirical strategy he employed to test this idea involved systematic destruction of the cortex in a variety of animals, followed by observation of their behaviors to see what, if anything, was interrupted. He varied both the location of lesions and the amount of cortex ablated. He argued that no specific behaviors were interrupted, regardless of where lesions occurred. Instead, once a certain size lesion was performed, animals lost their general abilities—they could no longer perform coordinated behaviors. However, the deficits were not restricted to solely (e.g.) sensory or motor aspects of behavior. Since no

---

[5] He posited the cerebellum, medulla oblongata, and tubercula quadrigemina to subserve locomotion, respiration, and sensory input, respectively, based on the fact that these were interrupted when the areas were lesioned.

specific areas in the cortex, according to Flourens, correlated with the loss of specific behaviors, he concluded that the hemispheres were indivisible, acting as a unit. Loss of behavior was solely due to loss of a sufficient mass of cortex, not to location.

What is important for our purposes is Flourens' reasoning regarding localized function, which is undeniably absolutist. The distinctions he draws are between the unified mind and the parceled mind of the kind described by phrenology, between a unified cortex and a cortex where each area matches a division in the mind. If there were functional specification in cortex of the kind that would suggest a divided mind, then each area of the cortex would, when lesioned, interrupt one type of behavior. His data supporting that it was mass, not location, that mattered in lesions, and that it was general behavior, not specific, which was affected, supported a brain/mind in which no divisions could be made. Flourens' conclusion: "The faculty of feeling, of judging, of willing any thing, resides in the same place as the faculty of feeling, judging, or willing any other thing, and consequently this faculty, essentially a unit, resides essentially in a single organ. The understanding is, therefore, a unit" (Flourens, 1846, p. 35).

Flourens effective and trenchant criticisms were part of a significant anti-localizationist movement stemming from dissatisfaction with phrenology's methods and conclusions (Tizard, 1959). In the late 19th century, however, continued investigation of lesion-induced deficits in mammals, along with the nascent technique of electrically stimulating particular parts of the brain, renewed the debate. A particularly strong case for localization came from Broca's (1861) seminal studies of a patient with a lesion in the temporal cortex. The specificity of the patient's language deficits—he could understand speech but could not produce it articulately—strongly suggested a specific role for "Broca's area," namely the area in charge of producing language. Ferrier employed stimulation experiments in an attempt to discover a localization of visual perception, with some success (although there was debate about the precise location; see Bechtel, 2008). Others, such as Goltz, denied localization. Goltz introduced a new empirical finding that he took to question localization—the recovery of function after lesion. His reasoning is aptly paraphrased by Phillips, Zeki, and Barlow: "Restitution of function must be by the activity of surviving areas of the cortex. But according to the theory of localization, every area has its unique function. Then after restitution, one or more areas must have more than one function. And if after the operation, why not before?" (1984, p. 329). The interest in localization via lesion studies was taken up further in the Russo-Japanese war and World War I, where soldiers with lesions to the striate cortex were shown to have significant visual impairments (see Bechtel, 2008, for discussion of these developments). Stimulation experiments were extended by Penfield, among others, in the early 20th century, and Penfield (Penfield & Boldrey, 1937) famously used them to map a bodily homunculus onto the surface of the somatosensory cortex.

Dissent regarding localization continued, however, into the mid-20th century. Its strongest and most influential critic was Lashley (1929, 1930, 1950), who employed similar methods to those employed by Flourens, but in conjunction with detailed behavioral analysis. Lashley trained animals (mice and rats) to complete complex tasks, such as navigating mazes, opening latched boxes, and perceptually discriminating levels of light intensity. He then performed lesions in different parts of the cortex and reassessed the animals' abilities on the tasks. The reasoning, while more fine grained than Flourens', is similar. If each ability—or, as Lashley put it, "memory" or "engram"—of how to perform the task is localized, then lesioning the location should disrupt the behavior, and not others. Lashley admitted only one localization: that primary visual cortex (Brodmann's area 17, subsequently referred to as V1) was necessary for visual discrimination tasks. In rats, lesions to frontal cortex left these established behaviors untouched, whereas for monkeys severing Brodmann's area 18 (subsequently referred to as V2), immediately anterior to area 17, left "perfect retention of … visual habits" (Lashley, 1950, p. 7). He thus denied that visual cortex could be divided further after area 17. For "association areas"—areas between visual and motor cortices—he reported results similar to Flourens', where small lesions had no effect on behavior and large lesions had very general ones.

Importantly, Lashley targeted specific areas of the brain in an effort to disprove existing conceptions of their functions. Removing motor cortex entirely in monkeys, for instance, did not affect latch-box tasks after the monkeys recovered from temporary paralysis. Summarizing his analysis of the motor cortex, he wrote: "The traditional view of the function of this area regards it as the region of final integration of skilled voluntary movements. My own interpretation … is that it has no direct concern with voluntary movement, but is a part of the vast reflex postural system which includes the basal nuclei, cerebellar and vestibular systems." (Lashley, 1950, p. 3).

Lashley went further in proposing a "mass action" view of the cortex—a kind of equipotentialiy theory. Lashley thought that particular *patterns* of activation across the entire cortex were important for specific memories and behaviors, but that only the organization was important. Only the resources of the whole brain, he argued, could underlie performance of a behaviorally relevant function, situated with regards to particular spaces and times (Lashley, 1950, p. 27). As such, he ended up at a view that paralleled Flourens', namely that the brain acts as a unit. "Cerebral organization," he wrote, "can be described only in terms of relative masses and spatial arrangements of gross parts, of equilibrium among the parts, of direction and steepness of gradients" (1930, p. 18). Any specific effects on particular parts of the brain are manifestations of "some qualitatively unitary thing" (1929, p. 565).

While Lashley was primarily concerned with failure to find specific areas with specific tasks, he also suggested that data showing that a single area performs multiple functions undermines localization (Lashley, 1930). In summary, then, the failure to find an area specified for a certain task, or the suggestion that a single area contributes to multiple tasks, should be inferred as suggestive of equipotentiality. "A review of symptoms suggests that no logically derived element of behavior can be shown to have a definite localization; no single sensation, memory, or skilled movement is destroyed alone by any lesion. On the contrary the various parts of the functional areas seem equipotential for such elements, and either a whole constellation of them is affected by the lesion, or none at all." (1930, p. 13). In short, the dialectic is an absolutist one.

From this extremely brief historical illustration, we can extract a few lessons. First, the question of localization is always tied to a particular way of conceptualizing the functions that are supposed to be localized. Reconstruing the functions (e.g., switching from traits to tasks) allows the question of localization to be re-raised. Similarly, scientists can and have drawn conclusions about function on the basis of different methodologies. However, despite differences in method and changes in conceptualizations of function, the conceptualization of localization along absolutist lines has remained consistent. That is, if a single function can be found for an area, then localization is true. If not, then equipotentiality is. In the next section, I suggest that the same dialectic is still at work in at least some current debates about localization in neuroscience.

1.5. Current Debates about Function.

Lashley, to be fair, did not think that there were *no* meaningful differences in the brain. He argued that each area or unit had a particular place in a relational structure comprising the whole brain. Current theorists who deny localization, similarly, often believe that there is *some* kind of differentiation in the brain; however, their dialectic in thinking about those divisions is still strongly shaped by absolutism. While Lashley's criticisms were influential for a time, from the 1950s onwards a wide-ranging enthusiasm for localization emerged that has continued to the present day. In part, the change was inspired by the emergence of electrophysiology and imaging as recording techniques, and by successes in understanding vision at the neural level (detailed in chapter 2). Cataloging all of the aspects of the modern acceptance of localization that has occurred since Lashley would be difficult. Localization is currently pursued as an explanatory goal in many projects in systems neuroscience, in motor systems and other perceptual modalities besides vision. In cognitive neuroscience, localization has been the primary goal of imaging studies, which attempt to measure activity in awake humans during reasoning and perceptual tasks.[6] Often, cognitive neuroscience focuses on "higher-level" cognitive activities, such as reasoning, decision, and action planning (Uttal, 2001). Despite the recent enthusiasm, there is still considerable debate surrounding the issue of localization.

Current dialectic about localization has, arguably, advanced little in terms of the general shape of the positions and the dialectic that occurs between them. Shallice (1988), after considering a wide range of worries for localizing independent, unitary functions based on lesion data, decides that it must be done on pragmatic grounds: "Avoiding being deceived by a tempting shadow will be less important than dimly making out the existence of some possible path" (Shallice, 1988, p. 266). Uttal, criticizing Shallice's view among others, cites a variety of considerations based on imaging data—namely that a number of areas are likely to be active during any given task, and the fact of complex interactions between areas of the brain—as undermining localization. While he denies equipotentiality writ large, he offers only primary sensory and primary motor cortices as examples of areas with relatively successful

---

[6] Considerable debate has accrued around localization in imaging studies in recent years, both in terms of whether localization is in fact shown, and in terms of whether imaging evidence is a good indication of it (Hanson & Bunzl, 2010). I will not discuss these debates in detail here.

localizations. He denies that any other structural differences in the brain can amount to localized functions, and ends up at a view significantly like Flourens', which denies Shallice's appeal to pragmatism, and the "analyzability" of the mind into components. It is worth quoting him at length:

> Compartmentalization of mind into a set of components is a convenient concept to help us organize what would otherwise have to be considered a composite and integrated as well as unified 'mind'. Indeed, it may be that it is extremely difficult to study mentation in any other way. Yet exactly because of this pragmatic convenience and ease, it is possible that we are being led astray from a truer, more valid, and more realistic conceptualization of the unified nature of mental processes based on widely distributed brain mechanisms. If our search is for truth and an actual rather than a convenient or pragmatic understanding of mind and brain, then it may be necessary to consider what the alternative—an indivisible, composite, aggregated, and unanalyzable mind—might be like. This is the main alternative raised by those of us who argue against the uncritical acceptance of mental analyzability and the cerebral localization of high-level cognitive (i.e., nonsensory and nonmotor) processes. Indeed, though seemingly mute these days, there have always been critics who have challenged analyzability and suggested that what is being measured in different experiments are the measurable properties of a unified mind rather than separable entities. (Uttal., 2001, p. 135)

So, while the brain may have divisions—for instance in structure—they are not useful for understanding the mind. This is surely a major blow to the explanatory aims of localization. And we are stuck in the same dialectical place as in the traditional debates: either we accept staunch absolutist localization, or we arrive at holism and an undivided mind. The more explicit recognition of and debate about pragmatic value, reflective of the discussion of decomposition in section 1.2, does not change the dialectical options.[7] Other theorists have cited similar considerations as

---

[7] To be fair, much of Uttal's critique is based around the vagueness of psychological notions. His claims about "analyzability" of the mind, or lack thereof, are often focused on whether psychological traits can be differentiated and validly operationalized. He proposes a rejuvenated behaviorist approach

arguing against TA. Meta-studies, for instance, which look at results across a large number of tasks in individual studies, have shown that cortical areas tend to be active in a number of different situations, where different tasks are being performed. However, rather than abandoning absolutism, several theorists have taken the results to motivate the formulation of different *kinds* of absolutism, thus showing how strongly absolutism has taken hold. I discuss these views in chapter 4.

It is also worth noting how the assumption of absolutism has shaped some more specific empirical debates about particular areas. A now-classic example is that of the Fusiform Face Area (FFA). In 1997, Kanwisher, McDermott, and Chun (1997) performed an imaging study which measured brain responses to visually presented faces. In virtue of the increased activation of the FFA, they posited that perception of faces is localized in that region. In very short order a variety of other studies were published purporting to show that the FFA was *not* face specific: studies showing FFA activity to houses, birds, cars, and places were offered, as well as studies suggesting that FFA activity is increased with more knowledge of the object (e.g., in car experts versus non-car experts; see Gauthier, Skudlarski, Gore, & Anderson, 2000; Gauthier & Tarr, 2002). The alternative interpretation is that the FFA is a *general* mechanism—that it responds to any perceptual information for which the subject has perceptual expertise or experience. A variety of strategies have been employed back and forth in

---

to psychology, for which I have considerable sympathy. The problem is that Uttal offers no perspective for understanding how brain and mind relate on his preferred approach, opting instead for his own vague claims about unity and lack of analyzability. In my view, an appropriate understanding of how the brain can be decomposed can help in overcoming the vagueness of our psychological terminology. However, for that benefit to accrue, we must first have a sense of how to do localization, and Uttal offers nothing in this regard.

the debate, including using inhibition techniques (transcranial magnetic stimulation) to gauge the specificity of interruption of the FFA for facial processing (Pitcher, Walsh, Yovel, & Duchaine, 2007), and using high-resolution fMRI to see if there are divisions *within* the FFA that *are* face specific versus some that are not (for different views on these results, see Grill-Spector, Sayres, & Ress, 2006; Haxby, 2006).

In keeping with the absolutist framing of the debate, in a recent review article Kanwisher (2010) writes: "This article focuses primarily on the question of functional specificity, because this is the question that is critical for understanding the architecture of the human mind" (Kanwisher, 2010, p. 11164). She then goes on, however, to list only four areas for which she takes sufficient evidence of localization to have been established, including the FFA.[8] Even if she is right about these areas, which is heavily debated, finding only a handful of functionally specified areas seems to show that it is *not* "the question" that is important for understanding how the mind is organized. If localization must be read in an absolutist way, then finding only a small number of localized functions shows that localization is not in fact very important, since most of the brain operates without it. Moreover, Kanwisher says something puzzling about the data for these areas: "None of these regions is the only one with its defining selectivity. For faces, selective responses are found not only in the FFA but also in a nearby but more posterior occipital face area, as well as other regions in the superior temporal sulcus … and anterior temporal pole" (Kanwisher, 2010, p. 11165). But precisely this kind of data—showing that a particular area is *not*

---

[8] The others are the posterior parietal area, which Kanwisher argues is specified for perceiving places, the extrastriate body area, for perceiving biological bodies and their parts, and the temporoparietal junction, underlying reasoning about others' mental states.

the only one activated or affecting a proposed function or process—has often inspired anti-localizationist claims, since if multiple areas are relevant or necessary for some function, then that function is not localized in one of them. I take it that something has gone seriously wrong with the dialectic regarding localization and specialization, and I propose that absolutism is its root cause. I will spend the rest of the dissertation fleshing out why, and proposing a contextualist alternative.

These debates are internal to neuroscience, but absolutism has also crept into dialectic surrounding cognitive architecture in cognitive science writ large. For instance, critics of psychological modularity often present results showing multiple informational influences on brain areas, as well as plasticity, as evidence against the thesis (Buller & Hardcastle, 2000; Prinz, 2006). Philosophers have taken to citing localization data as evidence in debates about the neural correlates of consciousness and the nature of conscious perception (Block, 2007; Clark, 2009). I will not expound these debates here; suffice to say that if the current trend towards absolutist localization is overturned, then many in these fields will need to rethink some of their argumentative strategies. In the next section I offer some movements that I take to be on the right track. I see many advances in neuroscience as progressing slowly towards a rejection of absolutism and an embrace of context as partially constitutive of neural function.

1.6. Movement towards Contextualism.

Absolutism has gone deep enough that it is often simply taken as definitive of localization. Consider the following quotes from Hardcastle and Stewart (2005):

> Brain plasticity and concomitant multifunctionality belie any serious hope of localizing functions to specific channels or areas or even modalities. (p. 28)
>
> Searching for the function of any area is a fool's errand. The same area could be doing different things, depending on what else is happening in the rest of the brain. Or, perhaps more accurately, the brain might emphasize or privilege one process over another in the same area, depending on circumstance. … It doesn't make sense to ask, as many cognitive neuroscientists are prone to do, What does this area really do? For the answer will always be that it depends. (p. 36)

Here in the second quote, seemingly, is recognition of the importance of context for understanding function. But the initial quote says that the same results show function cannot be localized![9] I know of one place in which a philosopher has taken a positive stance towards context as contributing to functional localization. Klein (2012) analogizes the "pluripotency" of functional areas to a diesel engine equipped with an engine break—what the pistons do by pumping depends on whether the brake is engaged. This is fine as far as it goes, but is at best a starting point. The analogy suggests that we can learn about how things function by looking at external context, but falls well short of saying whether a contextualist view can meet the desiderata on a theory of functional localization. For instance, Klein writes, somewhat sanguinely, that "if we restrict ourselves to the same context, we can infer that a

---

[9] In this case as well as in Uttal, the worries about localization are based on the inability to find pre-conceived psychological categories in the brain. These theorists don't seem to think that this failure undermines decomposing the brain. Even if this is so, and even if they are right, we are left without a manual for how to do localization, given that our ways of understanding functional decomposition are deeply bound up with psychological notions. Moreover, if one abandons psychological notions entirely, as Uttal advocates, then interpreting the tasks becomes very difficult, and this difficulty redounds to understanding how the brain should be decomposed. A major advantage of the contextualist theory I propose in chapter 5 is that it shows a way for taking initial decompositions based on intuitive notions—in this case, intuitive notions of perceptual properties such as motion, color, etc.—and showing how to employ them in an attempt to generate neurally plausible contextualist function ascriptions.

specific cognitive function is employed" (2012, p. 956). It is highly questionable that this will satisfy anyone who takes projectability seriously, since it confines all of our conclusions about function to specific contexts.

Klein also embraces "network" approaches towards understanding the brain, and it is indeed projects in this vein that I take to have made the biggest steps towards a contextualist theory. The network perspective is nebulous, and involves both a shift in focus regarding how to understand the brain, and the application of formal methods new to the neurosciences. The basic shift in focus is to privilege the patterns of interaction among parts of the brain. With the massive amounts of anatomical data that have been gathered in several model organisms, and the development of computational tools to analyze it, there is hope that we can understand how the brain is organized anatomically prior to any particular conception of how its parts function. The tools of graph theory—which provides measures for how closely connected different parts are and how the closely connected groups tend to be connected to others—provide detailed measurements of brain structure without presupposing a functional view.

Of course, one must then interpret how function is realized in the anatomical structure, and neuroscientists working broadly within the network perspective have taken a range of positions about how revisionary a view of function is suggested by a focus on networks. Some take network analysis of the visual system, for instance, to reveal how the MFH theory of visual system function, discussed in detail in the next chapter, is implemented (Meunier, Lambiotte, & Bullmore, 2010; Müller-Linow, Hilgetag, & Hütt, 2008). Others have taken up a half-way position. Sporns (2011), one of the leading figures in graph-theoretic approaches, asserts continually in his seminal book, *Networks of the Brain*, that function is both "segregated" and "integrated." It is segregated because each area, due to its place in the network, contributes something unique, and has unique physiological properties, but is integrated due to the high level of connectivity between all parts of the brain. This is an appealing position, in that it seems to walk a middle ground between an absolutist description and something like Lashley's relationalist view. It is unclear however, what precisely it means to be both segregated and integrated, and how to analyze functions of this sort. Moreover, it is unclear precisely how such a view is supposed to avoid collapsing into full-fledged holism, and how it is supposed to connect up with physiological data. While I will not discuss the network perspective in detail (see, however, section 5.4), the contextualist view I propose can aid in navigating what it is for function to be both segregated and integrated, since it focuses on what differentiates the functions of areas that are highly dependent on factors outside of them.

A full-out commitment to contextualism has been defended by McIntosh in a number of publications (McIntosh, 1999, 2004). McIntosh proposes that functionality of a particular area is sensitive to a combination of "neural contexts," the other parts of the brain that are operative at a given time, and "catalysts," the behavioral and perceptual circumstances in which the co-activations occur. He writes:

> The conflict between localization and distributed functions no longer exists in this view of brain function: functions are localized in the sense

> that their expression requires the integrity of certain regions or that a particular type of information is available to an area (e.g., visual, motor), but the actual expression of any function results from the actions of several regions. Such an integrated view of brain operations will likely bear more fruit as the notions of context and catalysts are tested and refined. (2004, p. 179-180)

The goal of this dissertation is to show how such a promise can indeed bear fruit. Specifically, I will focus on catalysts—the perceptual circumstances that determine what perceptual areas respond to and represent.[10] If the contextualist perspective is to genuinely aid in understanding the brain, there must be a methodology in the offing that will be both rigorous and meet the goals of a theory of functional localization. I argue that such a methodology is already implicitly at work in perceptual neuroscience, and I aim to articulate it. Doing so, however, requires showing in detail how absolutist approaches fail to adequately localize function. This is what I undertake in the next several chapters, focusing on area MT.

---

[10] Of course, the concept of representation is a troubled one, and much philosophical ink has been spilt about the nature of mental representations, and how they are situated in the natural world (for a recent review, see Rowlands, 2009). My project here is not a philosophical analysis of the notion of representation. In perceptual neuroscience, representation is generally tied to the kind of information that a neural response represents "explicitly," and I will analyze this notion in chapter 2. Representation, on this view, is taken to be a property of both information and use—a part of the brain represents the world in virtue of a privileged informational relation between a feature of the world and its responses, such that information about that feature can be used in further processes. As such, the notion of representation here matches up well with Millikan's (1989) notion of "consumer semantics."

Chapter 2. Absolutism and Visual Cortex Function

2.1. Introduction.

The standard application of TA in the visual system is the "modular functional hierarchy" (MFH) theory. The theory adopts TA as a guiding principle, positing that each area of the visual system represents a specific feature of the perceived world, and explains the function of the visual cortex as a whole in terms of a hierarchy of functionally specified areas. Assessing the MFH theory requires an understanding of what kind of theory it is and what its commitments are. In this chapter, I argue that the MFH theory is best understood as implementing what Griesemer (2012) calls a "formalism"—the use of an abstract structure to describe the structural divisions within a particular system of study. Reading the MFH theory as implementing a formalism both gives a good explanation of the kinds of reasoning undertaken by the theory's proponents, and a clear articulation of how and why the MFH theory is committed to TA.

The notion of hierarchy discussed in this chapter involves the *flow of information* between distinct parts and levels. The explanation of the entire system's operation is due to specific information being processed at each location in the hierarchy. Higher levels of the hierarchy operate over the outputs of lower levels, and sometimes send commands back to them. I will refer to such systems as "processing hierarchies." Processing hierarchies are important in technology, particularly in the design of computer systems (Bechtel & Richardson, 1993), as well in a variety of areas in biology (Love, 2012). In neuroscience, the hierarchical approach to studying the visual system has influenced the study of other systems, including other sensory systems and the systems underlying action control (Pacherie, 2008).[11]

In this chapter, I will attempt to characterize the *theory* that lies behind hierarchical explanations of the visual system. What type of theory is a hierarchical explanation? What kinds of evidence support a hierarchical explanation, and how do they support it? How does the theory deal with seemingly countervailing evidence? And, most importantly, what are the commitments of a hierarchical theory? What sorts of evidence speak against such a theory, and why? I will argue that Grisemer's notion of a formalism provides good answers to these questions. Employing a formalism consists in drawing a form/content distinction, wherein the particular content of the theory—in this case, a description of the visual system—exhibits an abstract structure that is divestible from particular cases. Since the formalism is abstract, it must be *implemented* to explain a particular case. I will argue that the MFH theory's alliance to TA, are due to its means of implementing the hierarchy

---

[11] There are a variety of related notions of hierarchy, which must be kept distinct. A compositional hierarchy is one in which a whole is comprised of its assorted parts, and those parts are in turn comprised of still smaller parts. Proponents of mechanistic explanation (e.g., Craver, 2007) have focused on the importance of compositional hierarchies in neuroscientific explanation, and Churchland and Sejnowski (1992) have explained the importance of the notion for neuroscience as a discipline. Compositional and processing hierarchies are importantly different, however, in that compositional hierarchies arguably do not involve the flow of information between their levels, and parts at one level arguably do not causally interact with their components (Craver & Bechtel, 2007). Processing hierarchies must also be distinguished from categorical or type hierarchies. Type hierarchies are semantic, or perhaps (depending on one's ontological predilections) metaphysical. They divide up the world into types, such that particulars at one level are instances of a more general type, which is further an instance of a still more general one. But as with compositional hierarchies, information is in no sense exchanged between different categorical levels (although finding out that x is a member of category Y might carry some information for an observer), which is distinctive of processing hierarchies.

formalism.[12] I will use the formalism approach to illustrate the history of the MFH theory, and articulate its commitments.[13]

In articulating the MFH theory, I will rely strongly on the work of two researchers, Semir Zeki and David Van Essen, whose results were foundationally important in developing the view. While many researchers, of course, contributed to the theory's development, the breadth of Zeki and Van Essen's work, their many theory and review papers, and their occasional disagreements are helpful for understanding the commitments of the view. An important advantage of viewing the MFH theory as a formalism is that the notion of a formalism itself is neutral as to the ultimate epistemic standing of the theory. That is, the formalism notion accounts for the kinds of reasoning undertaken by the proponents of the MFH theory without presupposing that theories are either fundamental truths or heuristics. The formalism interpretation thus is compatible with both the large amount of credence that theorists have granted to the MFH theory, and the subsequent heuristic employments of the theory that I will emphasize in chapter 5.

2.2. Formalism and Abstract Theory.

Abstract ideas often guide scientific inquiry, but they are notoriously slippery, due to their very abstraction. What are the empirical commitments, really, of the idea that objects have an appropriate place or orientation in nature, as Aristotelian physics held? Of the idea that "perfect" motions, such as those that should be found in the

---

[12] I will often simply refer to the "MFH formalism"; this should be read as referring to the implementation of a hierarchical formalism by the MFH theory.

[13] Other descriptions of the development of the MFH theory can be found in (Bechtel, 2001, 2008). My exposition will be selective—I will focus mainly on aspects of the theory that clearly bring out its absolutist commitments.

heavens, are circular? The latter claim was influential for all of natural philosophy until Kepler and Newton, and some historians have claimed that adherence to the fundamentality of circular motion prevented Galileo from discovering the law of inertia (Cohen, 1985). The commitments of such views are difficult to pin down, since they are constituted by strongly-held, only partially empirical assumptions. For that same reason, when they *are* overtuned, the effects are sometimes revolutionary.

The idea of a processing hierarchy is an abstract one, as evidenced by the wide variety of social, biological, and technological systems to which it has been applied. But to understand whether the idea is empirically adequate for describing a particular system, we need a view of the role the idea is playing, theoretically, in a scientific investigation of that system. Following Griesemer (2012), I suggest that the MFH view of the visual system is a *formalism*. Formalisms articulate an abstract structure, the form, which is independent of the content, or the particular entities, activities, and properties that play roles in the structure. One of Griesemer's examples is Mendel's distinction between "factors," which obey particular mathematical principles (Mendel's "laws"), from "characters," the particular traits that are determined by the factors. The abstraction from content in the form allows for the study of the mathematical laws of assortment in *any* particular trait. Other examples Greisemer discusses include Frege's distinction between logical form and mathematical content, Darwin's logic of evolution versus particular examples of evolutionary change, and Wiesmann's theory of embryonic development (discussed further below).

A formalism consists of two aspects: (i) a list of the basic *types* of entities involved in the system, and (ii) what I call "in-virtue-of relations," which explain how causal processes will work in systems implementing the formalism.[14] The commitments of theories based on formalisms stem from the need to "align" (Griesemer, 2012, p. 299) the formalism with empirical data regarding the system of interest. In order to implement a formalism, one must show that the system being studied has components that instantiate the types specified in the formalism, and that those components stand in the appropriate in-virtue-of relations. If these conditions hold, then the system is adequately described in terms of the formalism. If they do not, then it falls outside of the formalism's "domain"—that is, the theory is false when applied to the system of interest.

To illustrate formalisms and implementation, consider Griesemer's (2012) discussion of Weismann's theory of embryonic development. Figure 2.1, drawn by Weismann (1893), charts the embryological development of a worm from a single fertilized cell. The different shadings of groups of cells represent distinct cell types, and the nodes between cells show the chain of development. In the diagram, however, is a more fundamental distinction, which is the beginning of a formalization of development. The cell line whose branches progress through the center of the diagram (before splitting apart in the top third) is what Weismann called the "germ" line. This is the line that he observed to create branches of new types of cells. Once branched, the new types did not further differentiate into still further different types, but only

---

[14] These elements of a formalism are at implicit in Griesemer (2012), but I believe are important for fully fleshing out the notion of a formalism. I will argue for this largely by illustration.

divided to create more cells of their specific type. Moreover, each germ cell divided into both a new germ cell and an instance of a new cell type.

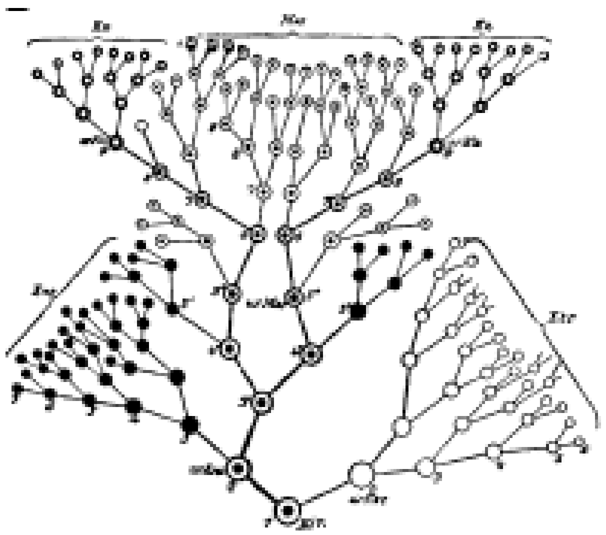

Figure 2.1. Weismann's (1893) diagram of cell development in the worm.

The fundamental distinction between types of entities proposed by Weismann, aspect (i) above, was between germ cells, progenitors of new cell types, and "soma" cells, cells which can only generate further cells within their particular type. This distinction is fully extracted in Figure 2.2, produced later, and thus represented as a formalism.

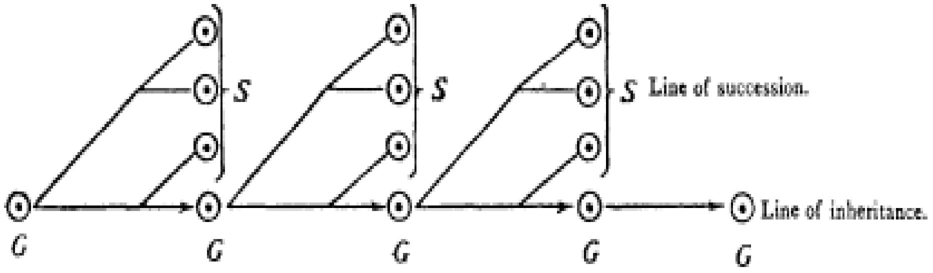

Figure 2.2. The formalization of "Weismannism" (Wilson, 1900).

The diagram captures only the split of each germ cell into another germ cell and a soma, thus eliminating information about any of the particular cell types that stem from the soma line. The formalism is thus “a distinction of genealogical form—the pattern of cell lineage relationships—from embryological content—the specific nature of each of the cells of the developing body” (Griesemer, 2012, p. 304). The “pattern” of relationships is guaranteed by the in-virtue-of relations in the system. Development is explained by the fact that *new cell lines* come only from germ cells—i.e., occur in virtue of germ cells—and that differentiation within cell types come only from soma cells. So the overall pattern of development is explained by the distinction between types of entities in the system and the different processes that occur in virtue of those distinctions.

While Weismann’s original diagram was vital for developing the formalism, it can also be seen as an implementation of it. It shows that the types of entities and in-virtue-of relations occur in the system. Similarly, for any new system to which the formalism is applied, evidence must be given that the distinctions exist and the in-virtue of-relations hold, regardless of the specifics of the entities playing those roles. Just as vitally, it is easy to see when applying the formalism to an organism would be *incorrect*. It would be incorrect, for instance, if there were no clear distinction between germ and soma cells in the system, or if new cell types occasionally came from soma cells instead of germ cells.

When a theory implements a formalism, the commitments that are essential to the formalism—i.e., the need to produce evidence for the part types and the in-virtue-of relations that the system posits—can be considered the *core* or fundamental commitments of a type of theory. As such, the formalism can account for the *flexibility* of abstract theories, both in terms of allowing for internal debates and incorporating new evidence. Debates are internal to a theory if what they disagree about does not involve the tenets of the formalism, but instead involves details about how the formalism is implemented—e.g., which entities fill which system divisions, or how, specifically, an in-virtue-of relation is carried out. New results can be accounted for if there is a way of describing how they fit within the general structure posited by the formalism. The formalism also explains how potentially countervailing evidence is handled by the theory. Accepting a formalism means that one's theory is empirically adequate (recall, from the last chapter, that this is a desideratum on theories of localized function) only if all of the evidence can be explained, or explained away, in ways compatible with the basic entity-types and in-virtue-of relations posited in the formalism being maintained. Finally, viewing an abstract theory as a formalism does not *itself* make any presuppositions about the epistemic status of the theory—for instance, whether it is best seen as a contextually-bound heuristic or as an exceptionless truth about the domain. Either view is compatible with the formalism reading of the commitments of a theory. This is helpful in the neuroscience case, since in this and the next chapter I will show how the theory's epistemic use has shifted: once considered a truth to be established, it has since transitioned into a more heuristic type of theoretical role.

When looking at a particular theory and its history, there is good evidence that the theory is based on a formalism if it can be articulated in terms of distinctions between types of entities and in-virtue-of relations, and if the history of the theory reflects attempts to maintain those distinctions when dealing with new evidence. If so, then the formalism reflects the commitments of the theory. In the next section, I argue that the MFH view is an application of a hierarchical formalism to the visual system, and that its TA-type commitments stem from this implementation.

2.3. The MFH Theory of Visual Cortex Function.

The idea that the visual system is hierarchically organized is actually a very old one, stretching back at least as far as the empiricist philosophies of the 17th and 18th centuries. Consider, for instance, Locke's (1700) view of sensation and ideas. Locke thought that the basic sensations consisted of sensory qualities—color and shape, primarily—and that "ideas" of objects were the mereological sum of basic sensory qualities. Associationism of this type has had a large amount of influence on neuroscience (Tizard, 1959). The idea of a processing hierarchy in perception recapitulates Locke's intuitive notion, on which perception works by first independently representing distinct types of features, then combining those feature representations at later stages. The functional decomposition of the visual system, on the MFH theory, is predicated on finding anatomical and physiological correlates of these individual representations.

Here I will argue that the notion of a processing hierarchy, as implemented in the MFH theory, is a formalism. First, it involves a particular type of system-division, and a way of typing entities. Entities are functionally typed by their ability to represent *specific*, univocal types of information, at a specific place in the hierarchy. Second, the MFH theory employs a characteristic in-virtue-of relation, which I call the "use-relation," on which representations at lower levels of the hierarchy are in turn used as inputs for processing at higher levels, leading to further specific representations. Since the implementation of a particular formalism is definitive of the MFH theory, it is committed to traditional absolutism—that is, the theory must interpret the parts of the visual system to be functionally defined by representing, univocally, one type of perceptual information. This commitment has shaped the interpretation of evidence within the MFH perspective.

2.3.1. A (Brief) Primer on Electrophysiology.

I have been somewhat loose in discussing the brain data employed in debates about localization. Since I will focus heavily on electrophysiology in the succeeding chapters, a few more words about it are in order. Electrophysiological techniques consist in measuring the electrical potentials in certain parts of the brain. While this is done in a variety of ways, one powerful method involves the implanting of electrodes in particular locations to measure the potentials of particular cells.[15] When an electrode is implanted, some aspect of the stimulus is changed, and the results are recorded. Generally, a cell has a background level of activation at a given time, and the property of interest is how much the electrical activity changes as a result of the

---

[15] Other electrophysiological techniques include electroencephalography, which measures large patterns of electrical activity in the brain via electrodes external to the skull, and the measurement of local field potentials, summed activity from large numbers of neurons. I will not discuss these techniques in detail.

change in the stimulus. Cells can be recorded either intracellularly (with the tip of the electrode inside the cell) or extracellularly (outside), and the measurement is of changing electrical potential inside or just near the cell membrane. Arrays of electrodes can be used to record extracellularly from multiple cells at the same time. Cell activity is measured in action potentials, the brief spikes in membrane potential from a resting, depolarized state to a polarized one via which cells propagate signals down their axons to other cells. When a cell is activated by a stimulus, its firing rate of action potentials increases from its baseline to some higher level. Since the activity level of the cell is dependent on the stimulus, its activity is generally taken to encode information about that stimulus.[16,17]

Often a cell's response properties are not precise. If a cell responds to motion, for instance, it will show the greatest response to motion at a particular velocity (combination of direction and speed), and somewhat less response to velocities in the same neighborhood. These patterns of response are called "tuning curves," and are generally taken to provide the functional characterization of the cell, with its primary

---

[16] Electrophysiological methodology is not free from worry generally. One major concern, specifically with single-cell electrophysiology, is "selection bias." No studies find that all cells in a brain area of interest respond to the stimulus of interest, even for classically successful discoveries. There is thus a worry that the cells that are selected for study are the ones that match the preconceived notions of the experimenter. It is not my purpose here to engage in deep methodological critiques of current recording techniques; I will primarily follow the conclusions of working scientists, for whom such concerns as selection bias are standard pragmatic worries.

[17] 'Information' is no less troubled a notion, philosophically, than 'representation' (see chapter 1, note 9), and again I will spend relatively little time on it. The neuroscientific approach to information tends to bridge purely mathematical and "semantic" notions of information. The former can be cashed out in terms of joint probability distributions between two variables, X and Y (e.g., an aspect of a perceptual stimulus, and a neural response), such that a change in the probability of X entails a change in the probability of Y. The latter involves a privileged relationship between a signal and a particular designation or reference (Piccinini & Scarantino, 2011). Again, as with representation, the notion of "explicit" information is meant to bridge the gap between the notions in perceptual neuroscience, and I will discuss this in detail in the next section.

function to represent the stimulus that causes it to become the most active. Evidence for functional localization in an area, in physiological research, is based on the finding that all or most of its cells have tuning curves reflecting a particular stimulus property—motion, color, faces, etc. The area as a whole represents that property; cells and groups of cells within it represent specific variations or values of that property. So, if the stimulus contains motion of a particular velocity (or motion in multiple velocities) the specific response of a motion-detecting area as a whole will reflect that velocity (or velocities).[18]

Individual cells in the visual system have "receptive fields" (RFs). These are often defined both spatially and functionally. A cells' "spatial RF" is the area of the visual field in which light of the right sort causes the cell to be activated. Individual visual areas have cells with spatial RFs of different sizes, and the combined RFs constitute a "map" of the visual field that is both topologically distinct at different areas and distinct in terms of how much detail is represented. Area V1 in the visual system has a detailed, fine-grained map of the visual field due to the small RFs of its cells, whereas areas in mid-level areas such as V4 and MT, and high-level areas such as the temporal cortex, have progressively larger RFs and thus represent considerably less spatial detail. A cell's spatial RF is often further divided into "classical" and

---

[18] It is important to note that conclusions such as these are always statistical. There are no universalities in as complex a domain as systems neuroscience. Neuroscientists compute a large variety of statistics, and perform detailed controls, in an attempt to establish their particular conclusions, but this does not eliminate the fact that conclusions from a neuroscientific study are always abductive. An entire dissertation could be written about the methodological assumptions underlying electrophysiology, and the statistics employed, and it is not my purpose to give a full methodological critique here. Again, I will largely follow the scientists, since my goal is to trace the way that empirical results have shaped conceptual progress surrounding localized function. I will flag specific statistical considerations only as relevant.

“non-classical” RFs. The classical is the area of the visual field that activates the cell, whereas the non-classical are areas in which light can modulate, cancel or accentuate that response. Generally, non-classical effects are due to lateral interactions (often inhibitory) between cells in the same area, which are taken to sharpen responses of the area as a whole.[19] The functional component of an RF is *what* stimulus at the appropriate RF drives the cell. I will be largely concerned with this aspect in what follows.

2.3.2. Modern Physiology and the Use Relation.

Kuffler (1953), building on the early physiological studies of Hartline’s (1938), showed that cat retinal ganglion cells had spatial RFs. Ganglion cells fell into three types: ones that responded to the shining of a light in the center of their spatial RF, ones that responded to the light turning off, and ones that responded to both. Moreover, there was variation within the RFs of individual cells, with the center of the RF causing the strongest response (of whatever type), areas at the edges causing less activity, and occasionally intermediate regions that differed in response properties. These results suggested a very sophisticated and specific function for retinal ganglion cells—responding to particular types of stimulation at specific places in the visual field. In the 1950s, Lettvin et al. (Lettvin, Maturana, McCulloch, & Pitts, 1959) showed that in the retinal cells of flies, brief, small movements of dots in the RF elicited a strong response. Importantly, they suggested an ecological and behavioral

---

[19] Chirimuuta and Gold (2009) review evidence that the spatial RFs of individual V1 cells are dynamic—that is, that they can change depending on the stimulus—and consider these results to be in tension with the idea of a classical spatial RF. The considerations I will evince are roughly parallel, but for the functional aspect of the RF.

function for this signal—since the stimuli regularly elicited tongue flickings towards the dots from the frogs, the signal could be seen as a 'fly' detector. Their findings were highly influential in popularizing the notion that important aspects of perceptual stimuli could be represented in the responses of single or small groups of neurons, in particular parts of the brain (Barlow, 1972).

The MFH view got its initial motivation from Hubel and Wiesel's Nobel prize-winning research, which extended physiological investigation into the mammalian cortex. Hubel and Wiesel (1959), largely by accident, noticed that cells in the cat's striate cortex (Brodmann area 17, subsequently called V1) had different response properties within their RFs than either retinal ganglion cells or cells in the lateral geniculate nucleus (LGN), the first synaptic connection upstream from the retina, which provides input to V1. As in the retina, some cells responded to light onset in the center ("on" cells), and some to darkness in the region ("off" cells). Striate cortex cells, however, responded vigorously to extended light discontinuities or "edges," consisting of bars of light or shade of specific widths and orientations within their spatial RFs.[20] As with Lettvin et al.'s study, there was an obvious ecological import to this finding—cells that respond to extended light continuities could provide useful information for eventually recognizing the form of objects in the visual field. In continuing investigation, Hubel and Wiesel (1962) subsequently showed that there were still more complex receptive fields in area 17. Unlike the "simple" receptive

---

[20] Cells in V1 also have surrounds within their RFs that are inhibitory. That is, if an appropriate stimulus falls in the center of the RF, the cell is activated, but if in the surround, then not. This property is important for the orientation selectivity of other types of V1 cell (Hubel and Wiesel, 1962; and see below).

fields of cells that responded only to edges within the center of their RF, "complex" cells responded to specific orientations anywhere within their RFs, which were generally larger than in simple cells. "Hypercomplex" cells had specific orientation preferences, but were inhibited unless a bar ended within the center of their receptive field. Further, some cells had strong responses to motion (displaced bars) within their receptive fields.

What's important for understanding the MFH view is the *hierarchical* relationship between simple, complex, and hypercomplex cells that Hubel and Wiesel proposed. On this view, simple cell responses to edges are determined via anatomical connection to particular LGN cells, specifically ones whose RFs are arranged in the shape of the orientation and width to which the simple cell responds. So, the physiological properties of simple cells downstream from LGN cells are determined by the physiological properties of LGN cells, plus their arrangement. Similarly, Hubel and Wiesel proposed that simple cells with similar orientation preferences provide input via anatomical connections to complex cells, so that the complex cell will respond to that orientation at any of the simple cell RFs. This produces both the larger RFs of complex cells, and their responses to an appropriately oriented edge anywhere within their RFs.[21]

A few things to note about the view of function implied here. First, function is tied to physiological response to stimuli, where these responses are constrained by anatomy (Hubel & Wiesel, 1968). As discussed in chapter 1, the function of particular

---

[21] In a discussion of cat cortex, Hubel and Wiesel (1968) distinguish between "lower-order" hypercomplex cells, which have spatially specific RFs, and "higher-order," which are like complex cells in not preferring stimulation of a specific area in their RF.

cells is individuated by the feature they represent, and this is discovered by finding a privileged relationship between a certain perceptual feature and the cells' responses. For cells other than those at the retina, physiological response is determined by the physiological responses of cells closer to the sensory periphery, plus their anatomical connections to the cell(s) in question. Thus, the cells earlier in the pathway provide a specific type of *information* to the latter cells—about particular stimuli at specific places in the visual field—and the latter cell's physiological response, while different from the input, is a function of that input. The latter cell's physiological response, reflecting its specific information, can then potentially be "used" by other parts of the visual system, as Hubel and Wiesel point out in a famous quote, which ends their 1968 paper:

> Specialized as the cells of 17 are, compared with rods and cones, they must, nevertheless, still represent a very elementary stage in the handling of complex forms, occupied as they are with a relatively simple region-by-region analysis of retinal contours. How this information is used at later stages in the visual path is far from clear, and represents one of the most tantalizing problems for the future. (1968, p. 242)

The *use* condition presented here is the in-virtue-relation posited by the MFH view, which can be phrased as follows:

- *Use relation*: for any neural unit X, X performs its function in virtue of the information represented in the earlier levels of the hierarchy that provide it with input.

I use the term 'unit' here because the use relation can apply both to individual cells and to distinct areas in the brain, which I will discuss below. The use relation is constitutive of a processing hierarchy, where the functional divisions are based on the passing of information between distinct levels (e.g., LGN cells, simple cells, complex cells). As such, the use relation provides a way of differentiating the contributions of different cells, and how they relate to each other in performing an overall visual function. For instance, Hubel and Wiesel discuss how complex cells' responses, while technically carrying less detailed information about the stimulus (information is carried about a specific orientation, but at a less specific place of the visual field), than simple cells, can be derived from the responses of simple cells. Carrying visual information at different levels of detail might be a vital function of the visual system. Hubel and Wiesel imagine specific cells for representing "specific forms" (1962, p. 146), without representing their detailed properties. An individual cell might represent, for instance, that the perceived object is a face, without representing any of the details of its shape, its position in the visual field, etc. Importantly, Hubel and Wiesel suggest that the same conjunctive logic that holds between (e.g.) simple and complex cells could be extended to account for the relationship between "lower" and "higher" visual areas (Hubel & Wiesel, 1965, p. 275) to account for this kind of more complex form recognition.[22]

The MFH theory built directly on the successes of Hubel and Wiesel's studies. While Hubel and Wiesel focused primarily on V1, the MFH theory extends the logic

---

[22] Another thing to note is that Hubel and Wiesel explicitly claim that the hierarchy view is speculative in nature, and qualify it in many places, calling it "oversimplified" (1968, p. 217) and tentative. Among other considerations, they point out that the division into simple and complex cells is not hard and fast (they say that "new subtypes are continually appearing"; 1962, p. 109), and that complex cells do not universally have larger receptive fields than simple cells. A common theme we will see is that different theorists qualify statements to different degrees and in different ways. The challenge is to show that, despite being qualified, laden with provisos, and differently interpreted, theoretical views such as those based on processing hierarchies can have specific commitments. This is what the formalism reading is intended to provide.

of their studies to provide a theory of function for the entire visual cortex.[23] This involved further differentiating the cortex into more areas, based on a combination of anatomical and physiological evidence, and expanding the hierarchical logic to account for whole brain areas and their interactions, as well as particular cells. It also involved extensions of the view of V1 proposed by Hubel and Wiesel.[24] I will discuss some of the key developments here briefly, without elaboration, before undertaking a detailed discussion of how TA-style function attributions were applied to MT and V4 (for a review, see Livingstone & Hubel, 1988).

Different groups of retinal ganglion cells differ in their properties, both in terms of their RF size and their receptivity. Magnocellular or "M-pathway" cells respond faster and have larger RFs, while parvocellular or "P-pathway" cells respond more slowly and have smaller ones. These divisions project to different layers of the LGN, and in turn to distinct parts of V1. At V1, the P-pathway is further split into two pathways, called "blobs" and "interblobs" due to the specific shapes in which their inputs cluster in V1. The standard interpretation of the three anatomically segregated pathways, as laid out canonically by Livingstone and Hubel (1988), is as follows. M-pathway cells respond to certain orientations exhibiting movement in particular directions, as well as disparity between the two eyes (discussed below). Interblob regions respond to orientation, but not to motion or disparity. Blob regions are color specific, responding solely to particular wavelengths. Livingstone and Hubel thus

---

[23] The view is also a rejection of one of Hubel and Wiesel's original proposals, based on studies of the cat, that the hierarchy consisted of three levels—areas 17, 18, and 19—which were themselves not divided functionally (Hubel & Wiesel, 1965).

[24] Most of this subsequent work was performed in the macaque monkey, and I will discuss studies of the macaque primarily in what follows.

describe the divisions as a “motion” pathway, a “form” pathway (the interblobs), and a “color” pathway (the blobs). The term ‘K-pathway’ is occasionally applied to the pathway mediated by the interblobs. While each pathway responds to luminance contrast, they respond to different aspects of it (i.e., form versus motion/disparity). Both in the LGN and V1, cells receive input from particular points of the retina, thus underlying their spatial RFs, and ensuring that the spatial layout of the visual field is recapitulated, if with abstractions and deformations, at V1. This is true for cells in each pathway, which allows for each pathway to represent its specific type of information at any point in the visual field. This is broadly referred to as “retinotopy,” and is continued, albeit with increasing RF sizes in individual cells, until the anterior regions of the visual system.

Importantly, the division of V1 into multiple divisions does not undermine feature specificity. V1 is standardly seen as a “segregator” of information, in which the different pathways have specific representational functions in distinct subdivisions. The MFH view posits that these pathways remain distinct throughout visual cortex, and that as the signal progresses through levels of the hierarchy, different specific features are extracted by further areas. V2 maintains the separation between the three pathways in its own architecture, with distinct parts responding to disparity, luminance, and color. The next stage further splits the pathways into their own distinct areas. The M-pathway projects to area MT[25], with motion-selective cells

---

[25] There has, historically, been serious debate as to the nomenclature of MT, or the “middle temporal” area. Zeki, in most of his publications, refers to it as ‘V5’, while Van Essen has traditionally preferred ‘MT’. This disagreement is not important for current purposes; I will use ‘MT’ throughout, since this is the current consensus label.

projecting directly to it from V1, and disparity-selective cells projecting to it from V2.[26] MT is posited to extract motion information specifically from the M-pathway input (see below). Area MST then uses this information to represent more complex types of motion (e.g., spiral motion). Information about these types of motion is then used in parietal areas to make "decisions" about the spatial trajectories of objects, and in the frontal eye fields (FEF) to guide eye movements.[27] The P- and K-pathways are posited to send information to distinct areas of V4, which represent shape and color, respectively. This distinct information is then sent on to further areas in the temporal cortex, which use it to represent more complex shape patterns and to eventually recognize objects. For instance, individual neurons in the inferior temporal area have been shown to respond preferentially to particular types of objects such as faces or hands (Gross, Bender, & Rocha-Miranda, 1969; Gross, Rocha-Miranda, & Bender, 1972), an impressive confirmation of Hubel and Wiesel's "form cell" speculation. The distinction between M-pathway projection through the dorsal part of visual cortex to the parietal cortex, and P-pathway projection through the ventral part to the temporal cortex, has been incredibly influential. Ungerleider and Mishkin have defended a "where" and "what" functional distinction between the dorsal and ventral streams, respectively (Mishkin, Ungerleider, & Macko, 1983). The MFH version of

---

[26] Area V3 intercedes between V2 and the MT/V4 split, but its function is still relatively poorly understood—it is not clear, for instance, what it contributes that is not already performed at V2 (Anzai, Chowdhury, & DeAngelis, 2011).

[27] It is an open question where in the brain conscious perception occurs. Zeki (2001) argues that it occurs directly in feature-specific modules, while others assume that motion information must be "read off" further down the processing stream in order to perceive specific features. I will stick to the notion of representing information here, and not make any guess as to where in the brain conscious perception occurs.

the hierarchy suggests that each functional area within each stream represents a specific feature, at a specific level of detail.

So, to sum up, the MFH view posits that each area beyond V2 is specific in the type of information in represents.[28] What is it, however, to represent a type of information specifically? The answer, in literature surrounding the MFH view, is generally that a certain type of information is *explicit* in the responses of cells in the relevant area, where explicitness of information at an area is defined as information (1) that is *distinct* from the information available at other areas, and (2) extractable by the area based solely on its inputs. Consider complex cells in Hubel and Wiesel's framework. The information their responses carry is distinct from that of simple cells (condition 1). While both simple and complex cells respond to edges, the information they carry about where in the visual field the edge is differs, because complex cell responses to edges occur over their whole spatial RF, not at a specific place within it, and because complex cell RFs are larger (condition 1). This information, however, is derivable from the inputs from simple cells (condition 2). Explicitness, thus defined, is an absolutist notion, in the traditional sense. According to the MFH view, each area is supposed to convey one type of feature information explicitly, and individual cells within that area reflect particular values of that feature. So, for instance, MT represents linear motion explicitly, and activation of individual cells within MT

---

[28] Many more areas have been isolated in visual cortex, and discovering their functions is one of the key goals of visual neuroscience going forward. Fuller diagrams involving all of the areas can be found in Van Essen and DeYoe (1995) and other locations. I will be focusing on the cases of MT and V4, because these are the hallmark areas for functional specification, and for differentiation of the two visual streams. My claims in this project can be read as a suggestion that if feature-specificity and TA are not the best ways of analyzing even hallmark success cases of functional explanation in the visual system, then exploration in the rest of the visual cortex should not proceed along those lines, either.

reflects different directions and speeds of motion (see below). Only motion information is carried explicitly. Thus, Zeki and Shipp (1988, p. 19) state that explicit information is “exclusionary” (cf. Bartels & Zeki, 1998, p. 2330) and that there can be only one type for each functional area. This information is then available for use at later hierarchical levels.

A variety of functional considerations are advanced to support this view. Van Essen and De Yoe (1995) analogize the perceptual system to a factory, which is organized to perform large tasks by splitting them into specialized subtasks. Zeki and Bartels (1998) discuss the need to represent different properties of the stimulus at the same time, suggesting that specialization allows for maintenance of information that “would otherwise be lost.” (p. 2329). The ideas here, while different, both reflect the notion that feature-specific representations and the use relation are vital to how the brain achieves visual perception. That feature specificity is what is meant by these theorists is easily seen in earlier quotes from Zeki (1978), who says that different areas are “functionally specialized to analyze different features of the visual environment” (p. 423) such that “at every area a different type of information is analyzed” (p. 428). As an example, area MT, on Zeki’s earlier view, is a “motion area” (p. 426).

The hierarchical arrangement proposed between areas is shown in the following diagram from Van Essen and Anderson (Van Essen & Anderson, 1995).

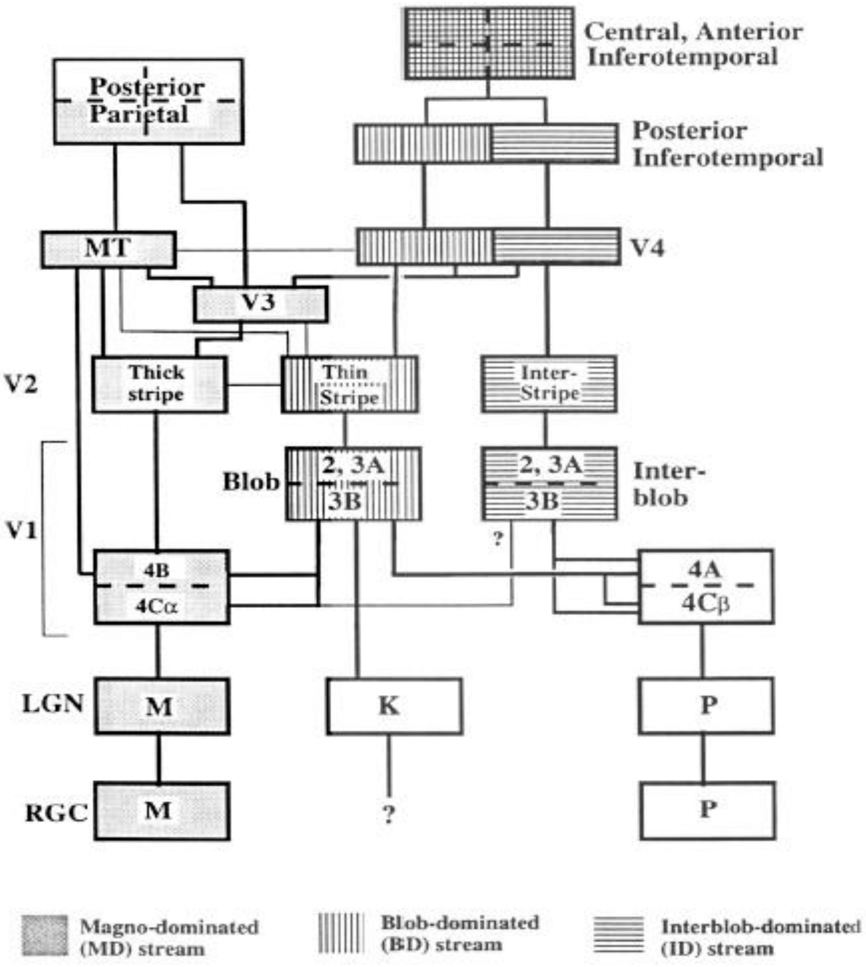

Figure 2.3. A simplified diagram of the visual hierarchy. From Van Essen and Anderson (1995).

The inherent tension at the heart of the MFH theory, which I will focus on throughout, is the relationship between the commitment of feature-specific, explicit representations as definitive of localized function, and the complexity and high degree of interconnectedness in the visual system. This can be seen even in the highly simplified diagram above, in which connections are present between the different areas and streams. If the distinct areas are connected with each other, how can it be that each processes only one sort of information explicitly, to the exclusion of others?

This issue has been faced by MFH theorists throughout the history of the view, and their ways of handling it show the flexibility and power of the formalism. Every major review article (and many research articles) qualifies the processing hierarchy, as well as the distinct streams within it, by discussing the high degree of anatomical and physiological interaction between streams and areas. In 1991, Felleman and Van Essen (1991) cite an average 19 connections between each area and other areas in the hierarchy, and an even higher average (27 connections) in the levels of the hierarchy between V1 and V4/MT. Physiological results also showed that responses are "multiplexed" (Van Essen & Deyoe, 1995, p. 393), carrying a variety of "subsidiary" types of information, along with the information represented explicitly. Finally, as in the factory analogy, one specialized subprocess might be needed for multiple functions (consider, for instance, the production of ball bearings, which go on the play a role in many other processes of production). Thus, a way of communicating the explicitly represented information for use in a variety of different functions is important (Van Essen & Deyoe, 1995; Zeki & Shipp, 1988). The challenge is to account for this complexity in a way that maintains feature specificity, and this commitment is knowingly taken on by proponents of the MFH theory.

Consider the following quote: "The physiological properties of any given cortical neuron will, in general, reflect many descending as well as ascending influences. Nevertheless, the cell may represent a well-defined hierarchical position in terms of the types of information it represents explicitly and the way in which that information is used." (Felleman & Van Essen, 1991, p. 32).[29]

The quote can be read as acknowledging the commitments of the MFH formalism—both establishing feature-specific representation, and showing how that information can fulfill the use relation—as well as the need to make sense of complex interaction. It is possible, however, to get more specific about the types of interactions between areas that are compatible with explicit feature representation at each area. Three—input, output, and modulatory interactions—are already implicit in the above quote.

- *Input interactions*: a variety of types of information may serve as "cues," for explicitly representing a specific feature.
- *Output interactions*: the explicit feature representation may be passed onwards from an area to multiple other areas for use in distinct further functions.
- *Modulatory interactions*: interactions with other areas/types of information in the hierarchy can shift the particular interpretation of an explicitly represented feature, without changing what feature is represented.

Input interactions are based on the idea that multiple more basic features earlier in an area's functional stream might be relevant for determining the

---

[29] True, Felleman and Essen use the plural "types" here, but this shouldn't be taken to mean that they think individual areas are not feature-specific. While motion is the standard feature-specific function attribution to MT, there are multiple aspects to motion, including speed and direction (see below), which contribute to motion representation in MT. In other work, Van Essen (Van Essen & Gallant, 1994) clearly posits motion representation as MT's function.

interpretation of an explicitly represented feature (Van Essen & Anderson, 1995). This is intuitively obvious for object recognition—determining that an object is a horse could depend on many more basic features, including its shape, color, movement patterns etc. But it is also true for individual features. Determining color, for instance, depends not only on wavelength but also on luminance.[30] So, in general, multiple lower-level features might be *used* by the area to compute its explicit feature value. Output interactions are based on the idea that, once an explicit feature representation is established, a variety of different further areas might use that representation in performing their own functions. I have already discussed one example above: an explicit representation of motion direction and speed at MT might be used in either the parietal cortex to keep track of object location, or by the FEF to guide eye movements. Similar relations could occur for other feature representations. Modulation interactions can come either bottom-up or top-down. The idea is that an area represents the same feature in every case, but it might represent that feature in a specific way or in a specific situation based on input regarding other features or from object representations at higher hierarchical levels. I discuss each of the strategies much further through specific examples below.

There is one further strategy which is relevant, which we have already seen at work in descriptions of V1 and V2—the "differentiation strategy."

---

[30] Despite the importance of the input strategy, it is rare for theorists within the MFH perspective to posit that cues from different pathways can interact. Van Essen and Gallant (1994) stress the independence of processing for motion and form streams. Zeki and Shipp (1988), while accepting that generally multiple "attributes" might be relevant for computing a particular feature, suggest that this is unlikely to hold for form and color interactions. Van Essen and Anderson (1995) are a unique case, in admitting that color and motion might "indirectly" influence each other as inputs. However, they do not spell this out in detail, and the idea is not generally taken up in the literature.

- *Differentiation*: an area formerly thought to represent only one feature can be decomposed further into subcomponents, each of which represents only one feature.

Differentiation will be relevant below, and in the next chapter. In TA, and the MFH theory, it is feature-specificity that is of primary importance. Whether the feature-specific areas match precisely with anatomical areas is secondary.

The picture that emerges is one where an area takes the lower-level information that is relevant in computing its explicit feature representation, represents a specific value of its specific feature, then passes that representation on to multiple areas. Zeki and Shipp refer to this as "marshalling and distributing," (1988, p. 22) saying "An area performing a specialized … function will tap any source of information that is useful" (1988, 21) and then distribute it to multiple areas, suggesting both input and output interactions (Phillips et al., 1984; Zeki, 1978). The strategies are absolutist—they attempt to deal with interaction by showing that an explicit, feature specific representation can be localized at each functional area, and then clarifying how that representation interacts with other representations at other areas. The strategies also are anti-contextualist in the way distinctive of absolutism. They attempt to show that the functional localization, in terms of a feature representation, does not vary depending on the context in which an area functions. Any changes in an area's responses due to context are due to one of the interactions—differences in input, output, or modulation—and *not* to a difference in the type of representation the area contributes to processing in that context. On such a scheme, what an area represents does not change with context, but multiple influences can contribute to details of that representation in specific situations.

We can now also begin to see what genuinely disconfirming evidence for the MFH view would be—the discovery of non-feature-specific representational functions (i.e., representations that systematically confound multiple features, or areas that involve explicit representations of multiple distinct features at different times), where the evidence for that representation cannot be accounted for in the ways described above. In the next chapter I argue that such evidence exists in MT. In the rest of this chapter, I selectively discuss some of the classical evidence in favor of the MFH view, and show that at every stage one or more of the strategies have been used to account for potentially countervailing evidence. The evidence taken to support the MFH theory has generally been taken to establish precisely feature-specificity and the use relation, which are definitive of the view on the formalism reading. I first show how anatomical data and physiological evidence have used the strategies discussed to establish feature specificity. I then show how a variety of types of evidence have been used to support the use relation. I focus primarily on MT and V4, since these are often taken to be primary cases of functional localization, and since I will discuss MT further in the next chapter.

2.4. Some Classical Evidence in Favor of the MFH Theory.

2.4.1. Anatomical Data.

The locus classicus of anatomical work within the MFH perspective is Felleman and Van Essen’s (1991) meta-study of histological work in macaque visual cortex. They studied 32 cortical areas, and combined connectivity data from a large number of histological studies to discern the relationships between the areas. The areas studied exhibited 31 percent of “full connectivity,” where ‘full’ refers to what would be expected if all areas were connected to all others. Moreover, most of the connections between areas were mutual—that is, if an area provided input to a second area, that area was likely to provide input back to the first area. Felleman and Van Essen employed known anatomical properties of the cortex to show how a hierarchical organization could be intelligible in such a highly interconnected system. The basic principle relies on the layered organization of the cortex. Each part of the cortex has six horizonatal layers exhibiting specific connection patterns. Each layer has stereotypical patterns of connections other areas of the cortex. Felleman and Van Essen split the layers into three categories. Layer 4 they referred to as the “granular layer,” and layers above or below layer 4 they referred to as “supragranular” or “infragranular,” respectively. The stereotypical pattern of inter-area connectivity exhibited by the layers is as follows. “Feedforward” connections are those originating from either supragranular or both supra- and infragranular layers of one area and connecting to layer 4 of a second. “Lateral” connections originate in both supra- and infragranular layers and terminate in all three layers of their target. Finally, “feedback” connections originate in either supra- and infragranular layers or just in the infragranular, and terminate in both supra and infragranular layers.

A hierarchy of levels between the connected areas can then be defined on purely anatomical grounds. Two areas are at the same level of the hierarchy if they exchange only or predominantly lateral connections, or if they each receive feedforward and feedback connections from the same other levels. An area is at a higher than another if it receives only feed-forward connections from that area (and others at that level) and provides only feedback to it. An area is at a lower level than another if the reverse is true.

The hierarchy, as explored by Felleman and Van Essen, is purely an anatomical distinction. It accounts for most of the relationships between areas, and divides them into 10 hierarchical levels (Felleman and Van Essen admit that by changing the stringency of conditions on levels, other outcomes are possible). Despite the anatomical nature of the hierarchy, Felleman and Van Essen tie it explicitly to the visual streams hypothesis, and construe it as supporting feature-specific localization. As noted in the quote above, Felleman and Van Essen take explicit information to be possible despite the large degree of interconnection between areas. They use a combination of input/output and modulatory strategies to account for this prevalence of "cross-talk" (Felleman & Van Essen, 1991, p. 39). Their account of relationships between streams closely follows the input/output description given above. Regarding the M-pathway, for instance, they suggest that, "once motion is extracted" (Felleman & Van Essen, 1991, p. 41) from a variety of cues, it can be used for a variety of further computations. For top-down connections they suggest a modulatory relation, specifically one that allows for stimuli outside of a cells' receptive field to affect that cell. I will discuss this kind of claim further in section 2.5.[31] Next, I turn to

---

[31] Another possible consideration is the strength of connections between distinct areas. Felleman and Van Essen suggest that only 30-50 percent of connections may be "robust," while Zeki (without citing

physiological evidence. I will discuss both V4 and MT here, since the MFH theory posits that the two visual streams diverge at this stage, and that each area represents a specific type of feature.

2.4.2. Physiological Evidence.

The physiological evidence supporting the MFH view has classically been interpreted as establishing the explicitness of feature-specific representations and the use relation. Establishing explicitness involves showing that a physiological response meets the two conditions discussed in section 2.3—the response must carry information that is *different* from the information available at earlier hierarchical levels and other areas at the same level of the hierarchy, and this information must be derived from the area's inputs. The general process for determining explicitness is to vary multiple parameters of the stimulus and determine those for which the cell is *specifically* sensitive (Maunsell & Van Essen, 1983a). A cell's response should change only with changes in the relevant parameter, thus insuring that it carries information specifically about that parameter. In general, this is an oversimplification, and the strategies are often employed in separating the explicitly represented information from other information that affects responses.

2.4.2.1. Physiological evidence of feature-specificity in V4.

Zeki's (1977, 1983) pioneering studies of V4 showed that a particular type of information is present in the responses of area V4, and therefore available for further

---

data) suggests that the MT to V4 connections are "not strong" (2001, p. 76). It is unclear where such considerations fit into the strategies discussed above, and it is perhaps best to see this move as a way of "backgrounding" certain anatomical data (Griesemer, 2012). In any event, recent anatomical studies (Ungerleider, Galkin, Desimone, & Gattass, 2008) show strong anatomical connections between MT and V4, so I will not discuss this potential strategy here.

use, which is not present at previous levels of the hierarchy. In an early paper, Zeki (1977) established that there were connectivity differences between V4 and MT, despite their proximity in the superior temporal sulcus. He had also established that V4 cell responses were driven by color, not motion. A follow-up paper (Zeki, 1983) further explored how color responses in V4 differed from responses at V1. As mentioned, cells in both the P- and K-pathways exhibit spectral sensitivity. In V1, cells receiving input from these pathways also show wavelength opponency in their receptive fields—their responses are increased by certain wavelengths, and inhibited by others. In V4, some cells also have wavelength opponency, thus exhibiting similar responses to V1 cells. Zeki therefore sought to determine the difference between V1 and V4 cell responses to color.

Zeki asked whether cells in V4 showed responses primarily to “natural” colors—colors matching those perceived by actual perceivers—or merely to wavelength. It is well known that perceived colors (‘red’, ‘green’, etc.) are not strictly tied to particular representations of wavelength; that is, different wavelength combinations can be perceived as the same color. Zeki constructed a variety of stimuli and asked human observers to sort them by color. Some of the colors, as expected, were sorted into the same category even if they had different wavelength combinations. He then showed those stimuli to monkeys within the spatial RFs of particular V4 cells. Some cells responded only to wavelength, in that their responses were always the same to particular wavelengths (e.g., in their opponency responses) regardless of whether the perceived colors were the same or different. Other cells had responses that were tightly tied to the presence of a particular natural color. They would respond, for instance, when ‘red’ was presented, regardless of what combination of wavelengths was included in the stimulus. Moreover, they did not respond when another natural color was presented. Zeki interprets the results as showing that natural colors are represented at V4, but not before. He also suggests that this “transform” (1983, p. 764) might be due to interactions within V4 between cells of the two types—those that exhibit wavelength opponency similar to cells found in V1, and those that represent natural colors. So, while wavelength opponency signals are an input to V4, and even though those signals are recapitulated in some V4 responses, what individuates V4 functionally is the emergence of a new sort of sensitivity—to natural colors—which can then play further roles in the hierarchy. This reflects a combination of the input and output strategies for determining the explicit information represented in an area. While some V4 cells show similar response properties to V1 cells, this is not what defines V4’s function, which is determined by the information that V4 represents uniquely on the basis of those inputs. Once natural color is explicitly represented, it is then available to any number of outputs.

While Zeki claimed to have discovered no responses in V4 cells (of either type) to changes in orientation or motion in their RFs, it soon became clear that V4 in fact has many orientation selective cells. Desimone and Schein (1987), while not denying Zeki’s claims about color representation in V4, established that a large number of cells in V4 showed responses to the length, width, orientation, and spatial frequency of the stimulus, which they took to suggest that V4 “appears to be as much a form area as it is a color area” (p. 861). Subsequent analysis isolated influences of the P-pathway and K-pathway to distinct parts of V4, thus reflecting a differentiation strategy, as shown in Figure 2.3 above (DeYoe & Van Essen, 1988; for a recent argument to this effect, see Tanigawa, Lu, & Roe, 2010). The employment of a differentiation strategy does not undermine either absolustism or feature-specificity, but simply draws a more fine-grained distinction between the parts of the visual system.

2.4.2.2. Physiological evidence of feature-specificity in MT.

Motion perception, like perception of other properties of objects, occasions its own unique set of problems and challenges. Motion is an inherently dynamic property—the change in spatial location of an object over time—and this is difficult to track using cell with constricted spatial RFs. For one, motion across the whole extent of a cell’s RF could be due to motion in multiple distinct directions, since each would produce the same pattern of stimulation in the RF. This is called the “aperture” problem, and is shown in Figure 2.4 below.

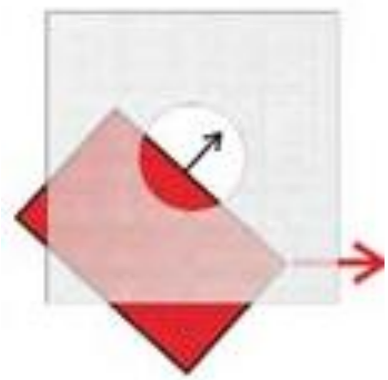

Figure 2.4. The aperture problem. Modified from Born and Bradley (2005). The stimulus is moving in the direction defined by the red arrow, but is perceived as moving in the direction defined by the black arrow, due to the circular receptive field of the filter.

Second, a single stimulus often contains movement in many different directions. Moreover, the motion stimuli from a single object may not be spatially contiguous, as objects may be oddly shaped or occluded. This is an issue of “segmentation,” or deciding which motion signals in the scene are due to one object moving in one direction. Partially, the solution to these problems involves computing “pattern motion,” the overall movement in a particular part of the visual field, which comprises the combination of particular motions, or “components”. The MFH theory posits that MT’s role in the hierarchy is to represent linear pattern motion in the scene. This information is different from what is represented both in V1 and in later levels of the dorsal stream.

Zeki (1974) was the first to propose based on physiological evidence that the primary stimulus driving cells in MT is motion in a particular direction, and Maunsell and Van Essen (1983) subsequently extended these results to show that MT cells had both direction *and* speed preferences. Zeki explored in detail the directional responses of MT cells, and showed that most individual cells in MT had preferences for particular directions of movement within their receptive fields—each would respond most strongly to motion in its preferred direction. Zeki suggested that MT cells were "overwhelmingly" (1974, p. 559) direction-sensitive, and moreover that this direction selectivity was greater than that present in V1, suggesting that motion becomes "relatively more emphasized" (1974, p. 568) at this stage of processing. Maunsell and Van Essen expanded these results by showing that MT cells had detailed tuning curves for direction, with their preferred responses eliciting the greatest response. They also explored the speed preferences of MT cells. Most exhibited a tuning curve for speed as well as for motion—if direction was held constant in a cell's preferred direction, and speed varied, cells showed greater responses for their preferred speed than for others.

These discoveries formed the core conception of MT as an area specialized for representing motion at particular direction and particular speeds. In both papers, the authors stress the finding that none of the MT cells studied responded to specific wavelengths in the stimulus, but did respond strongly to particular aspects of motion. Moreover, motion responses in MT were at least roughly retinotopically organized, with cells RFs covering particular places in the visual field (larger that the RFs of M-pathway neurons in V1). Importantly, the responses to motion were distinct from and more specialized than representations of displacement in V1. As Zeki says, "in this area not only are directionally selective cells far more common than in area 17 … but, in contrast to area 17, a good many cells respond to movement independently of contour" (1974, p. 568). Notice the "a good many" language. Some recorded cells did in fact change their responses to certain orientations of moving bars—they had a specific direction preference, but fired more for certain orientations moving in that direct than for others. Why is this not an ambiguity of feature information, reflecting both motion and orientation rather than motion explicitly?

Both papers employ the input strategy to suggest that orientation is not as important in determining cells' responses as motion (Maunsell & Van Essen, 1983). Zeki suggests that the influence of orientation is a residual effect of MT's receiving input from V1. The cells representing motion regardless of orientation, on the other hand, are truly "selective" for motion, since they respond to only direction within their RF, and not to any other feature. As discussed regarding V4 in the previous subsection, the input strategy allows for some information to be shared between hierarchical levels, so long as there are responses that are distinct, and that carry distinct information, at the higher level. Zeki analogizes the cells showing orientation effects to simple cells in V1, which reflect both the orientation of a presented edge and their specific spatial location in their RF. Motion-only cells he analogizes to complex cells, which privilege *only* orientation, at the expense of the detailed spatial information. What is distinctive about MT, he suggests, is that there are cells that carry only motion information.

Maunsell and Van Essen argue in a slightly different fashion, employing a combination of input and modulatory strategies. First, they show that a stimulus that is only a point (i.e., not an elongated edge with an orientation) will drive an orientation-sensitive MT cell if it moves in the cell’s preferred direction at its preferred speed. So, at best, certain orientations can increase or decrease the response within the cells preferred motion conditions. They also argue that when a cell has a preferred orientation, that orientation is generally perpendicular to the preferred direction. Thus, as an input, perpendicular orientation serves to maximize the amount of movement in the cell’s preferred direction. A parallel orientation bar, for instance, would only enter the RF moving in the preferred direction at a single point, whereas a perpendicular bar would enter across the entire width of the RF. Thus, while details of orientation in the input might modify the response, motion is represented explicitly.

Both arguments contend that, even though some influence of orientation is present at MT, its motion responses are distinctive and separable from those influences. Moreover, each paper suggests that the motion-representing properties of MT are importantly distinct from those in V1. Therefore, in the parlance of MFH, motion is explicitly represented in MT. This reasoning was further supported by a subsequent finding that MT, but not V1, exhibits significant motion *opponency*. Snowden et al. (Snowden, Treue, Erickson, & Andersen, 1991) showed that if MT cells were recorded when *both* motion in their preferred direction and the opposite direction was shown within their RFs, their responses were considerably lessened. V1 cells did not show this property. Snowden et al. suggest that opponency of this sort can be important for determining the overall pattern motion in the stimulus (this will be discussed much further in chapter 4), thereby suggesting that pattern motion is explicitly represented in MT, and not in V1.

Maunsell and Van Essen (1983) also use the input and modulation strategies in combination in discussing the responses of MT cells to binocular disparity. Information stemming from similar parts of the visual field, but hitting the two eyes in different places, is conjoined for the first time in V1. Once it is determined what points in visual space are affecting what parts of the two eyes, the *difference* between those two points can be a relevant cue for *depth*. Consider an object 5 feet in front of you. Light reflecting from this object hits your two eyes at a particular angle. Now move the object to 15 feet away. I will discuss depth and disparity in more detail in chapter 3. What is relevant for now is that while Maunsell and Van Essen showed that some MT cells had preferences for motion in a particular direction and speed *at a particular disparity*, they did not take this to indicate representation of depth in MT.

Their reasoning was based on the idea that motion is the primary function of MT. If MT is to represent depth, they suggest, then it is likely to represent *motion* in depth. Movement in depth inherently involves *changes* in disparity over the time course of the stimulus, for instance those that indicate that the object to be moving towards or away from the perceiver. They tested motion stimuli at changing disparities, and found no significant MT cell tuning for particular changes. That is, MT cells did not reflect changing disparities: they responded much more strongly to frontoparallel movements (movements at a fixed depth plane) than to stimuli that changed disparity during presentation. While the response was stronger to movements at the preferred depth, cells responded to motion at a variety of different depths. The authors thus conclude that there is thus no depth representation in MT, but instead that the disparity signals modify an existent motion representation. This reflects a combination of the input and modulation strategies, because disparity information is known to be fed forward to MT from V2 cells (Livingstone & Hubel, 1988), but is posited to only modify the existing motion representation, not be used as a cue for perceiving depth specifically. As Maunsell and Van Essen note, “In no case … was the response to a motion in depth better than to frontoparallel motion at the best disparity. … We do not believe this type of response is correctly interpreted as selectivity for motion in depth.”[32]

I will discuss disparity and depth in much more detail in the next chapter. For now, the foregoing can be summarized by noting that the strategies allow for the idea of an explicit, feature-specific representation attribution for an area to be maintained despite the influence of a variety of other types of information. The picture that emerges here is one in which neural areas within the hierarchy are dedicated to representing specific features explicitly, which constitute the output contributed by the area to other parts of the hierarchy. Other responses reflect inputs to an area or modulation effects. The TA element of the view is thus dependent on isolating a particular feature that is represented at an area, such that all other influences fall into

---

[32] Another potential argument, which Maunsell and Van Essen do not pursue specifically, could be based on the fact that disparity tuning curves in MT do not differ significantly from those in V1. Several categories of disparity tuning curve in V1 were uncovered by Poggio and Fischer (1977). The categories include cells that respond most to “zero”-disparity (i.e., where there is no difference between the angles presented to the two eyes compared to the “plane of fixation,” the depth at which both eyes are focused; see chapter 3), others that respond to any disparity that *isn’t* near zero, and still others that respond to a range of disparities, but only on one or the other side of zero disparity. Recall the distinctness condition on explicit representation, on which explicit information must be different from that available at other areas. The fact that the tuning curves in MT were not found to be significantly different from those in V1 might be implied to undermine the distinctness condition. Since Maunsell and Van Essen do not pursue this argument in detail, and since there is now physiological evidence against it (see chapter 3, fn. 7), I will not discuss it in detail.

one of the other categories. There are also a variety of types of evidence that have been used to support the use relation. I here discuss these briefly, before moving on to show how viewing the MFH theory as a formalism makes sense of a variety of extensions to, and debates surrounding, the theory.

2.4.3. Establishing Use.

The use relation is generally established in one of two ways. One is by showing that there is an explicit kind of information available at an area, tracing the area's outputs, and inferring that the areas projected to must use the explicit information in their own functioning. This is one of the points of the "output" strategy discussed earlier.

The oldest way of establishing use is by lesion studies, and lesion data has occasionally been taken as evidence for the MFH theory. The idea is that if lesion to an area interrupts only processes requiring representation of a certain perceptual feature, then the area must represent that feature. Zeki and Shipp, for instance, suggest that human lesion data show that "different attributes of the visual scene, such as form, color, and motion, are processed in separate, anatomically distinct regions in the visual cortex, each executing its function with considerable autonomy (1988, p. 311). However, there are a large number of challenges and debates in interpreting lesion data. For one, evidence of plasticity makes it difficult to interpret lesion-induced deficits. For instance, even with complete MT lesions, with continued training monkeys can re-develop motion detection abilities (although not quite to pre-lesion levels) within a matter of weeks. Further, the specificity of deficits from lesions to both MT and V4 has been questioned; Schiller (1993) has argued that both MT and V4 lesions cause a variety of deficits, some of which overlap. Due to these issues, and due to the presence of other methods for determining use, I will not discuss lesion-studies in detail here.[33]

A less direct way of positing the use relation is to compare particular neural responses with behavioral measures. If manipulation of a particular neural response (either activating or inhibiting it) correlates with changes in behavioral measures, then there is evidence that the perceptual system uses information from the area for that behavioral response. Often the behavioral measures involve perceptual judgments employing the perceptual information in question. Movshon and colleagues pioneered a powerful method of this type in the early-to-mid-1990s. They trained monkeys to discriminate direction of motion in random dot stimuli (RDS). In such stimuli, dots move in random directions until the experimenter increases the amount of “correlated” movement—the number of cells that move in the same direction across each timeframe while the rest continue to move randomly. The monkey’s psychophysical sensitivity to certain amounts of correlated motion can then be measured by the number of times it guesses the direction correctly. Movshon and colleagues’ insight (Britten, Newsome, Shadlen, Celebrini, & Movshon, 1996) was that the responses of MT cells could be measured the same way to the same stimuli, and the neural and psychophysical performance compared. They first measured MT cells and showed that indeed, they increased their responses with increased correlated

---

[33] For a general discussion of difficulties in interpreting lesion data, see Young et al. (Young, Hilgetag, & Scannell, 2000).

motion in their RFs at their preferred direction. They then employed a technique from signal detection theory—receiver-operator characteristic (ROC) analysis—to correlate the neural response with the behavioral one. Essentially, what ROC analysis does is takes a signal and guesses the outcome of an event for which that signal provides information.

Britten et al. used ROC to show the correlations between MT neuron responses and monkeys' behavioral responses, the idea being that neurons' firing rate should be predictive of responses indicating the cells' preferred directions. They found that, on average, individual cells had a "choice probability" of about .56; that is, from increased firing in the cell, it was (again, on average) 56% likely that the monkey would make a decision in the cell's preferred direction. While this may not seem like a lot, the result was statistically significant, and impressive given the noise present in neural systems, and the fact that ROC stimuli must be near-threshold to generate choice probabilities (since the animal must get some answers wrong to generate a distribution of correct versus incorrect answers, that can then be compared to the neural data). Moreover, it is likely that populations of individual cells contribute to the motion signal (see below). Thus, to get such a significant predictive effect of behavior from individual cells was an important result. While this is not a direct establishment of use, it does suggest that positing a "decision unit" *somewhere* within the monkey, making a decision about direction of movement that leads directly to the behavioral response, could explain the monkey's choice patterns based purely on the motion signal in the MT cell. Moreover, since the RDS stimulus contains nonconnected parts of the visual field moving in a concerted direction, it is inherently a "pattern"-type motion stimulus, thus establishing MT's relevance for pattern motion.

Researchers have further sought to find the particular areas in which decisions of about motion are computed on the basis of MT signals, as well as where MT motion signals are used to compute other features. Unsurprisingly, the singled out areas, MST and the parietal cortex, are those that are at a higher level of the hierarchy than MT, and receive feedforward inputs from it. Physiological work in MST showed that MST responds to more complex motion stimuli than MT, including expansion, rotation, and shear (parallax), which are "non-linear" motions. Thus, the pattern motion in MT can be seen as providing input for more complex processes of motion analysis (Maunsell & Van Essen, 1987). Shadlen et al. (Shadlen, Britten, Newsome, & Movshon, 1996) focused on the lateral intraparietal region (LIP) as a potential decision area. LIP receives input from both MT and MST and projects to the frontal eye fields and superior colliculus, areas involved in the generation of eye movements. In a modification of the ROC studies just discussed, they had monkeys perform a saccade to a specific area to indicate the direction of motion in the stimulus. Certain LIP cells began to respond when the saccadic targets were presented, and continued until the saccade was performed. However, they were neither purely sensory nor purely motor in nature. They were not sensory, since they responded in a way that predicted a decision on saccade direction when monkeys were forced to make a choice for a stimulus with 0% coherence (i.e., when there is no directional motion cue present). They also were not implementing a purely "pre-motor" function, since their responses were *more* predictive when there was correlated motion in the stimulus. Movshon and Britten thus interpret LIP responses as implementing a decision about the direction of the stimulus based on input from MT and MST, and using that decision to drive motor behavior. So, there is evidence on multiple fronts for the use relation holding between MT and further cortical areas at higher levels of the hierarchy.

The results and interpretations discussed in this section show in part the validity of interpreting the MFH theory as a formalism. Reading the view as committed to feature-specific accounts of the parts of the visual system, plus the use relation, gives a good account for why the data has been interpreted in the ways that it standardly has been. It also shows why the MFH theory is committed to TA: the MFH theory's means of implementing the hierarchy formalism involves positing a specific type of feature-information as explicitly represented at each area. The area's function is then described in terms of the specific feature. In the final section of the chapter, I show some other benefits to reading the MFH theory as a formalism. For instance, it explains the *flexibility* of the theory, both in terms of its ability to tolerate internal disagreement and its ability to incorporate newly discovered evidence.

2.5. The Flexibility of the MFH Theory.

2.5.1. Internal Debates.

The massive consensus surrounding the MFH theory, which I have been at pains to articulate, should not be taken as a sign that there have not been serious, occasionally acrimonious debates about some aspects of the theory. In addition to debates about classification and nomenclature (see, for instance, the debate about using ‘MT’ versus ‘V5’ to refer to the middle temporal motion area, in footnote 9), proponents of the view have disagreed about the evidential status of particular methodologies (e.g., lesion results) and a variety of other issues. Viewing the theory as a formalism makes sense of the ways in which these debates are “internal” to the theory: proponents of a formalism can disagree about details of how the system-divisions and in-virtue-relations of a system are implemented, and about what constitutes good evidence for those implementations, without disagreeing about the nature of the formalism itself.

A seemingly weightier difference between MFH proponents is over the nature of feature representations themselves. Van Essen, in a number of papers, argues against the “detector” view of feature representations, on which a feature representing unit sends a precise signal to other areas of the visual system. He lists Barlow (1972) among others (e.g., Marr & Ullman, 1981), as those who have thought in feature-detection terms, citing Barlow’s “single-neuron doctrine.” The main reason Van Essen gives for rejecting the feature-detector view is the *breadth* of neural responses (Van Essen, Anderson, & Felleman, 1992). Neurons do not tend to respond in an all or nothing fashion to *only* their preferred stimulus. As discussed above, cells generally exhibits tuning curves, with their maximal responses to one particular stimulus (e.g., a particular direction of motion) and somewhat lesser responses to similar stimuli (e.g., slightly different directions). Therefore it is unlikely that there is a single, distinct code that is transmitted from some neural area, for instance MT, to its outputs in each case of, for instance, motion in a particular direction. Van Essen takes such considerations as motivation for adopting a "filtering" (Van Essen & Gallant, 1994) view of the responses of individual cells. Unlike a precise detection signal, a filter simply allows certain information to pass through. Filters are thus compatible with the signal differing slightly in each instance.

While there are important differences between the detection and filtering views of representation, the outcome of the debate does not affect the truth or falsity of the MFH view. Since the core commitments of the view are to feature-specificity and the use relation, the details of how those properties are instantiated can vary. So, what is important is not whether a feature is filtered or detected, but *that* it is a specific feature that is filtered or detected. So long as the representation is feature-specific, and further stages of representation (e.g., successive filtering) use that specific output in fulfilling their function, how the "representation" is construed is purely an internal question. The formalism reading can therefore make sense of how occasionally serious and important debates can occur between proponents of the same theoretical framework.

2.5.2. Extensions of the MFH theory.

I will now discuss two important developments in neuroscience that have proven to fall relatively easily into the MFH framework. The first is the importance of population coding—the idea that the relevant signals for neural processing are generated not by one, but by a group of neurons. There are a variety of motivations for positing population codes. One, which I've already mentioned above, is the tendency for neurons' responses to be best characterized by tuning curves. If tuning curves characterize cells' responses generally, then it is difficult to make fine grained discriminations about stimuli from single cells' responses; from a group of responses, however, the specific stimulus value can be calculated. In simplistic terms, each cell can be seen as casting a vote for their primary stimulus, where the outcome is determined democratically. The eventual representation is determined via the distribution of responses across cells with different tuning curves. Population coding is also important since population codes can be robust despite interruption—individual cells could be removed from a population code without dramatically lowering the discriminatory ability of the population (for reviews of population coding, see Parker & Newsome, 1998; Sanger, 2003).

It is easy to see, offhand, how population coding can be compatible with the MFH formalism. Since the MFH theory's categories involve areas producing specific feature representations, a population coding view will be compatible with the MFH theory as long as *what the population represents* is a specific feature. Population coding studies of MT generally model MT population codes as implementing representations of motion (Purushothaman & Bradley, 2004). The idea is to understand *how* specific motion representations are processed in MT, not whether. It is also important to note that the compatibility of the MFH view and population coding admits some variation in the details of a theory of population coding. There are at least three parameters that can be varied in modeling a population code, and they trade off with each other in interesting ways: the number of neurons modeled, the range of tuning curves included in the pool (e.g., all tuning curves, or only ones that are relevant to the current stimulus), and the amount of general correlation in firing between the cells included in the population. It is possible to view every cell in a population as contributing independently to a subsequent decision level, such that the responses of every unit are linearly summed by cells in the output area. However, with increased correlation between cells, there is little benefit to having all of the cells involved, since signal to noise ratio does not increase beyond a certain point. Shadlen et al. (1996) suggest that, in MT, a population of 100 weakly correlated cells could account for monkeys' behavioral responses to motion signals in the paradigm introduced by Britten et al. (1996). Another scheme, proposed by Purushothaman and Bradley (2004) for a fine-discrimination motion task, suggests that only cells with tuning curves close to the perceived direction and the opposite direction are informative about the monkey's eventual decision. The important point here is that which of these models is correct doesn't matter for whether MT can be characterized in MFH terms, so long as a motion-specific signal is the output of the population encoding mechanism.

A second major development that the MFH theory has proven capable of accommodating is the increased focus on top-down effects in the visual system. The conjunctional logic of the MFH hierarchy, as exhibited in Hubel and Wiesel and carried on by such figures as Van Essen and Zeki, is essentially bottom-up, with simpler features being transformed to extract more complex features (and eventually objects) at higher levels. However, as discussed in the introduction, the idea of a processing hierarchy has a significant top-down component. The distinction there was between *information* being carried upward through the hierarchy and *command* signals being conveyed downwards. The MFH version of this distinction consists in a bottom-up parsing of the scene into initial feature and object interpretations, and the subsequent *modulation* of those responses by top-down feedback. Lamme and Roelfsema (2000) review a range of studies to argue that the initial "feed-forward sweep" through the visual system, which contains only bottom-up information flow, is complete within roughly 100 ms from stimulus onset. However, they suggest that that time is sufficient for feature-dedicated areas of the hierarchy, as well as object-selective areas (e.g., for faces), to encode initial values of the features or objects that they explicitly represent. Then further processing can modify these representations in a number of ways. For instance, if a represented feature is part of a bound figure (e.g., a particular orientation representation in V1 representing part of a square) the firing rate of the neuron(s) representing that feature might be increased. This does not *change* the feature being represented, but merely modifies a previously existing feature representation to reflect its increased salience, a clear example of the modulation strategy discussed in section 2.4.1.

Attention has been a flagship example of a top-down effect that can be explained via modulation. In MT, Treue and Martinez Trujillo (1999) showed that both *spatial* attention (attention to the area within a cell's spatial RF) and *feature-based* attention (attention to the cell's preferred direction of motion) increased the firing rates of MT cells in an additive way—firing rates increased on average 10% for spatial attention alone, 5% for featural attention alone, and 15% for both. Cells selective for other directions of motion showed *decreased* responses in the attention conditions (this shows that modulation can cause both increased and decreased activation). Treue and Maunsell (1999) showed that, for both MT and MST, varying the type and difficulty of the task could substantially increase the percentage of firing rate increase. Such "gain" models of attention, while not the only extant type of model, show clearly how a modulation strategy can account for top-down influence in a way compatible with feature-specific representation. Once again, there is room for disagreement within gain models that do not affect this compatibility—for instance, Treue and Martinez Trujillo find, contra some other studies (Spitzer, Desimone, & Moran, 1988), that increased gain does not accompany a sharpening of the tuning curve (i.e., a narrowing of the range of stimuli to which the cell responds), but instead shifts the entire tuning curve upward in response rate. Which of these details is correct, of course, is irrelevant to whether it is specific feature representations that are being modified.

2.6. Potential Arguments against the MFH Theory.

A few examples exist in the literature of theorists offering evidence that they take to be in strong contradiction to the MFH theory (interestingly, these occasionally include former proponents of the theory). Importantly, the arguments involve either questioning the feature-specificity of specific areas, or stressing contextual change in responses. The fact that they are taken as arguments against the standard MFH theory is thus further support for reading that theory as an implementation of TA. In my view, none of them conclusively establish the falsity of the MFH theory—there remain strategies open to a defender of the MFH theory to explain the data. I do not claim that these instances are not in tension with the MFH hierarchy, or that the defense strategies could not be argued against. It is helpful to go through a potential defense, however, to make the MFH case as strong as possible, and to thereby define what kinds of results are genuinely in contradiction of it. In chapter 3, I discuss what I take to be one such case for area MT. If a definitive case can be made to "break the model" of the MFH theory, then the defense strategies offered here will seem much less appealing, thus opening up a motivation for a new view of function that more directly captures the kinds of evidence discussed here.

As laid out in chapter 1 and above, the MFH version of TA is inherently anti-contextualist. Albright and Stoner (2002) offer an impressive range of results regarding "non-classical" influences on visual cells' response properties to suggest that those responses are context-sensitive, and in contrast with general "elementarism" about perception. Elementarism is the view that perceptual responses are to specific elements of the retinal image, which for our purposes is close enough to the feature-specific absolutism of the MFH theory. While the spirit of their alternative "controlled contextualist approach" (Albright & Stoner, 2002, p. 344) is very much in line with the view I will propose in chapter 5, many of the cases they discuss fail to evade the strategies I have talked about for saving feature specificity. I will focus on two of their examples here—border ownership and completion phenomena. Both of these are instances of "spatial context" in which what is going on in the visual field outside of cells' spatial RFs can affect how those cells respond.

Determining border ownership involves the assignment of particular features to particular objects, especially where those objects border the background. Border ownership is important for figure binding, which is in turn important for separating figure from ground. Zhou et al.'s (Zhou, Friedman, & Von Der Heydt, 2000) studies of border ownership in V1, V2, and V4 show that a cell's response to its preferred orientation can vary significantly depending on whether the stimulus is on the border of a figure and the background, and to which it belongs. Certain cells, for instance, preferred the right-hand border of an object, while others preferred particular contrast patterns at the border of an object. Finally, others responded strongly to a border which occluded another object. Albright and Stoner take this to show that influences outside of the cells standard receptive field can change their responses. Ironically, this evidence is similar to the grouping evidence taken by Lamme and Roelfsema as entirely compatible with the MFH hierarchy. Lamme and Roelfsema's interpretation is based on a modulation strategy, specifically top-down modulation. It should be noted, however, that other mechanisms besides top-down ones can implement modulation.[34] The important point is that modulation, *as such*, doesn't undermine the feature-specificity inherent in the MFH view. So long as an observed modulation modifies a feature-specific representation, then it is possible for an MFH theorist to claim that the modulation is compatible with the standard hierarchy. Since in Zhou et

---

[34] There are a variety of theories about how interactions between cells in the same area or hierarchical level can cause their respective feature representations to be bound into a perception of a single object. One historically popular, but currently debated view is the "synchrony" hypothesis, where cells fire in synchrony when their explicit representations are part of the same object (Singer & Gray, 1995). This is an example modulation relation (cells go from firing non-synchronously to firing in phase) that does not change what each cell or group of cells represents.

al.'s results, the type of feature representation is not changed in the distinct conditions, (i.e., it is object form in each case), this type of response seems plausible.

The second type of evidence that can possibly be accounted for within the MFH theory is completion phenomena. There are many varieties of completion phenomena, but for expositional purposes I will focus here on only one: modal completion, for instance via illusory contours. In the Kanizsa triangle, seen in Figure 2.5 below, the triangle perceived in the foreground (i.e., as overtop of the circles at the corners) is perceived as having complete sides (the parts between the circles), despite there being no stimulation in the visual field at those points.

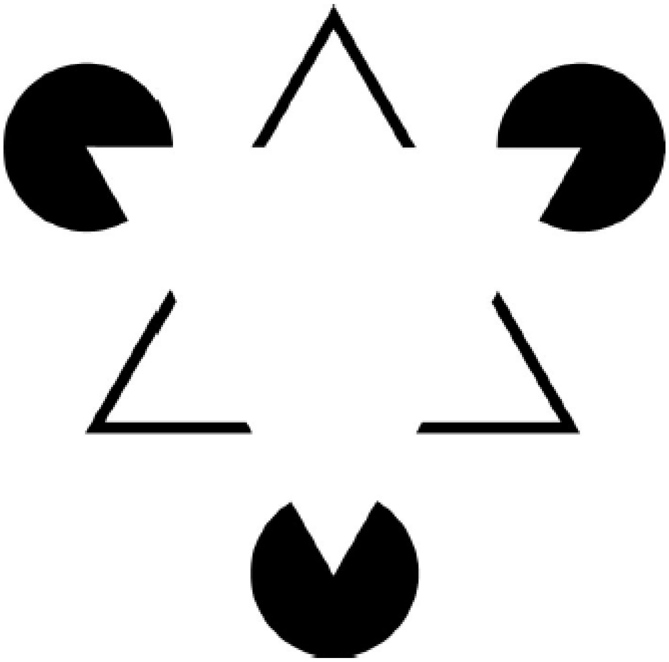

Figure 2.5. The Kanizsa triangle. From Stoner and Albright (2002).

In such illusory contexts, cells in V1 with receptive fields in those areas reliably respond as though the edge were actually present, suggesting that they do not in fact need to be stimulated "from outside" in order to represent a visual feature.

However, Lee and Nguyen (2001) found that in such conditions responses in V1 lagged behind responses in V2 by roughly 25 ms, which, as Albright and Stoner note, suggests feedback from V2 as a cause of the illusory contour. Feedback of this sort, however, is amenable to a modulation interpretation. The fact that a higher-level perception of the stimulus can cause a specific lower-level activation does not mean that the lower-level activation is not feature specific. The modulation strategy tells the following story about modal completion in the Kanizsa triangle. Certain cues from V1 (the arrangement of circles, and the relative orientation of the pie-slices cut out of them) are *used* by V2 to construct a representation of a bound triangle shape. V2 then completes the V1 representation by activating the edge-representing V1 cells as though they had been stimulated from outside. Such a view fits well with the processing hierarchy, and doesn't undermine either feature specificity or the use relation.

A final example of evidence that stresses but does not break the MFH theory comes, ironically, from two of its progenitors: Van Essen and Felleman, the two authors in the seminal 1991 paper. Hegdé and Van Essen (2007) constructed a wide variety of shape stimuli, grouped into simple gratings (varying in frequency and orientation) and a range of "contour" stimuli—complex shapes varying from bars bent at a variety of angles to spiral, radial, and star-shaped patterns. They then scaled the stimuli to match the RF sizes of V1, V2, and V4, respectively, and compared the responses of cells in those areas for each type of stimulus. Startlingly, they did not find that a greater number of cells in V4 (versus V1 or V2) were selective for more complex stimuli. Hegdé and Van Essen take this result as evidence against the conjunctive logic in which simpler features are conjoined to produce more complex features at higher levels, and Hegdé and Felleman (2007) take the results as showing that the three areas play “non-hierarchical roles in the processing of … shape stimuli” (2007, p. 418).

While the evidence is indeed in tension with the conjunctive logic, we have already seen that the input strategy can deal with some features having affects at multiple distinct levels. In Zeki’s study of V4, for instance, there were some opponent-wavelength cells found in V4 that had similar response properties (albeit with larger receptive fields) to V1 cells. What was important, however, was that *some* property, namely natural colors, was represented in V4 that had not been represented previously. While it is not decisive, a very sophisticated version of the input and output strategies might be applied to Hegdé and Van Essen’s results to analogize it to the Zeki perspective on natural colors in V4. Hegdé and Van Essen *did* find a difference in the population responses between areas, using multi-dimensional scaling (MDS). Briefly, MDS starts from a correlation matrix listing all of the correlations between cells’ responses to each particular stimulus, constructs dimensions that capture covariation between responses to different stimuli, and then re-plots the population’s response to each stimulus along those dimensions. Clusters can then be determined to show when the population’s responses differentiated between different stimuli and when they didn’t. Details aside, what is relevant is that the V1 population responses were not different for stimuli of varying degrees of complexity, whereas clusters separating simple from complex stimuli began to emerge at V2, and became pronounced at V4. The suggestion, then, is that while individual cells are selective for complex stimuli at V1, detailed population-level responses that distinctly represent those stimuli emerge only at V4. Since I've already discussed the ways in which population coding can be compatible with the MFH theory, a defender of the view could lean on the difference between population responses to suggest that complex stimuli are only explicitly represented at the V4 level.

This is, of course, not decisive. What I have tried to stress is the impressive flexibility the MFH theory possesses in dealing with new evidence and potential arguments against it.

2.7. Conclusion.

It would be hard to overstate the influence that the MFH theory, and hierarchy-based thinking in general, has had in neuroscience and psychology writ large. Theories of object recognition in computational psychology of vision rely heavily on the idea of feature specificity and hierarchical processing (Riesenhuber & Poggio, 1999; Ullman, 2007). Other sensory systems, including olfaction and audition, have been posited to exhibit processing hierarchies (Savic, Gulyas, Larsson, & Roland, 2000; Wessinger et al., 2001) similar in structure to the one posited by the MFH theory. Models of action control in cognitive neuroscience are predominantly based on hierarchies (Grafton & de C Hamilton, 2007; Haggard, 2005; Pacherie, 2008), extending the view that motor cortex is itself hierarchically organized (hierarchical views of the motor cortex are reviewed and criticized in Graziano & Aflalo, 2007).

Since the MFH theory is in many ways the progenitor of these perspectives, analyzing its absolutist commitments is a good way to test absolutist interpretations in neuroscience write large.

I have attempted to construct the strongest case possible for the MFH theory, both by explaining its successes and showing how it can hope to respond successfully to seemingly countervailing evidence. In doing so, I have relied at every step on the notion that the theory implements a formalism, and that feature-specific representation and the use relation are definitive commitments of the view. I thus conclude that the formalism both individuates the theory and captures its deep commitments. It is also clear what needs to be shown to genuinely undermine the MFH view—one must produce evidence that representations at particular areas are not feature-specific, or that the use relation does not obtain in the system. In the next chapter, I offer what I take to be the best arguments against the MFH theory in the case of MT—namely, its detailed responses to color and depth.

Chapter 3. Evidence against the Absolutist Reading of MT

3.1. Introduction.

In the previous chapter I articulated the core commitments of the modular hierarchy version of TA. In order for the hierarchy view to be true, functional areas in the visual cortex must process specific types of information explicitly, in virtue of inputs from earlier areas. This information must be feature-specific. Moreover, the information must then be used by areas higher in the hierarchy to perform their own functions.

Thus, the kind of evidence needed to argue against the MFH theory would either show that feature-specificity is false, or, minimally, that feature-specificity and the use relation cannot both be maintained.   Showing feature-specificity to be false involves showing that a brain area is responsive to several distinct types of information, where none of the input, output, modulation, or differentiation strategies are available. If such cases can be found, then the MFH theory is not empirically adequate—it cannot account for all of the evidence while retaining its absolutist commitment to function ascriptions that do not change with context. According to the characterization of ‘context’ I gave in chapter 1, contexts are particular instantiations of sets of explanatorily relevant properties. Showing that absolutism is false involves showing that context matters for what function a part of the brain is performing. The evidence I will evince in this chapter, I argue, suggests that contexts—the particular combined instantiations of perceptible properties in the environment, including their values and the relations between them—is partially determinative of what function

MT plays. What I will focus on is how different manipulations of the perceptual stimulus being studied—i.e., different values and combinations of the perceptible properties used in investigating MT responses—have revealed different variations in MT function. Since perceptual stimuli, as used in the studies, model different values and relationships between the perceptible properties that comprise particular contexts, this is strong prima facie evidence that the function of MT varies depending on context. Then in chapter 5 I will argue that a contextualist account can explain these variations in a way compatible with the desiderata on localization.

Recall the way in which feature-specificity and the use relation work together on the MFH view. The in-virtue-of relation for information processing goes bottom-up. Higher levels of the hierarchy perform their functions in virtue of information at lower levels. Top-down modulation influences are just that—as we saw with the example of perceptual filling in, higher levels of the hierarchy can change the specifics of a representation, or the situation in which it occurs, but do not *determine* the functions of lower levels, which are fixed independently. As such, feature-specificity must be specified "bottom-up," according to a privileged informational relation to the environment. At least, this much the MFH theory is committed to. Showing that the feature-specificity aspect of the view is false, then, involves undermining the notion of a single privileged informational relationship to the world. If this aspect of the MFH view is false, then the use-relation is also called into question, or at least complicated, since later areas cannot perform their functions in virtue of explicit, feature-specific information coming from earlier ones.

In this chapter, I introduce evidence that area MT responds to diverse kinds of information, depending on the perceptual circumstances. The first case study I undertake is of MT and color. The traditional MFH view is that MT is “color-blind” (Livingstone & Hubel, 1988). More recent advances, however, have shown that in specific contexts color information can have strong influence on how MT responds to a motion stimulus. The qualification that the stimulus must be moving is important, as it opens the way for a sophisticated application of a combined input and modulation strategy to maintain feature-specificity. While I will not argue here that the strategy can ultimately be overcome in this case, undertaking the case study anyway serves three purposes. First, it shows that the combined input and modulation strategy is the strongest strategy on behalf of the MFH theory, but that it must be applied *very* widely—not just to more standardly accepted information relevant to motion (i.e., luminance or disparity), but also to information from the color pathway. Second, and relatedly, it shows that evidence that cannot be accounted for via this kind of strategy would provide the strongest case against the MFH view. Third, the reasoning undertaken in determining color influence on MT is highly illustrative, and exploring it in detail provides a framework that I will expand on in developing a contextualist view of function in chapter 5.

The second case study involves MT and responses to depth. That MT responds to binocular disparity—one important cue for depth—has been known since the classic Maunsell and Van Essen studies discussed in the last chapter. Generally this was seen as modulating aspects of the motion signal. However, over the last 15 years investigations have uncovered ever more sophisticated responses to depth in MT, including, importantly, the ability to represent depth in the absence of motion. In light of this evidence, the field has come to recognize that MT's function is more diverse than that of representing motion. I argue that once the input and modulation strategies are unavailable, none of the other strategies are sufficiently motivated to preserve feature-specificity in the overall account of MT function. I then look at a variety of ways of attempting to save feature-specificity, and argue that they amount to giving up the use relation. I suggest that the evidence is strong enough to abandon the absolutist commitment in this case.

3.2. Color and MT.

Recall from the last chapter that Zeki's famous studies of MT showed no influence of color on MT responses. A range of psychophysical results established, however, that it is possible to see moving colors, although generally the movement is perceived less strongly than for similar motion defined by luminance contrast, and occasionally the motion is judged as being slower than for luminance-defined stimuli (for review and citations, see Dobkins & Albright, 2004). Experiments in the early 1990s showed that MT cells show a similar pattern, with color-defined motion eliciting weaker responses from MT cells than luminance contrast-defined moving stimuli at similar directions and speeds (Dobkins & Albright, 1990; Gegenfurtner et al., 1994). There are a variety of ways that colored stimuli could affect MT cells, however, without showing that they really respond to color information. That is, the response to color might be due to some other element of the stimulus, rather than to information about what colors are present. Two of these possibilities are as follows. First, MT could in fact be responding to luminance cues in these stimuli, since color stimuli, of course, also have specific luminance values. If MT in fact responds to a luminance cue in this case, then the evidence does not challenge the MFH interpretation of MT, since the MFH theory already posits that MT responds to luminance-defined motion (Livingstone & Hubel, 1988) as part of the M-pathway. The second possibility is that MT might respond to a *border* between colors, rather than to color *identity*. The basic idea is that displaced color boundaries could be used at early stages to define the object that moves, after which (by the stage of MT processing) no specific information about color is processed.

Clearly, the way to answer the question of whether color information specifically influences MT response is to devise stimuli that independently manipulate luminance cues, color-defined boundaries, and cues defined by color identity, and determine which compel MT responses. A variety of studies with such stimuli were undertaken by Albright and Dobkins (for review, see their 2004). While some of their effects could be interpreted in terms of luminance-defined motion, eventually they discovered a context in which color identity did partially determine MT responses. The combined results, then, stress the importance of variations in perceptual stimuli (read, perceptual context) in reasoning about the functional properties of MT cells.

In the first study, Dobkins and Albright (1994) constructed a grating stimulus that put color-boundary-defined and color-identity-defined motion into conflict. Normally, when a multi-colored stimulus moves, the boundary between colors moves in the same direction as the colors themselves. This is shown in Figure 3.1 (panel A). In the figure, the dashed line shows the direction of motion defined by the color identity (in this case, red), and the solid line motion defined by the boundary between green and red, both of which move towards the right in a series of time steps (t1-t3). Figure 3.1 (panel B), however, shows the first experimental stimulus used by Dobkins and Albright, in which, in addition to the grating moving, the colors of each bar were also reversed in each step.

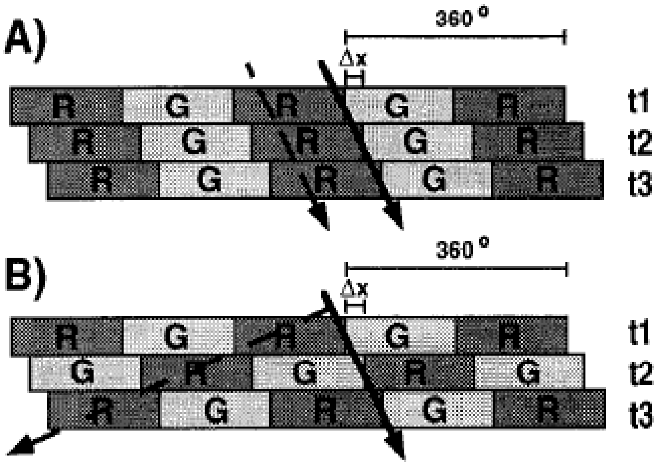

Figure 3.1. The stimulii from Dobkins and Albright’s (1994) study.

In the stimulus, the identity-defined movement is from right to left, as also indicated by a dashed line. There is still a border moving from left to right, however, as indicated by the solid line. This border has an advantage in the stimulus, in that *proximity* is important for motion correspondence—that is, all else being equal, motion perception will tend to follow the displacement that is closest at each time step (Dobkins & Albright, 1994). Within this stimulus, the amount of luminance contrast was varied across trials to show the relationship between color border, color identity, and luminance contrast.

MT cells responses were calculated via a type of measure we will see throughout the chapter, a "selectivity index" or "direction index" (DI). Recall that most cells have a background firing rate, which is modulated by incoming stimuli. A cell's preferred stimulus (for an MT cell, standardly, a particular direction and speed), is the one that increases it furthest from this baseline. A measure of the strength of the response of a cell to a particular stimulus, then, can be given by comparing how closely the response approximates its response to its primary preferred stimulus, versus an opposite, "null" stimulus. The selectivity index is this ratio. If a cell shows a significant DI to a stimulus, then it is reasonable to conclude that the stimulus strongly drives the cell.[35] Dobkins and Albright arbitrarily defined DIs in the direction of color-border motion as "positive" and those in the direction of identity-defined motion as "negative." In studying each cell, the stimulus in Figure 3.1 (Panel B) was alternated, so that at different times the boundary-defined and identity-defined were moving in the cell's preferred direction. DIs could then specifically be compared for each type of movement, to show to what extent that type of movement drove the cell.[36]

---

[35] The equation for DI employed by Dobkins and Albright is the ratio between two quantities: the cell's response to its preferred direction minus its response to its non-preferred direction over the preferred response plus the non-preferred response; DI=(P-NP)/(P+NP). Different studies employing selectivity indices use different equations depending on the stimulus being modeled.

[36] As Dobkins and Albright note, the fact that DIs are comparative between preferred and anti-preferred directions means that they are not measures of cell responses to a single stimulus. However, the *population's* combined DIs are an important measure of occurrent responses, since, for any given stimulus, the group of cells with positive DIs for that particular type of stimulus will respond. Dobkins and Albright thus take indices of this type as primarily a population measure, and I will assume this throughout.

The results of Dobkins and Albright's first experiment showed that near isoluminance (i.e., with the two colors the same luminance, and hence no luminance contrast), MT cells respond to the motion defined by the color border, as shown by positive DIs. This suggested that, absent a luminance cue, color border is what determines MT cell responses. When luminance contrast was changed, however, to reflect a contrast—i.e., where either red or green was significantly brighter than the other—this preference switched, shown by negative DIs in those conditions. What does this mean? First, it means that the color boundary is the preferred determinant of motion when boundary-defined and identity-defined moving edges are the only conflicting information in the stimulus—i.e., at isoluminance. This speaks strongly in favor of the color border interpretation of MT activation to moving colors. Second, it means that the color boundary response is easily overturned by information about luminance contrast. In the heightened luminance conditions, the color-identity border is *also* the luminance contrast border. Since the combined results can be accounted for with a combination of responses to luminance and color border, they suggest that color identity is not important for MT responses.

The results of Dobkins and Albright's second experiment, however, were more difficult to account for with either purely luminance or purely color-border responses. In this experiment, the stimulus matched the spatial displacements of the color identity-defined and the boundary-defined stimuli (whereas, in the stimulus shown in Figure 3.1 (Panel B), you'll notice that the spatial displacement was greater for identity-defined borders than boundary-defined ones). Thus, in the stimulus shown in

Figure 3.2, the boundary cues are inherently ambiguous in terms of direction of motion.

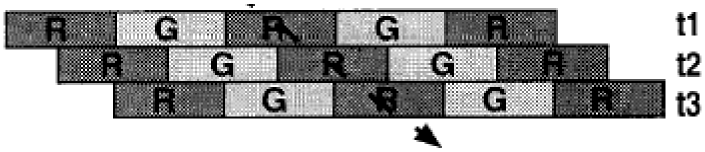

Figure 3.2. From Dobkins and Albright (1994).

At any time step a given border will be equidistant between two previous borders, and therefore it will be ambiguous which previous border it is moving towards, and which it is moving away from (and thus which direction it is moving in). In such a stimulus, any direction perception must be due either to identity-contrast or luminance contrast (as shown by the rightward arrow in Figure 3.2). Unsurprisingly, when a stimulus contains no luminance or color contrast, no motion is reflected in MT cell responses. However, when *only* identity-defined motion is present, cells still have significant DIs, although generally these are less than for luminance-defined. As such, MT cell responses recapitulate the psychophysical finding that perception of motion can occur for contrasts solely defined by color identity, but less strongly tha for luminance-defined motion.

The presence of an effect of color-identity on MT responses is highly unexpected from the standard MFH perspective, and Dobkins and Albright consider a variety of further mechanisms that might implement the response, including responses to color polarity in the stimulus at the level of the LGN, and the presence of a small number of color-selective M-pathway cells in earlier levels of the hierarchy. However, by this time it was already known that direct connections exist between ventral pathway areas, including V4, and MT, and Dobkins and Albright consider the possibility that these connections implement a direct influence of color identity information in the case of purely chromatic motion.[37] Dobkins, Stoner, and Albright (1998) further tested this possibility in a later study, which in fact showed that the MT responses to certain colored, moving stimuli are due to P-pathway input.

The reasoning of the study was based on the idea that any process sensitive to color identity must be sensitive to sameness and difference of colors in the stimulus. The stimuli in the study were also gratings, but now there were two of them. This setup allows for not only variation of the properties of the individual gratings, but also variation of the relationship between them. For instance, if one grating is passed over another, whether the gratings are perceived as one sliding above another, or as one unified plaid moving "coherently" depends on particular features of the two gratings. Specifically, Dobkins et al. varied the relationship between luminance and color contrast in the two gratings. In the "symmetric" stimuli, the luminance and color-defined contrasts matched in each grating—so, either the reds were bright in both gratings and the greens dark, or vice versa. In the "asymmetric" stimuli, the red of one grating was bright and the red of the other dark, while green showed the opposite pattern, or vice versa. In human subjects, Dobkins et al. showed that the symmetric condition stimuli are perceived as coherently moving—inherently a pattern motion response—and the asymmetric stimuli non-coherently moving, a component response. Interestingly, the distinction has upshot for the *direction* of movement as well, as

---

[37] Subsequent evidence has at least suggested the possibility that MT receives direct input from P-pathway cells in V1 (Nassi, Lyon, & Callaway, 2006). Where in the P-pathway color information comes from, while interesting, doesn't directly affect the discussion here.

shown in figure 3.3, below. If a stimulus is perceived as a coherent plaid, unidirectional movement is perceived in a direction in between that of the two individual movements.

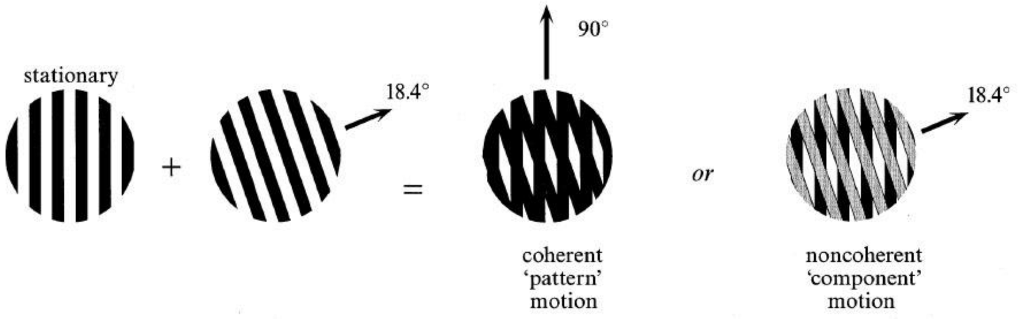

Figure 3.3. Component versus pattern motion in overlapping grating stimuli. From Dobkins, Stoner, and Albright (1998).

Dobkins and Albright compared MT cell responses in symmetric and asymmetric conditions. Within each condition, they further compared the responses to pattern versus component motion. This was done in a similar fashion to the studies discussed above, where the stimuli were alternated so that in some stimuli the pattern-defined motion was in the studied cell's preferred direction, and in some the component-defined motion was. The relevant measure was the ratio of mean pattern responses to mean component responses in each condition, referred to as P/C in Figure 3.4 below. Results above one indicated pattern motion preference in that condition, while results below one indicated component motion preference.

MT response was significantly modulated depending on whether the gratings were symmetric or asymmetric, in that cells reflected different P/C measures in the two conditions. Exemplary results from two cells are shown below in Figure 3.4. The cell shown on the left goes from having a strong preference for component motion in the asymmetric condition to reflecting much more influence of pattern motion in the symmetric condition. The cell on the right, with a less strong preference for component motion in the asymmetric condition, switches entirely (going from less than one to above one in the P/C ratio) in the symmetric condition, now preferring pattern to component motion. Significant modulations of this type were noticed in most of the cells studied.

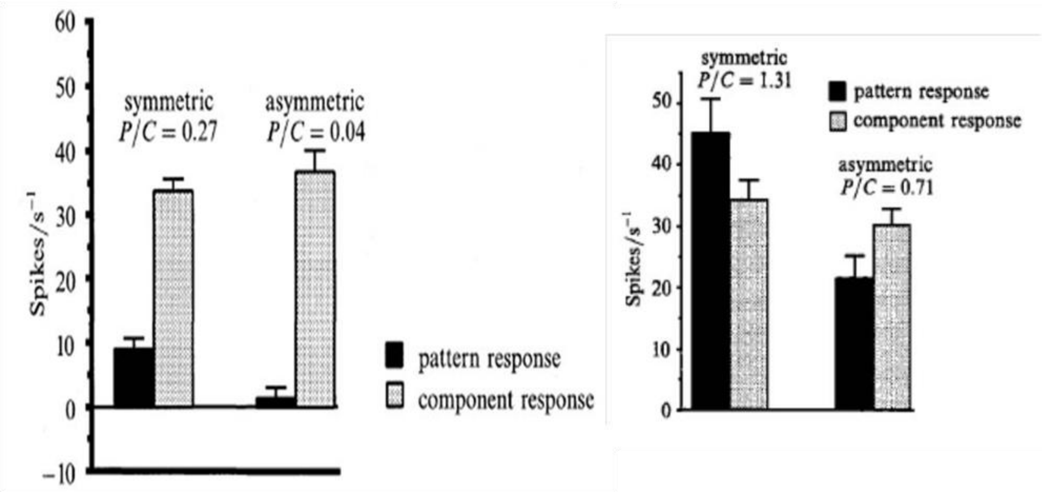

Figure 3.4. Results from two MT cells in the symmetric and asymmetric conditions. From Dobkins, Stoner, and Albright (1998).

The pattern motion interpretation of the stimulus relies on the fact the *same* color is bright in both gratings, while the component motion interpretation relies on the fact that two distinct colors are bright in the two gratings. Color identity information is solely a property of the P-pathway; earlier M-pathway neurons are not sensitive to it. So, unlike the alternative assumptions in the first study, there is no solely M-pathway interpretation of the effects. As such, MT responses are sensitive to information in the P-pathway. Dobkins et al. thus conclude that “the motion system has far more access to color information than previously believed” (1998, p. 689).

The discovery of the influence of color on MT follows roughly the same pattern as the establishment of motion influence—parameters of the stimulus are varied and the responses of the neurons are shown to be modulated based on the manipulation. Moreover, the results in MT match psychophysical results; color segmentation also affects pattern versus component perception in human subjects (Kooi, De Valois, Switkes, & Grosof, 1992). There is thus strong evidence that color information has a functional effect on MT processing. Can any of the strategies articulated in the last chapter for dealing with outside influence be applied here? The two major possibilities are a modulation strategy based on top-down influence, and a combined input/modulation strategy based on treating color as a cue for motion segmentation. I will now briefly argue against the top-down modulation strategy, before articulating why a very sophisticated version of the input/modulation strategy might work. This will set up the case in the next subsection, for which neither strategy is available.

The top-down modulation strategy suggests that some top-down process, such as attention, is responsible for the component-to-pattern shifts in the color-symmetry stimuli. There are a variety of studies that provide evidence against an attentional interpretation of color affects on MT, however. I will discuss only one here (see Thiele, Rezec, & Dobkins, 2002 for another example). Dobkins, Rezec, and Krekelberg (2007) tested possible roles for attention in a psychophysical study of chromatic motion in humans, based on the motion aftereffect. Aftereffects are due to sensory adaptation. After a stimulus is presented for a period of time, the neurons sensitive to that stimulus will decrease firing. Then, residual firing from neurons tuned for the opposite stimulus (e.g., red versus green, leftward v. rightward motion) will cause a faint perception of the opposite stimulus. Importantly, studies in monkeys have shown that MT is the neural basis of motion aftereffects (Kohn & Movshon, 2003). Dobkins et al. therefore reasoned that if chromatic motion (and MT responses) had to be mediated by attention, aftereffects for chromatic motion, but not achromatic motion, should only be present when the original motion stimulus is the prime focus of attention. As such, they had subjects observe a grating moving in one direction for sufficient time to produce adaptation, then a stationary grating. Due to the aftereffect, the second grating was perceived as moving in the opposite direction from the first. The subjects pressed a button when the aftereffect ended. The idea was that, if attention is vital for chromatic but not achromatic motion, then aftereffects for chromatic motion should be much more heavily modulated by attention (indeed, they should only exist at all in that condition).

They then compared chromatic and achromatic movement across "full attention" and "poor attention" conditions. In the full attention condition, the subjects attended to the motion stimulus. In the poor attention condition, they performed a vowel recognition task on lexical stimuli presented in a different part of the visual field. The relevant comparison was between the change in aftereffect duration from poor to full attention conditions between chromatic and achromatic stimuli. While duration did increase for both chromatic and achromatic stimuli in the full attention condition, there was no significant *difference* in the amount that they changed. If chromatic motion required attention, while achromatic did not, then the effect of attentional changes should be much greater on chromatic motion aftereffects. Given that there was an appreciable aftereffect in the poor attention condition for chromatic gratings (although it was less than for achromatic) and that attention had similar effects on each (determined by the ratio of duration between conditions), the conclusion is that chromatic motion does not require attentional influence any more than achromatic motion. And, while it requires an inference of several steps, it is not unreasonable to conclude that a significant influence of color information on MT occurs without top-down attention. The physiological results in MT show the effect of color on MT neurons, the usefulness of that information for perception, and the responsibility of MT neurons for motion aftereffects. Moreover, Dobkins et al. cite cross-adaptation effects between chromatic and achromatic movement as evidence that they share a common mechanism. Thus, independent evidence suggests that MT underlies such effects as the ones studied, and the psychophysical results here suggest that attention is not needed for those effects.[38]

The strongest possible strategy for maintaining feature specificity in MT, then, is a version of the combined input/modulation strategy. As we saw with saw in the last chapter, the input strategy can be stretched to account for a wide range of influence, so far as it is reasonable to conclude that the explicit signal present in an area is a feature-specific one. The idea is that other types of stimulation can modify the particular values of the explicitly represented feature, in this case motion. Van

---

[38] Of course, this study makes a relatively strong claim about the homology of motion perception in monkeys and humans. There is considerable support for the idea that the homology is very strong in the case of chromatic motion perception. See (Thiele, Dobkins, & Albright, 2001) for one discussion.

Essen and Maunsell, as discussed in the last chapter, used this strategy to argue that MT responses to disparity were not a *representation* of depth, but instead as a modulation of a motion signal. It is possible, in light of the chromatic motion results, to go one step farther still and say that in situations like the Dobkins et al. (1998) results, MT responses represent *motion* of particular color segmentations in the scene (e.g., grating versus plaid segmentations). On this view, color, as well as disparity cues, can serve as inputs which allow for the specification of motion, where the motion signal is what is represented explicitly. This view is compatible with even P-pathway inputs having an effect on MT. Consider that in the psychophysical studies, subjects are required to report direction of motion, duration of motion aftereffect, etc. They are thus compatible with the idea of color occasionally serving as a motion cue to a feature-specific motion area.

The reasoning here has analogues in philosophical discussions of representation. Consider what Fodor (1990) calls “asymmetric dependence.” Asymmetric dependence is the idea that out of two possible contents, X and Y, attributable to a representation, X is the content if representing Y counterfactually depends on representing X. Why, for instance, is MT known as a motion area and not a speed area or a direction area? It is because both these features are aspects of motion, and could not be represented unless motion was also represented. Similar reasoning could be applied to disparity and color. MT, the response goes, would never reflect information about these features in its responses if they were not specifying a motion stimulus in the scene. Since motion is the universal feature of MT stimuli, that is what it represents explicitly, with all other features serving as inputs or modulations.[39]

I will grant this reasoning for the present, so as to make the case for the MFH theory as strong as possible. However, there is reason to doubt the strategy in this case. Consider, the change in motion response by cells in Dobkins and Albright's second study is a *categorical* change in the kind of motion represented. That is, the motion response in MT as a whole (if the results generalize) switches from representing component to representing pattern motion depending on the *categorical* color information in the stimulus. Consider, then, a user of the MT signal trying to read off information from the MT response. Should the signal be read as a component signal or a pattern signal? It seems that it depends upon the color segmentation in the scene, and thus that MT signals cannot be read as motion signals *simplicter*. That is, in order for the signal to be useful as a particular motion signal (signaling component or pattern motion), information about color segmentation must also be included—otherwise, the putative decision maker would not "know" which decision to make (i.e., a component-direction or a pattern-direction decision). I will return to this kind of argument later in the chapter.

For now, I take the foregoing to have established that the combined modulation/input strategy is the strongest one available for defending feature specificity, and it relies on one feature—the explicitly represented one—being represented always as other parameters vary. Breaking the MFH model, then, requires

---

[39] This analogy is only meant to highlight the kind of reasoning involved here. Fodor's discussion has a variety of philosophical commitments and aims that are not necessarily reflected in the thinking of neuroscientists.

showing cases in which multiple features can be represented at an area independently. In the next section, I show that in the case of *depth* derived from disparity, the input strategy is not available, since depth is represented, in some cases, independently of motion. Once this is established, I argue, the motivation for maintaining feature specification writ large is considerably lessened.

3.3. Depth and MT.

Binocular disparity is an important cue for the perception of depth. Most simply, disparity is the difference in the angles at which individual points of light hit the two eyes, as shown in the diagram below (Figure 3.5).

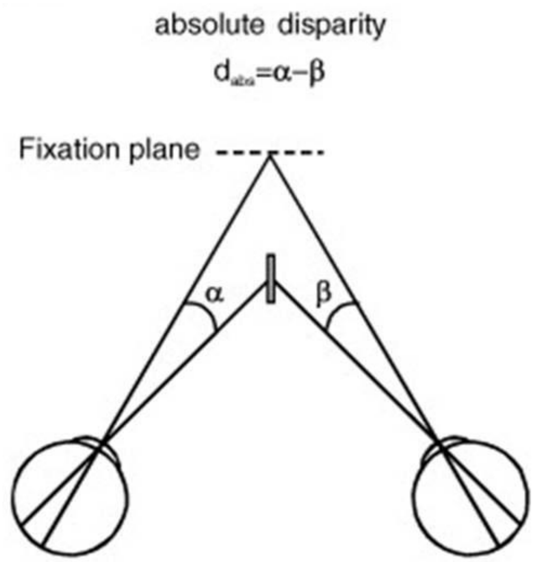

Figure 3.5. Absolute disparity. From Anzai and DeAngelis (2010).

The disparity is measured as the difference between the two angles at which light from one object reaches the eye, relative to the plane of fixation (the depth plane at which both eyes are focused).[40] Moving the object towards the perceiver would increase the angle, moving it away would decrease it. The above figure shows what is called “absolute” disparity. But more complex depth cues are available in the *difference* between the absolute disparities of two distinct objects, or two points on the same object. This is called “relative disparity,” and is shown below (Figure 3.6).

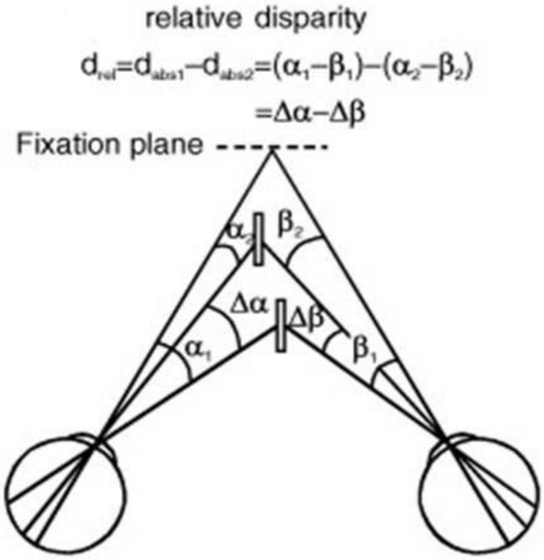

Figure 3.6. Relative disparity. From Anzai and DeAngelis (2010).

Standardly, the two kinds of disparity are taken as important for different kinds of depth perception. Absolute disparity is good for “coarse” depth perception, involving segmenting the scene into depth planes (compare the object in Figure 3.5 to the plane of fixation). Relative disparity is useful for more “fine” depth perceptions, involving, for instance, perceiving the shapes of objects.

---

[40] Perceiving depth via disparity involves solving the “correspondence problem,”—i.e., determining which two stimulated points at the two eyes are due to light from the same location in the visual field. The process of solving the correspondence problem begins at V1, where information from the two eyes is combined in single neurons for the first time. At V1, this information is subject to a number of ambiguities, so determining a unique solution to the correspondence problem may include processes in anterior areas (Anzai & DeAngelis, 2010). I will not discuss the correspondence problem in detail here.

Disparity can be manipulated in artificial stimuli, by presenting monocular images to each of the two eyes that are identical, except for points at which individual parts of the image are offset to the two eyes, creating an artificial disparity that is perceived as depth. Both kinds of disparity are important for depth perception, and both will be discussed in this section. As we saw in chapter 2, Maunsell and Van Essen (1983b) used an input/modulation strategy to relegate disparity responses to a subsidiary role in MT. In the last 15 years, however, continually more evidence has emerged that MT neurons show sophisticated and important responses to depth information in the stimulus. Interestingly, the recent history of the topic has continually reflected MFH biases—the tendency has been to downplay depth influence at MT, until forced by the evidence to admit a greater role for MT in depth perception. At each point, changes in the kinds of stimulus contexts investigated have led to new insights about MT function.

Interest in disparity responses in MT ramped up in the late 1990s with the studies of DeAngelis, Cumming, and Newsome (1998), who studied not only the disparity-sensitive *architecture* of MT, but also whether stimulation of disparity-selective cells could influence depth judgments. They showed drifting gratings to monkeys within the receptive field of particular MT neurons, and then varied the disparity over several presentations while keeping the motion signal constant. The experimental setup is shown below (Figure 3.7).

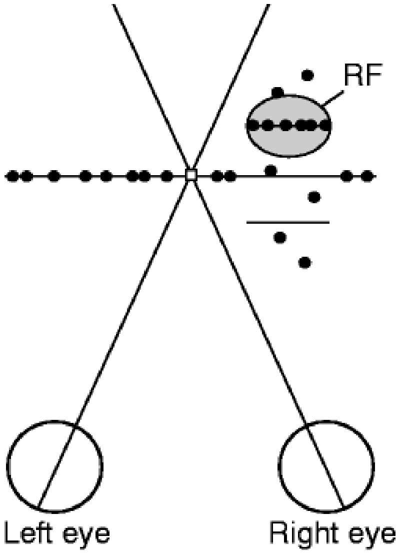

Figure 3.7. The experimental setup of DeAngelis et al. (1998).

Using multi-electrode recording, the researchers showed that many sites within MT had strong disparity tuning. They also explored the spatial layout of depth-selective cells. They showed that cells with similar disparity preferences were clustered together, and, importantly, that disparity preferences were arranged in columns across MT, providing a complete spatial map of disparity values in the visual field. This organization was in an important sense *independent* from the organization of direction-selective cells, in that there was no significant relationship between the depth-selective organization and the motion-selective organization. The finding is important for a number of reasons. First, it shows that disparity sensitivity in MT is organized in a potentially informative way—i.e., one potentially useful for perceiving depth. Second, it shows that the map of disparity is not tied in any obvious way to the map of direction, but instead that they are interspersed, and thus that disparity information might be useful independently of direction information. In support of this, a later study by Uka and DeAngelis (2003) showed that particular disparity selectivities in individual cells are not correlated with particular direction selectivities. That is, while disparity and motion maps are distributed across the same cells in MT, they are distinct and separable.

To test whether MT responses contribute to depth judgment, DeAngelis et al. constructed a task in which a disparity signal consisting of a number of dots at a consistent disparity was interspersed with “noise” dots at random disparities. This is a depth analogue of the RDS stimuli discussed in the last chapter, where it is levels of disparity and not of coherent motion that are manipulated. The stimulus was shown in the receptive field of a particular multi-unit recording, at the disparity (but also direction and speed) that the cells in that part of MT tended to prefer. The monkey had to indicate whether the depth of the disparity signal was near or far compared to the fixation plane, by making a saccade to a particular target for near or a different one for far. The area of the multi-electrode recording was then stimulated during the presentation of the stimulus. The results showed that monkey’s responses were significantly biased towards the disparity preferred by the MT area activated. Moreover, in the stimulation condition, less coherence amongst the disparity was needed to produce depth responses in stimulated cells’ preferred direction. Unlike Maunsell and Van Essen, then, DeAngelis et al. suggest that it is strongly possible that disparity information in MT is used to perceive depth.

DeAngelis et al. then performed one further variation that will be important in what is to follow—they tested whether similar tuning curves and stimulation biases occurred for *stationary* stimuli that varied in disparity. They showed that, for 9 of 12 tested disparity-selective sites, not only did the site retain activation for non-moving stimuli, but their tuning curves were the same. Moreover, when a version of the stimulation study was performed for stationary stimuli, similar perceptual biases were exhibited—the monkeys were more likely to make judgments at the preferred disparity of the stimulated site. This was the first evidence that MT cells could signal disparity in the absence of motion, and moreover that these signals could be used for depth judgment independently of any motion task. In a different study employing the choice-probability measure (see chapter 2), Cumming and Parker (2000) showed that similar correlations to perceptual depth judgment are not to be found in V1 responses, suggesting further that MT makes a distinctive contribution to depth perception.

The finding that information at MT cells about disparity was *used* for depth judgments, and moreover that it could be so *independently* of motion, is an extremely important finding for judging MT function with regards to TA. DeAngelis et al. put the matter succinctly: “Under natural viewing conditions, MT neurons may be active during viewing of stationary objects” (1998, p. 679), a far cry from what a TA posit of a motion-specific function would suggest. But the results remained suggestive at this stage. As Palanca and DeAngelis (2003) note, the microstimulation in the study may have caused a perception of both motion and depth, in which case it is unclear what information was causing the monkeys’ responses. have been used in the task. Palanca and DeAngelis set out, using single-cell recording, to examine in a fine-grained way how motion and disparity response properties were related in MT.

Palanca and DeAngelis (2003) employed a random dot stimulus in which disparity was added within the receptive field of MT cells while monkeys fixated at a point in the center of the screen. This is shown in Figure 3.8. The square in the center is the fixation point, and the circled area shows the disparity added to particular dots within a studied cell's RF. They then varied movement (both direction and speed) and disparity independently to compare tuning curves of MT cells—both in single unit and multi-unit recordings—to each, and in combination to show how the tuning properties affected each other.

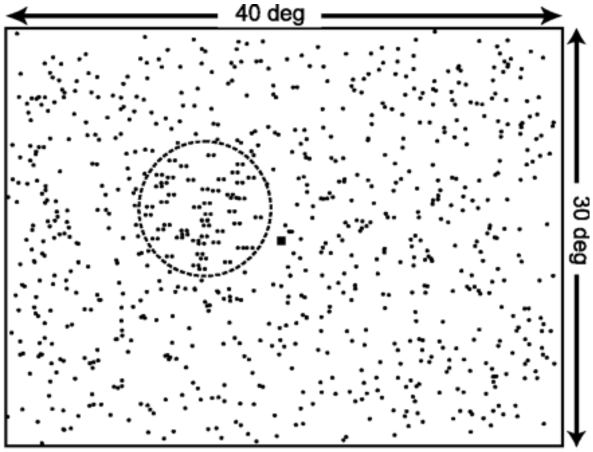

Figure 3.8. The stimulus from Palanca and DeAngelis (2003).

The following were among their important results. First, they showed that a significant percentage of MT neurons responded to disparity in the absence of motion. 34% of neurons responded at least 1/3 as strongly to stationary as to moving dots, and a significant subset of these *preferred*—responded more strongly to—stationary versus moving stimuli. Moreover, these cells' RFs covered all parts of the visual field (as the distribution of disparity-selective cells across MT found in the DeAngelis et al. study would suggest). Second, the tuning curves for disparity in MT cells that responded to both disparity and motion signals were similar whether the stimulus was moving or not, suggesting that while motion in the stimulus generally caused a stronger response, the cells' responses to disparity were qualitatively similar in either case. Palanca and DeAngelis also constructed a Disparity Tuning Index (DTI), similar to the direction indices discussed in the last section. Recall that these measure the ratio of responses to the preferred disparity versus the non-preferred disparity. The DTI was also not significantly different in moving versus non-moving conditions, suggesting that cells generally remain equally sensitive to disparity in either case.[41]

Finally, Palanca and DeAngelis analyzed whether responses to stationary versus moving dots varied in timescale. Measuring DTI (see note above) for both single unit and multiunit recordings that were sensitive to disparity, they found that cells' ability to discriminate their preferred disparities was just as strong in the motion versus stationary conditions, and importantly, exhibited high levels from early responses (around 70ms) to a few hundred milliseconds (although they did drop off at a second or more). Since saccades in non-experimental conditions occur every few hundred milliseconds, they conclude that "MT neurons can signal disparity for static scenes with almost as much fidelity as they do for scenes that contain moving objects" (2003, p. 7652). They thus conclude that MT contains signals that can be used to perceive depth in the absence of motion, and suggest that their study provides external

---

[41] A related measure, the Disparity Discrimination Index (DDI), which measures how well the neuron could discriminate between preferred and anti-preferred disparities across trials, was lessened in the motion condition. The authors attribute this to increased variability in the responses in the motion condition.

support for a depth-perception interpretation of the microstimulation studies by DeAngelis et al.

The combined results discussed suggest that disparity has a dissociable affect on MT responses. Many cells respond to disparity independently of motion, and there is at least strong evidence that these signals contribute to depth perception. Moreover, even for cells that do respond to both stimuli, the fact that neurons' tuning curves for depth don't vary with motion suggests that the two cues may make separate, dissociable contributions to MT responses. As Palanca and DeAngelis note, "the striking examples of MT units … that prefer stationary stimuli appear to simply reflect a continuum of speed preferences that includes zero speed" (2003, p. 7655). These properties will be important below, where I argue against the applicability of the strategies for defending feature-specificity. After the Palanca and DeAngelis results, it was generally accepted that MT responses contain signals relevant to depth, and do so relatively independently of motion.

Around the same time, several studies also further explored how motion and depth information in MT responses relate, and how depth information is used. I will only mention these briefly here. Dodd, Krug, Cumming, and Parker (2001) and Grunewald, Bradley, and Andersen (2002) measured MT neuron responses to random dot stimuli that involved both motion and depth information. Their studies were structurally similar to those of Britten et al. (Britten et al., 1996; discussed in chapter 2), in which both psychometric and neurometric functions were calculated, and MT neurons were studied to see whether they could predict behavioral response. Dodd et al.'s stimulus involved a rotating transparent cylinder, where disparity information between the front and back surfaces was varied. At every point of the figure, motion cues existed in both directions. With no disparity information, it is ambiguous which surface is in front, and therefore in which direction the overall cylinder is rotating. Perceptually, in both humans and monkeys this produces a bi-stable figure, in which the cylinder is alternately seen as rotating in one direction then another. Their stimulus is shown below.

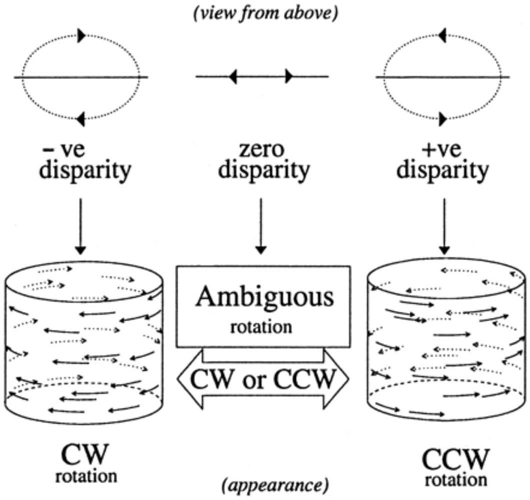

Figure 3.9. The bi-stable cylinder stimulus of Dodd et al. (2001).

Dodd et al. showed not only that the responses of MT neurons to the figure are modified when disparity is added, but that these modified responses are relevant to perceptual judgments. Monkeys had to judge in which direction the *front* surface of the cylinder was moving, and choices of a particular direction increased as a function of the amount of disparity cue added. So, if the 'front' face was moving clockwise, the amount of clockwise choices increased systematically with the amount of positive disparity. Similarly, firing rates from MT cells increased monotonically with the amount of disparity added, despite the fact that only relatively small range of disparities were tested (recall from chapter 2 that, in ROC-style studies, the range of stimuli must be one in which behavioral success is neither perfect nor at chance). This is important, since it showed that the cells conveyed information about the amount of disparity in the stimulus, and did not just respond to the direction of motion disambiguated by the added disparity. In a trial-by-trial analysis, the responses of individual neurons to depth also predicted the monkey's behavioral decision. Moreover, along a variety of analytical parameters these predictions were more significant than in other studies—a greater percentage of neurons had a significant chance of predicting the outcome, and the prediction values were greater—than in pure motion studies. In a similar study performed around the same time, Grunewald et al. (2002) compared the effects of MT disparity sensitivity on perception to those in V1, and found significantly greater CPs in MT.[42]

Uka and DeAngelis (2003) using a stimulus similar to that of DeAngelis et al. (1998), also performed choice probability measurements, but varied the amount of signal versus noise dots in the stimulus. They found that, not only were behavioral responses indicating depth judgments highly predictable from MT cells' disparity preferences, but that this was true regardless of how "strong" the signal was and how

---

[42] Recall that, in chapter 2, fn. 17, I discussed how Maunsell and Van Essen found the disparity tuning curves of MT cells to be mostly similar to disparity-tuned cells in V1. I noted there that, while Maunsell and Van Essen did not make the argument, it could be argued that disparity information was not explicit at V1 due to the lack of difference in their tuning curves (i.e., disparity would have failed the distinctness condition on explicit information). The Grunewald et al. results, along with those from Cumming and Parker (2000) discussed briefly above, show that, contra this potential strategy, information about disparity is used for depth perception in a way that is distinct from what is available at V1. As such, the depth signal more likely meets the conditions on explicitness discussed in chapter 2.

often the monkey got the depth judgment right. Since monkeys tended to get fewer trials correct in the low signal/noise conditions, this proves that MT is predictive of response regardless of success or signal strength, thus supporting that the claim that the MT signal is used for the depth judgment. Now, on their own, these studies' stimuli are confounded with motion in the stimulus. However, in connection with the results from DeAngelis et al. (1998) and Palanca and DeAngelis (2003), these studies suggest that not only is information present about depth in MT, but that it can be used for depth judgments.

Even given the impressive results regarding depth representation in MT, however, there was still the question of what *kind* of depth perception MT contributes to. A general consensus developed by the mid-2000s that MT largely responds to absolute depth, and that its signals are useful for depth ordering (or "coarse" depth perception), but not for relative disparity or fine depth discriminations (Anzai & DeAngelis, 2010; Born & Bradley, 2005). This is befitting MT's place in the dorsal stream, the traditional "where" pathway (Mishkin et al., 1983). As mentioned briefly in chapter 1, subsequent to the division of pathways established by Livingstone and Hubel, it was shown that V4 cells also have significant responses to disparity, and the consensus held that V4 cells' responses are to relative disparity, befitting its place in the "what" pathway (since relative disparity is useful for recognizing shapes and objects). Review articles from the mid-2000s to as late as 2012 (Roe et al., 2012) suggest a solely coarse-depth role for MT disparity responses, positing the ventral stream as the locus for fine depth perception. This was despite initial results from

Nguyenkim and DeAngelis (2003) which showed that MT neurons exhibited response to the fine-depth features of *tilt* (the position of the parts of an object relative to a horizontal axis at its midpoint) and slant (relative to a vertical axis), which was invariant and separable from their disparity tuning curves. Not only did MT cells have preferred tilts and slants which were not significantly correlated with particular motion preferences, but, for a small number of neurons tested, Nguyenkim and DeAngelis showed that these preferences remained even if there was no coherent motion in the stimulus (e.g., all motion was random). Still, the standard interpretation in the field was that, if MT had any responses to relative disparity at all, it plays a “reinforcing” role for motion responses (Born & Bradley, 2005, p. 175)—a clear use of the modulation strategy. Two more recent studies, however, have shown more robust and intricate responses to relative disparity in MT, suggesting a more substantial role for MT in fine depth perception.

Sanada, Nguyenkim, and DeAngelis (2012) compared MT cell responses to tilt as defined across multiple cues, including motion, disparity, and texture. Tilt can be defined by different disparities in the different parts of the object (e.g., greater disparity at the top of the object than at the bottom, or vice versa), by movement at different speeds in distinct parts of the object (e.g., in a rotating cylinder the closer parts of the surface will appear to rotate faster), and by texture. The stimuli exhibiting these distinct tilt cues are shown below; each panel of Figure 3.10 portrays tilt defined by a different feature, except for the lower right panel, which shows tilt determined by several features.

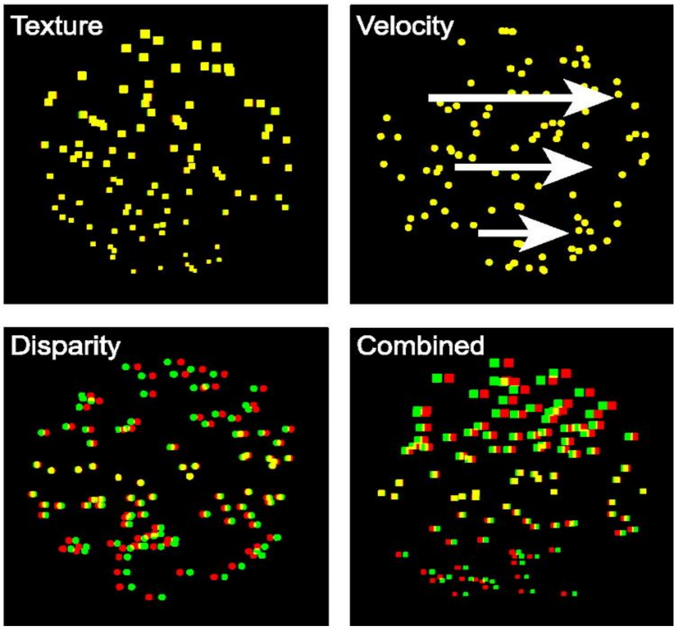

Figure 3.10. The tilt stimuli from Sanada et al. (2012).

Sanada et al. measured MT responses to each cue. In each case, the stimuli did move in the cells' preferred directions in this case. However, the direction itself provided no information about the tilt of the object—the only tilt cues came from the disparity, speed differences, or texture, respectively, and these were independently varied. The individual conditions were then compared to a "combined" condition, which included all three cues. They measured the preferred tilt angle of individual cells across conditions, as well as constructing a tilt discrimination index (TDI) to judge how informative the cells' responses were about their preferred tilt. The TDI measure compared the firing rates of cells to their preferred and non-preferred tilts. A significant TDI is inferred to show that a cell's response to its preferred tilt carries usable information about that tilt, as compared to its non-preferred ones.

Sanada et al. found a range of responses in MT neurons to the different manipulations. Some cells had very weak tilt preferences. Others had very strong ones. Importantly, many of the tilt tuning preferences were not the same across conditions. Many neurons had tilt preferences for tilt stimuli defined by motion and disparity, but not for texture. Amongst these, some had the same tilt preference for the two different cues and some did not.

At first, this seems a bewildering array of responses, but the authors applied a variety of population-level measures to make sense of them. Most cells (59%) showed significant tilt tuning (measured by TDI) across the velocity, disparity, and combined conditions, while smaller numbers had tilt tuning only in the combined and velocity (15%) or combined and disparity (10%) conditions. A much smaller percentage was selective only in the combined and texture conditions (2%). Thus, MT cells generally were tilt-selective, but their selectivity varied with the type of condition. Interestingly, while both disparity-only and velocity-only tilt selectivity were statistically correlated with the selectivity in the combined condition, the correlation between the two single cue conditions was considerably less strong. This result was due to the fact that some cells' tilt preferences were more strongly driven by disparity, some by velocity, and some were equally driven by both.

Sanada et al. fit a linear model to each cell's responses in the different conditions, which treated the response of the cell in the combined condition as the additive combination of the cell's responses to each of the individual conditions (the many cells that showed no effect of texture simply had a zero weighting for the texture parameter in the model). They showed that the linear model could account for virtually all of the data in the combined condition. Importantly, this suggests that MT, as a whole, contains a variety of depth-sensitive cells. Some do not respond to relative disparity (but do respond to absolute). Of the many cells that do respond to relative disparity, different ones weigh different cues differently, but all in a roughly linear way. To visualize this, see the figure below, which is a scatterplot of all of the cells that responded in the combined cue condition for relative depth. The fact that the upper right quadrant is most heavily populated shows that most cells have positive weightings for both kinds of cue.

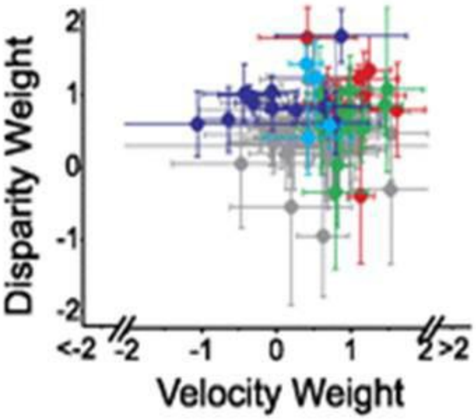

Figure 3.11. Weightings of disparity and velocity in tilt stimuli in MT cells. From Sanada et al. (2012).

The important conclusions from the study for current purposes are that many MT neurons are selective for tilt, an aspect of fine depth, that tilt interpretation is generally enhanced by the presence of multiple cues, and that individual neurons' responses for tilt are due to a combination of cues that is cell-specific. In their responses to relative depth, then, both disparity and motion are important, but neither one is primary across the population. Instead, individual cells are differentiated by how they weigh different cues. Moreover, the vast majority of cells contain responses to both disparity and velocity, and *both* contribute, albeit at different levels, in perceiving fine depth. Thus, information about multiple cues is present and functionally relevant in MT neurons, and the effects of those distinct cues are additive. I will discuss these properties more fully below.

The final study I will discuss tried explicitly to tease apart absolute from relative disparity preference in MT. Krug and Parker (2011) constructed a stimulus in which absolute and relative disparity could be varied independently, to see whether MT cells responded to particular relative disparities across changes in absolute disparity. Their stimulus-type is shown below.

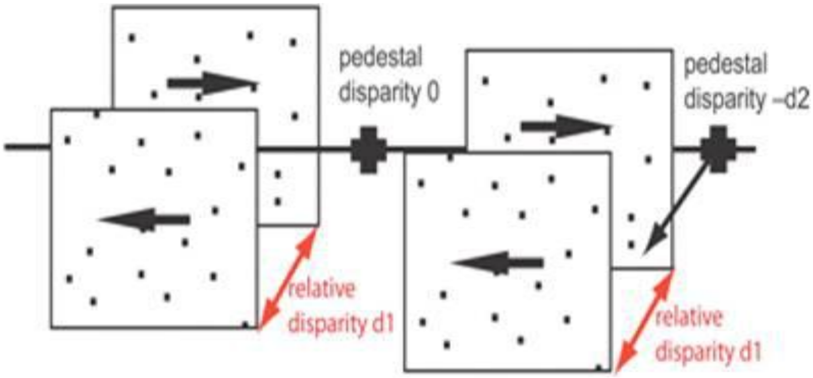

Figure 3.12. Krug and Parker's (2011) stimuli.

Two planes with different absolute disparities were superimposed on each other, creating a relative disparity between them. The key was to independently vary the *relationship* between the amount of this relative disparity and the backdrop, "pedestal" disparity. Varying the pedestal disparity meant, essentially, applying the same amount of absolute disparity change, in the same direction, to both planes, thus keeping their relative disparity the same. The authors first measured absolute disparity-selectivity by setting the pedestal disparity at a certain level and then varying the absolute disparity of the front plane only. Then, they would vary the pedestal disparity and perform the same manipulation. The reasoning was that if absolute disparity is what is important, then cells should prefer a particular absolute disparity regardless of the pedestal disparity and the location of the rear plane. However, if relative disparity is what is important, then cells' responses to absolute disparity changes should not be consistent, but should vary depending on the pedestal disparity (and therefore, the disparity of the other plane). Second they measured relative disparity-selectivity by keeping the relative disparity set and changing the pedestal disparity. The idea was that relative disparity-selective cells should prefer one relative disparity (and thus show a consistent tuning curve to relative disparity) regardless of the variations in the pedestal disparity, while absolute disparity-selective cells should show the opposite pattern. Responses for a cell with a strong relative disparity preference are shown on top of Figure 3.13 below, and those from a cell with a strong absolute disparity preference are shown at the bottom. Note that in the top cell, the absolute disparity response changes as an effect of the pedestal disparity, but the relative disparity preference doesn't. This connotes a specific preference for relative disparity, since the cell responds equally to the relative disparity (i.e., has the same tuning curve) no matter the absolute disparity. The bottom cell shows the opposite pattern, showing a specific preference for absolute disparity (in each case, the preferred disparity just happens to be zero, but this varied for different cells).

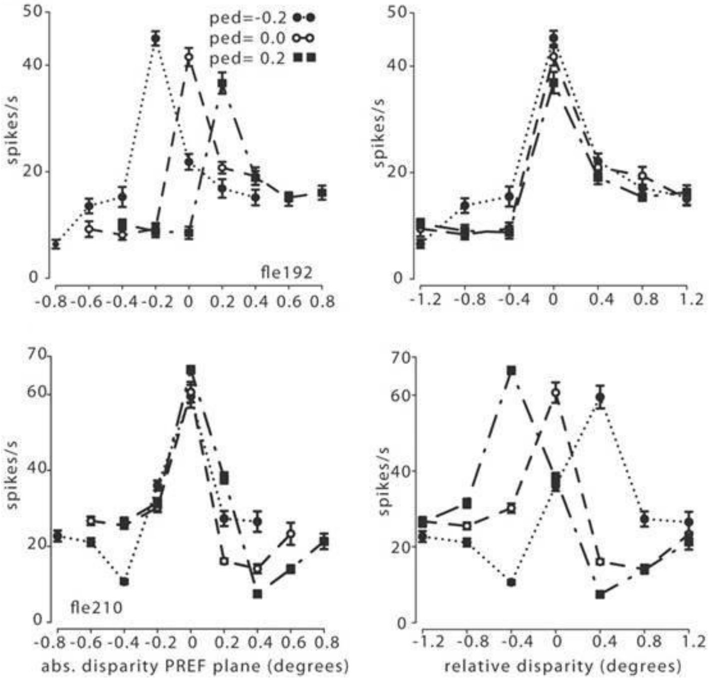

Figure 3.13. A cell with a consistent relative disparity tuning curve, and variable absolute disparity responses (top two panels), and a cell with the opposite preference (bottom two panels). From Krug and Parker (2011).

Since cells selective for relative disparity have tuning curves for absolute disparity that change with pedestal disparity modulation, the strength of selectivity for relative disparity can be measured by the extent to which the absolute disparity tuning curve shifts as pedestal disparity is changed. When these “shift” ratios were measured, Krug and Parker showed that the population average shift ratio is significantly non-zero, suggesting that the population in general has selectivity to relative depth (although of course the shift ratio varied between cells). In sum, then, the two studies discussed here suggest that MT cells indeed show detailed responses in the presence of relative, not just absolute depth.

A few caveats are in order here. First, the stimuli in these studies were moving in the studied cells' preferred directions (unlike the small sample of neurons tested for tilt and slant responses in stationary conditions in Nguyenkim and DeAngelis, 2003). However, there are aspects that resist an easy retreat to the input strategy. First, in Sanada et al.'s studies, motion was a backdrop consistent condition above which velocity and disparity were varied. Since some cells had a much higher weighting for disparity in their tilt judgments, the results suggest that disparity is much more important for some cells in determining fine-grained depth features. The situation is thus similar to Palanca and DeAngelis' (2003) study, which they summarized as showing that the cells that responded more strongly to stationary than moving stimuli were part of a continuum of depth- and motion-preferences, some reflecting more influence of depth, some of motion. As such, the reasonable conclusion is that both disparity and motion velocity contribute, but in dissociable ways, to MT responses, as, by extension, do the overall depth characteristics of the stimulus. Moreover, the rear plane in Krug and Parker's study was in fact moving in cells' non-preferred directions, while the front plane was moving in the preferred direction. The relative depth response, however, *depends equally* on both planes, since it is in essence a comparison of their absolute disparities. The motion signal in the rear plane, on its own, would in fact inhibit MT responses due to motion opponency (Snowden et al., 1991; see also chapter 2). Still, further studies should be expected to examine the strength of relative disparity tuning independently of motion.

Finally, establishing the psychophysical relevance of relative disparity signals remains to be done. Krug and Parker (2011), in fact, show that relative disparity selectivity is *not* vital to the rotating cylinder tasks studied by Dodd et al. (2001) and Grunewald et al. (2002). They measured choice probabilities for the cylinder task from both cells more strongly selective for relative disparity and those more strongly selective for absolute disparity, and found no significant difference between them. As they point out, however, this does not mean that the relative disparity information is never useful. Further studies are required which more explicitly test relative depth tasks in order to determine if and when this is the case. Given the trends in the research, it is a reasonable bet that such contexts will be found.[43]

I have examined studies of depth representation in MT in considerable detail. In the next section, I abstract a bit to the general findings I take to be important for assessing TA. I then argue that none of the strategies outlined in the last section are likely to overcome the growing evidence of depth-related functionality in MT. Finally, I show that the use relation cannot be relied on to save feature-specificity.

3.4. MT and Absolutism.

It is by now well accepted that MT cells exhibit significant responses to disparity. Most of these cells respond to both motion and disparity. Some cells, but not all, respond to disparity signals in the absence of motion, and even for relative

---

[43] Similarly, Uka and DeAngelis (2006) performed a microstimulation study which showed that stimulation of MT cells influence absolute but not relative depth judgments, but they based their study on the idea that MT cells only signal absolute disparity, and therefore did not attempt to stimulate cells with relative disparity preferences. Moreover, they used a variation on DeAngelis et al.'s (1998) task, which did not independently vary absolute and relative disparity in the same way as Krug and Parker's study, instead categorizing as relative disparity the behavior on a "fine" discrimination task between two absolute depths. As they say, somewhat prophetically, "relative disparity selectivity, in general, may be highly dependent on stimulus configuration" (Uka & DeAngelis, 2006, p. 6802)

disparity stimuli depth signals make a dissociable contribution to MT cell response properties. While more investigation is needed, there is at least strong evidence that the disparity signals at MT are used for depth perception. As such, there seems to be strong motivation to deny that MT is a feature-specific area. MT seems, at least, to be a motion and depth area, for which cues such as color make less frequent, but occasionally significant contributions. Is there any way of saving the TA-interpretation of MT by employing the strategies discussed in chapter 2?

The three factors most important for responding to each strategy are (i) the widespread significant responses to disparity in MT cells, (ii) the ability to signal depth independently of motion, and (iii) the dissociable contributions of depth and motion to MT cells in particular stimuli. By "dissociable" contributions I mean to stress such factors as that individual MT cells have independent, and non-correlated, tuning curves for both disparity and movement, such that any given MT response can be construed as due to combined input from different cues. While this is clearest in the Palanca and DeAngelis (2003) results, the Sanada et al. study discussed above also suggests that multiple different cues contribute to MT cells' responses in this dissociable fashion. For simplicity's sake, I will continue to speak in terms of linear contributions of each type of information to cell responses, with particular weighting functions on each that characterize each cell. Nothing I will say hangs on the combinations being linear, however. I will speak of "motion" and "disparity" signals, where motion ranges over both direction and speed preference, and disparity over both absolute and relative preference.

First, consider the combined input/modulation strategy. The reasoning relied on in the color case was based around something like Fodor's asymmetric dependence condition, the idea being that the function of a neural area is tied to the information that it always represents. Since the color influence on MT serves to modulate *what kind of motion* is in the stimulus—e.g., pattern versus component—the argument states that color only serves as a cue for representing motion, and is not itself represented in MT. The combined results in the depth case speak against this kind of reasoning, since they establish independent contributions from disparity and motion. Many MT cells, though not all, respond to disparity independently of motion. In these cases, there simply is no motion signal to modulate. If this information can be useful for judging depth, as the studies discussed above suggest, then depth information, present at MT, can contribute to perception independently of motion. In these cases, there simply is no motion representation to modulate. Given the presence, dissociability, and functional relevance of depth information in MT, the input/modulation strategy will not work. It is not true in every case that the functional relevance of depth information is to modulate a motion representation.

Once the input/modulation strategy is off of the table in general, there is little reason to judge the cells whose responses depend on both motion and depth as solely representing motion. For cells whose response profiles are constituted by some combination of weightings to motion and depth, there is a continuum of relative influences on the cells from the two kinds of information—Palanca and DeAngelis (2003) point out that neurons that prefer slower speeds respond more strongly in absence of motion. A similar property is shown in the tilt study by Sanada et al. Cells exhibit different contributions from motion and disparity in their tilt tuning curves (Figure 3.11), some more strongly influence by motion and some more strongly influenced by disparity. Tilt, therefore, is not tied solely to a motion cue—disparity, in many cases, contributes to it as much or more than motion. Thus, there is no universal priority relation for cells in MT; they exhibit genuine sensitivity to both disparity and motion, and both kinds of information seem to have functional consequences. The combined results stretch the input strategy farther than it can go.

What about a strategy based on top-down modulation? One might posit that a decision process based on motion feeds back to MT cells and modulates their firing in a way that subsequently reflects a depth cue in the stimulus. The decision process might make use of depth information from elsewhere—for instance, somewhere in the ventral stream—and use that information to modify MT responses, for instance in the rotating cylinder cases of Dodd et al. (2001) or the plane stimuli of Krug and Parker (2011). This strategy would attempt to read these affects along the lines of the perceptual filling-in case discussed in the previous chapter. Several of the results speak against this version of the strategy, however. Importantly, the timescale of disparity responses in general in MT is similar to those for motion responses. Palanca and DeAngelis (2003) showed that significant information about preferred depths in stationary conditions is present in MT-cell responses within the same time-frame as information about preferred directions of motion in motion conditions. Dodd et al. (2001) show that responses to disparity cues in MT occur just as quickly as responses to motion cues. Uka and DeAngelis (2004) show that choice probabilities for MT cells in the depth discrimination task are significant as early as 50-60 ms post-stimulus, which is around the timing of emergence of significant CPs in the Britten et al. study for motion-perception. Importantly, it would take time for a decision-mechanism to integrate information from other sources and then feed back to MT—the very fast responses of MT cells to disparity, and the comparable timeframe of choice probablilities based on MT responses to both depth and motion, speak against this type of top-down modulation interpretation of the results.

A second version of a top-down modulation response would posit a role for attention. If seeming MT responses (and choice probabilities) to depth were in fact due to an attentional effect—either spatial or featural—in a motion stimulus that happened to be correlated with a depth response, then there would be no need to posit selectivity for depth in MT. Spatial attention is unlikely to cause the effects discussed here, since generally when the stimulus contains both motion and depth cues these are in the same location in the visual field (i.e., they are features of the same object or objects). If feature attention were invoked, then importantly it would have to be attention *to motion* that caused the change in responses, since the objection is based on showing that MT cells do not respond genuinely to depth. But as Grunewald et al. (2002) point out, such a response is not possible for rotating cylinder stimuli, since motion is entirely ambiguous without the depth cue. Of course, positing that the depth is first disambiguated, then attention paid to direction-of-motion, is equivalent to the first modulation response, since it supposes that the decision is already made, and then feedback to the motion-representing cells occurs. But this version of the objection has already been dismissed. Feature-based attention to motion also cannot account for the sensitivity of MT cells in the Palanca and DeAngelis study, since these cell responses occur even in the absence of motion cues. Finally, it is unlikely that attention to motion could account for the detailed tuning curves that MT cells exhibit to different aspects of depth. If motion were what is relevant in a signal from MT, and if disparity only served as cues to distinguish motion such that attention could be paid to it, then there seems to be no reason why MT cells should show such detailed preferences and responses to specific depth cues.

Now, consider differentiation. DeAngelis et al. (1998) originally suggested that there were columns of neurons in MT that are selective for both disparity and motion, but that these are interspersed with purely motion-selective cells. A differentation strategy would contend that positing feature specificity for MT as a whole may be false, but that there are functional distinctions to be made *within* MT that correspond to feature-specific representation. Interspersed columns are not as clear a spatial distinction as, for instance, that proposed between color and form-selective regions in V4 (chapter 2), but that is no bar to the differentation strategy generally. In V1 and V2, as discussed, M- and P-pathway cells exist spatially segregated units within the same area. However, in a study with a considerably greater sample size than the original DeAngelis et al. (1998) study, DeAngelis and Uka (2003) found that 93% of MT cells show significant tuning for disparity, and report no significant spatial clustering of the ones which are have no disparity tuning.

They explain the “patchy” organization uncovered in the earlier study via the interrelationship between their single cell recordings and the multi-unit recordings from DeAngelis et al. (1998) drew their organizational conclusions. They contend that, rather than not being disparity selective, certain parts of MT have groups of cells with less correlated disparity preferences. In a multi-unit recording these areas would have less concerted signal than in others with more greatly correlated depth preferences.

Stepping back for a moment, consider just the three facts that (i) the vast majority of MT cells are disparity selective, (ii) that the non-selective ones are not significantly grouped together, and (iii) that MT cell responses are generally a combination of weights on motion and disparity cues. These properties allow for *many* different kinds of organization over and above columns for direction and disparity. We’ve seen that tilt preferences are grouped together in MT; similarly, the Palanca and DeAngelis (2003) study showed some clustering for particular speed/disparity. The picture that emerges, however, is not one in which motion and depth information are strictly segregated in MT, but instead one on which different MT cells combine motion and depth in different ways for specific purposes.

To summarize, the current state of physiological and psychophysical results suggest that MT is not purely a motion detector. It is now helpful to think about what these results mean for the use relation. Take a highly simplified case in which a group of MT cells’ responses are linear combinations of disparity and motion cues, for which the cells have standard tuning curves, and where each cue is equally weighted. In this situation, a “read-off” mechanism charged with making decisions based on input from MT cells will not know, for any change in MT cell response, whether it is a change in motion or in depth. A given input could be due to a purely motion signal at the cells’ maximal preference, or to a combination of motion and depth signals, neither at the peak of the cell’s tuning curve for that stimulus. Expanding to a population with different weightings does not seem to change the nature of the problem, but only complicate it. As long as MT cells are responsive to both disparity and motion, a putative decision maker will not be able to make motion-specific decisions based on the response, absent some further explanation.

Of course, since in some cases motion decisions get made based on MT response, and depth decisions in others (that is, if the choice probabilities are to be believed), *some* process must eventually disambiguate the information. How this happens is of considerable interest in neuroscience generally, and especially in the current situation. Further complicating things is the fact that, as one would expect from the combinational view of cue influence on MT response, depth influences MT cell responses even when depth is entirely irrelevant to the task at hand (DeAngelis & Newsome, 2004). Once again, it will be useful to idealize. Consider two diametrically opposed (and equally unrealistic) ways of making a perceptual decision. One is to have an omnipotent decision maker, which knows all of the details about its inputs, and chooses to use an ambiguous signal as a signal for only one feature. In this case, MT might only be ever be used for motion, but the decision mechanism would not work *in virtue of* the feature-specificity. Instead, it would be quite the other way around, with the feature-specificity being in virtue of a mechanism of use that operates on entirely other principles (in this case, the useful one of omnipotence). If the account of the commitments of the MFH theory given in chapter 2 is true, then the formalism would still be undermined by this decision procedure, because the procedure exhibits the inverse in-virtue-of relation to the one that is definitive of the MFH view.

On the other hand, we might focus on decision makers that are themselves highly specialized. Consider two dissociated decision mechanisms, one entirely for depth decisions, and one entirely for motion decisions. We could further assume effectors attached specifically to each decision mechanism that control behaviors solely related to motion or depth. These decision mechanisms blindly sum all of their inputs and cause their effectors to fire. As such, signals from the population of MT cells will be taken as only motion signals by one detector and only as depth signals by the other.[44] But this is not feature-specificity *for MT* at all! On this new scheme, feature-specificity is not needed to produce perceptual decisions; it thus simply denyies the TA aspect of the MFH theory.

Of course, it is equally unlikely that one decision process makes all decisions omnipotentaly, or that all decisions have their own unique mechanism. To make the considerations here more concrete, consider the possibility that, for specific tasks, perceptual decision-makers might “learn” to rely on whatever neurons are most informative for the task. MT cells selective for particular absolute disparities, for

---

[44] This would assume, of course, that something determines which decision is needed at the current time. Nothing hinges on this for the current example; just assume it’s random.

instance, will be the most informative for tasks about coarse depth regardless of what else they may represent. There are a variety of precedents for the idea that perceptual decisions are based on the most informative neurons for the task at hand. Recall from chapter 2 that on certain views of population coding, the population response is based only on the most informative set of neurons. There is no reason why such a principle could not be extended to multiple distinct tasks involving information represented in an area. In fact, Uka and DeAngelis (2004) suggest just such a scheme for the depth judgment in the original DeAngelis et al. (1998) task. They propose that, during training, whatever decision process ultimately implements the depth decision learns to base the decision on the MT neurons with near and far disparity preferences. Interestingly, this has negative as well as positive consequences. Uka and DeAngelis, in some trials, switched the stimulus to one where neurons selective for zero disparity were the most informative (because either the noise or signal dots were at the plane of fixation), yet still found that CPs were significant only for near- and far- tuned MT neurons. The authors contribute the effect to the monkeys "learn[ing] not to monitor" the zero disparity cells during training (2004, p. 306).[45]

Perhaps the best way to describe the possibility evinced here is that there are general perceptual learning mechanisms that can learn to perform new perceptual tasks, and that their learning principles allow them to eventually employ the most informative neurons from their inputs (this probably happens slowly, as monkeys must undergo considerable training to perform these tasks). This learning, however, doesn't

---

[45] DeAngelis and Newsome (2004) also suggest, based on similar reasoning, that their monkeys showed individual differences in "task strategy"—i.e., in which neurons were used to inform the perceptual decision.

seem to depend on feature specificity. It can be true that the MT cells get used for depth in the task above, but that doesn't mean that they are used in virtue of representing only depth. Rather, the decision process is based on learning that certain neurons are informative for a specific task, and employing them in that context. This does not prevent the neuron from also contributing to other contexts based on the other information that it responds to—it is compatible with the same neuron being used for motion representation in other contexts.

3.5. Conclusion: Towards a Contextualist Methodology.

Return to the desiderata discussed in chapter 1. Absolutism is a commitment about projectability—namely that a single functional description should cover all contexts. Read as such, however, the evidence discussed in this chapter suggests that the MFH theory's TA-style decompositions into feature-specific areas do not hold across contexts. That is to say, the MFH theory, if read in absolutist terms, does not meet empirical adequacy. The fact that cells' responses, in general, depend upon multiple parameters, which can vary in different stimuli, suggests that their actual responses are context-dependent. An MT cell will exhibit specific patterns of responses as depth and motion cues are varied, which reflect the weightings it applies to each cue. These distinct signals can then be used for distinct perceptual tasks, as the studies in this chapter have shown. Since in some circumstances there are representations of other features at MT which cannot be explained away in terms of one of the strategies, it seems that we must posit MT as representing distinct things in distinct contexts—this is incompatible with the ideal of absolute projectability. The

MFH view, as specified most clearly in the input strategy, argues that any variation within MT responses must be due entirely to variation in aspects of motion, such that the description of the area as a motion detector needn't change with that variation. But the arguments and results in this chapter suggest that that is not the case. Depth and motion are both represented in MT; which one depends on the context. Since anti-contextualism is a corollary of TA, and the MFH theory an implementation of TA, the current results are strongly in tension with the theory.

In chapter 5, I will expand this perspective into a contextualist view of localized function. For the time being, I want to highlight a few points about how the discoveries regarding MT and depth and color responses were made, which will be useful for the eventual account. Unsurprisingly, the methodology involved manipulation of context at every turn, and insight about potentially relevant contexts can come from a variety of sources. I have discussed several in this chapter. Psychophysical results suggested the relevance of color segmentation for motion perception, which was eventually also shown in physiological investigation of MT. Moreover, the search for color influence on MT paralleled the increasing recognition in anatomical studies that MT receives input from P-pathway cells both in V1 and V4. Finally, while it has long been known that MT cells respond to disparity, and that disparity is a cue for depth, intense interest in analyzing depth perception in MT was spurred largely by new physiological insights about just how prevalent disparity tuning in MT is. There are thus many sources of insight into context, and many ways of constraining search for contextual influence. I will explore these in detail in chapter

5, as a way of showing that a contextualist view can meet the desiderata on a theory of functional localization.

At each step in the analysis of color and depth influence on MT, a reasonable conservatism was exhibited, where results were interpreted by departing as little as possible from the traditional MFH view. The early study of Dobkins and Albright concluded that color influence could be accounted for with purely luminance and color-border responses. Even the strong depth selectivity discovered in MT was, for a considerable period, interpreted as only being relevant for coarse depth perception, not for fine depth judgments of the type relevant to object perception. In each case, there is strong evidence that a more revisionary account is needed, but this evidence was discovered in part by first assuming a less revisionary view. This kind of development speaks towards a more heuristic role for the MFH theory—namely as one that has guided exploration of context, without being *true* in all contexts. The science has also not ground to a halt with increasing evidence against the MFH view. Indeed, in addition to new discoveries about MT functionality in color and depth contexts, advances are being made in thinking about how perceptual learning can come to take advantage of specific information in neurons that are tuned along multiple parameters, as discussed at the end of the last section. The view I develop in chapter 5 is meant to describe and inform these developments.

Before advancing the contextualist view in chapter 5, one more negative argument is necessary. Several theorists have drawn similar conclusions about the failings of TA from similar evidence in other systems—namely evidence which shows the contextual variation of information represented at an area, or the diverse functions to which an area can contribute. However, rather than going contextualist, they argue for absolutism of a different type, specifically absolutist function ascriptions described in computational terms. In the next chapter, I argue that such views are not a significant advance beyond TA.

Chapter 4. Against Computational Absolutism

4.1. Introduction.

In the previous chapter, I argued that there is good physiological evidence that MT responses are tuned to multiple distinct features, and these distinct feature representations are functionally useful in distinct contexts. Therefore TA, which argues for a single, feature-specific account of MT's function, fails to be empirically adequate in this case. If we don't wish to abandon the project of functional localization, then, some other way of talking about function needs to be developed. One attempt that has gained considerable recent momentum is to argue that absolutist functional localization can be pursued despite the kind of results I've discussed. Under what I will call "computational absolutism" (CA), results undermining TA show, not that absolutism is wrong as such, but instead that the traditional way of individuating functions is flawed. Function should not be described in terms of carrying specific information, representing specific things in the world, or contributing to particular psychological faculties (language, memory, etc.), but instead in terms of carrying out a *computation* of a particular type. Put another way, functions should not be individuated in terms of what specific information particular areas process, but instead in terms of *what they do with it*. Even if a neural area processes many different kinds of information, perhaps it does the same thing with all of the different kinds. If so, then a particular functional attribution—in this case, one described in computational terms—will still hold across all contexts. If so, absolutism can be maintained.

The motivation for CA can be divided into two related parts, both inspired by absolutism. The first is the idea that, given the particular anatomical structure of each brain area—including both its pattern of internal connections and its connections with other areas—some particular functional description should be able to fully describe how that structure works. CA is thus partially motivated by the goal, which is standard in much of biology generally, of finding functional explanations for anatomical structures. The second idea is that the right kind of goal for computational neuroscience—i.e., what "successful" explanations in the field will look like—is finding univocal computations that explain in every context. The goal, as Price and Friston (2005) phrase it, is to find "*an* underlying function" (p. 263, emphasis added) for each part of the brain, where "the most useful functional labels are those that explain and predict how an area responds in different contexts" (p. 265).[46]

Good theories of functional explanation should be both *descriptively* and *normatively* adequate (Machamer et al., 2000). They should both describe the actual practice of the science and provide good normative principles for them to follow. I will argue that the motivations for CA do not constitute good descriptive and normative principles for computational investigation into MT. I will discuss only one type of model, motion-energy models, in detail, although I will mention other

---

[46] In later work, Friston and Price (2011) seem to soften their stance considerably, suggesting a view more friendly to contextual variation in function ascriptions, when they stress the importance of "understanding the context-sensitive nature of distributed but modular processing" (p. 241). However, they do not extend contextual variation to the functional account of particular areas: "Functional segregation refers to the specialization of brain regions for a particular cognitive or sensorimotor function" (Friston & Price, 2011, p. 242).

modeling projects as relevant. The great success of motion-energy models makes them a good test case for evaluating CA in perceptual neuroscience.

The arguments will proceed as follows. The first, anatomical claim of CA assumes that there is an extremely close relationship between the anatomical structure of a particular brain area and the computational account that describes its function. Motion-energy models, however are employed for different phenomena of motion-perception with different degrees of attention to MT's structural architecture. Some of these models discover functionally important properties of MT responses without relying on describing the anatomy to any particular degree of detail. This supports a general perspective, which has been voiced elsewhere (Piccinini & Bahar, 2012), that employing a computational model in neuroscience involves selectively idealizing from anatomical architecture for particular purposes.

I then argue that the second motivation fails to provide a reasonable explanatory ideal for projects in computational neuroscience. Even though motion-energy models provide a powerful framework for modeling MT responses in particular contexts, they do not explain how MT will work in *any given* context. That is, they fail to be projectable in the way that CA claims, and to obviate the need for contextual descriptions of MT function. Given these two arguments, CA's claims to achieve absolutist explanations of function are unsuccessful. I conclude the chapter by diagnosing the intuition behind CA as a kind of unificationist impulse that, while not flawed in principle, is not the right approach to localizing neural function.

In section 4.2, I discuss CA and its motivations in more detail. In section 4.3, I introduce a range of motion-energy models. In section 4.4, I argue against the anatomical motivation for CA, and in section 4.5 against the normative ideal of absolutist description. In section 4.6 I discuss the unificationist impulse and what role it properly can be seen to play in computational neuroscience.

4.2. Computational Absolutism.

In chapter 3 I argued that MT functionally responds to multiple different stimulus contexts involving multiple features, and therefore that a TA-style description of MT's function is inadequate. Proponents of CA have used similar reasoning in dismissing TA-style accounts for other areas of the brain. Generally, these theorists have criticized *task*-based analyses of function. On such analyses, the goal is to measure from or intervene upon a brain area during the performance of a certain task. Increased activation during a particular task, or interruption of the task when the area is intervened upon (either naturally or artificially), is then taken as evidence that the area's function is to perform the task. Anderson (2010) discusses Broca's area as an example—it is traditionally posited to be task-specific for the production of language. Anderson cites results showing that Broca's area is also active in non-language contexts—particularly in imagery and action preparation—as controverting the traditional functional ascription. Further, he cites meta-analyses of fMRI studies showing that the average cortical area is involved in nine different types of behavioral tasks, and therefore suggests that functions cannot be individuated by task-specificity.

Rathkopf (2013) cites the impressively large number of theories that have accumulated regarding the proper task-domain of the hippocampus to argue for a similar conclusion. He suggests that task analysis is flawed as a way of describing functions. The reasoning is that if an area's function ascription is tied to particular tasks, then the function ascription can, in principle, change every time the task does, producing a potentially infinite number of function ascriptions. As Rathkopf puts it, "there could be no grounds for claiming of any list of task-bound functions that it is complete. … Any functional contribution to behaviors yet to be taken into account would demonstrate the incompleteness of the list" (2013, p. 10). The conclusion drawn is that projectability is undermined by function ascriptions that may vary with context: "context sensitivity systematically prevents such [ascriptions] from explaining neural structure" (Rathkopf, 2013, p. 3). A correct functional account should be "complete" in that it does not admit of such variation. As in the quote from Price and Friston above, the right functional account should explain and predict what the brain area will do in any given context.

Successful functional localization, on CA, involves positing a particular type of computation that is carried out by a neural area in all contexts, regardless of the particular domain or task to which, in a given instance, the area is contributing. Put simply, no matter what information they receive, particular neural areas always perform a given computation, and it is this computation that defines and individuates their function. Bergeron (2007) and Anderson (2010) phrase the relevant distinction as between *use*, which varies with context, and *role* or *working* (respectively), which is invariant and univocal. As Anderson puts it, a working "is whatever single, relatively simple thing a local neural circuit does for or offers to all of the functional complexes of which the circuit is a part" (2010, p. 295).[47] The absolutist tenor of this style of explanation is clear—an area is supposed to perform one working in any of the "functional complexes" underlying particular task-performance in different contexts.

The two motivations for positing CA, as discussed in the introduction, are the desire to explain what a particular anatomical structure does, and to maintain the explanatory ideal of absolutist projectability. In what Rathkopf calls "structurally oriented explanations," a functional explanation is one on which "the structural properties of the device … make sense" (p. 6). That is, we isolate a part of the brain, such as MT, with a distinctive anatomical structure, and the functional explanation explains what that structure does. Price and Friston (2005) posit that an area's function is constrained by its "fixed," anatomical connections. Anderson argues that the computational working performed by a brain area is determined by its anatomy. Each of these theorists thus posits a very close connection between computational descriptions of a brain area and anatomical/structural ones. The goal, on CA, is to figure out what computation the fixed structure performs. The second motivation is to meet the explanatory ideal of "complete" explanations. As can be seen in the quote from Rathkopf above, complete explanations—ones that do not need to change with

---

[47] An aspect of Anderson's focus with which I am highly sympathetic is his focus on complexes and networks—he suggests that use is determined by other areas to which an area can be connected, where these can vary widely. However, I disagree with Anderson's idea of how we should *understand* networks. He thinks that networks should be understood by positing one single computational working for each part of the network, then explaining the different uses of that working by uncovering the other parts of the network with which it interacts. In my view, the function of each part of the network should be understood in terms of the patterns of contextual variation that they exhibit within the network (see chapter 5).

context—are assumed to be better explanations than ones that must change with context. Again, the absolutist intuitions underlying this motivation are obvious.

A few clarifications. First, while I will continue to talk in terms of information processing at MT, rather than specific task-involvement, the two are closely related. For one thing, the idea of a specific task-domain is often tied to that of a specific sort of proprietary information (Fodor, 1983)—e.g., a specialized language area processing language-relevant stimuli, or a face-perception area responding to face-relevant information . Similarly, I have shown how establishing MT's representational function has often involved establishing the use of its particular informational responses for specific tasks (e.g., motion tasks). I will thus assume that results showing that MT represents distinct featural information in distinct contexts are the kinds of results that CA theorists wish to rule out of functional explanation.

Second, there is some bit of variation in how theorists within CA describe the particular computations that they think comprise functional explanation. Bergeron (2007) argues that Wernicke's localization of speech perception in the temporal lobe was not described in terms of language specificity, but instead in terms of an "acoustic images center." Anderson suggests that Broca's area has a basic "sensorimotor" function that is relevant to both language and action. Rathkopf cites Levy, Hocking, and Wu's (2005) view that the hippocampus is an "associator of last resort," which involves connecting perceptual representations that are not obviously related to each other. Price and Friston (2005) suggest that the function of the posterior lateral fusiform area, commonly thought of as a "reading area," is in fact specialized for "sensorimotor" integration. These descriptions are qualitative. Rathkopf, however, suggests that there is a mathematical nature to such functions—he proposes a "signal transform" view of what particular neural areas do. That is, they take inputs (of whatever kind) and perform a particular mathematical or statistical operation over them, to produce particular outputs. Such descriptions are standard in what is commonly called computational or theoretical neuroscience.

More could be said here about the qualitative versus quantitative aspects of CA. It is unclear whether Bergeron or Anderson would agree that roles or workings should be described in signal transform terms (in other work, Friston strongly suggests the quantitative perspective; see Friston, 2010), but I will focus on the former primarily, since it is more consonant with modeling projects in MT. What is important for now is that the views each attempt to secure absolutism by switching from TA -style functional descriptions to CA-style ones. In the next section, I discuss motion-energy models of MT, which I will use to argue that CA is neither descriptively or normatively adequate as an account of functional localization in perceptual neuroscience.

4.3. Motion Energy Models of MT.

4.3.1. Modeling Motion Energy.

In perceptual neuroscience, computational models posit a type of computation that an area performs. The model is "successful" for a range of stimuli if, for those stimuli, the behavior of modeled cells is similar to the response of real cells. For instance, if the model predicts a certain level of increase in activation, this should be reflected in the firing rates of actual cells. Often, in MT, computational models attempt to capture the kinds of properties discussed above—e.g., direction selectivity, pattern motion, and opponency. The process of constructing a model involves positing a basic computation, designing the stimuli, and fitting the parameters of the model (such as the level of excitation for a certain stimulus property) to physiological data. Successful models explain the data by showing how the computation, performed over the input, produces the right kind of responses.

A significant trend in modeling motion perception, which subsequently formed the basis for several successful models of MT, began with Adelson and Bergen's (1985) classic "motion-energy" model. They started from the idea that a stimulus moving in two dimensions can be modeled in a three-dimensional space—in this case, two spatial dimensions and a temporal dimension. To visualize such a space, consider the following figure. In the left panel, y and x are spatial dimensions. The figure on the right shows these two spatial dimensions and a third dimension, t, representing time. When the bar in the left figure moves to the right (as suggested by the arrows), it takes the path through space and time shown in the figure on the right.

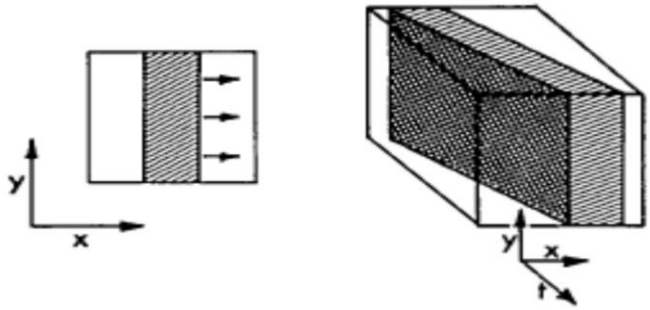

Figure 4.1. Motion in the spatio-temporal energy domain. From Adelson and Bergen (1985).

The idea of a motion filter is shown below (in this case, with one spatial dimension and one temporal dimension). A cell (for instance) that responded to the particular spatio-temporal pattern shown by the plus signs, and not to the one shown by the minus signs, would reliably signal the presence of motion at that spatial trajectory (and speed).

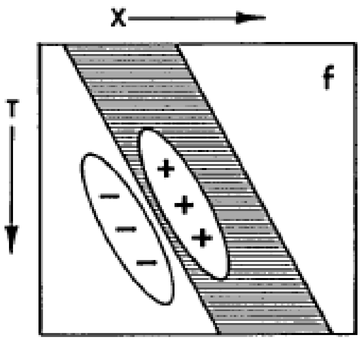

Figure 4.2. Spatiotemporal filters. From Adelson and Bergen (1985).

Motion filters are generally modeled in the spatio-temporal frequency domain. Spatiotemporal frequency can be determined by Fourier decomposition, which breaks down the overall stimulus into a power spectrum of frequencies in the dimensions of the stimulus (in the case shown in figure one, this would be two spatial dimensions and a temporal one). Each stimulus can be described as a combination of spatio-temporal frequencies, and their amplitudes. In the frequency domain, a direction-selective filter responds to particular frequencies in the spatio-temporal decomposition—they thus represent a particular “component,” of the decomposed “motion-energy” in the stimulus, which can be thought of as a combination of spatial frequency, direction, and speed (Bradley & Goyal, 2008). A filter for a particular component is shown below. It is constructed as a function of two individual functions, one each for space and time, which get multiplied at each space-time point, thus producing a filter for certain frequencies. A moving stimulus (e.g. a grating) whose spatio-temporal frequency matches that of the "plus" parts of the filter will strongly drive this unit.

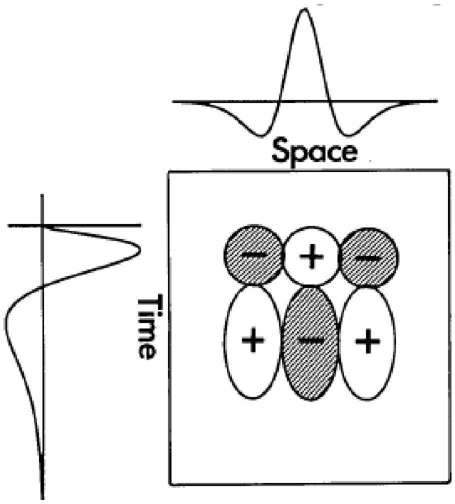

Figure 4.3. A filter in the frequency domain. From Adelson and Bergen (1985).

There are some further complications that I will not discuss in detail here.[48] The consensus surrounding Adelson and Bergen's model is that it is a good description of "early" vision processes, such as those present at V1 (Nowlan & Sejnowski, 1995). Filters of the type discussed above will not accurately compute pattern motion, due to the aperture problem—multiple different object directions are

---

[48] For instance, since the filters themselves are defined by frequency, their responses will be dependent on phase—that is, even for a stimulus frequency for which they are tuned, their responses will vary over the time course as the stimulus frequency stands in different phase relationships to the filter frequency. To solve this problem Adelson and Bergen constructed paired filters of the same frequency preference but different in phase, whose outputs were squared and then summed. The squaring cancels out any difference in sign between the responses at a given time, and the summation guarantees that the response will be consistent over the time course (since at each time one filter will be more active and one less, due to their relative phases). As Nowlan and Sejnowski (1995) point out, the schema on which multiple, more variable filters provide input to one less variant one matches the general schematic relationship between simple and complex cells in V1, further supporting the analogy of motion-energy models to V1 processing.

compatible with significant spatio-temporal energy in the preferred frequencies of these filters (Bradley & Goyal, 2008). As such, many MT models use motion-energy filters as *inputs* to modeled MT cells. Modeled MT responses are then intended to capture the particular features of motion perception—including pattern motion responses and opponency—attributed to MT, as compared to V1. The hope is that MT's function can be uniquely characterized in terms of "simple operations in the spatiotemporal frequency domain" (Bradley & Goyal, 2008, p. 686).

The hope of describing MT in terms of a particular computation over motion-energy inputs seems to match the CA characterization of function relatively well. In the following subsections, I discuss models that have expanded the motion energy idea to account for MT's particular response properties, and then argue that they do not in fact support a CA-style account of MT function.

4.3.2. Summation and Normalization Models.

Nowlan and Sejnowski (1994, 1995) modeled MT inputs from V1 using Adelson and Bergen's model, with MT receiving the motion energy representations from each point in the modeled visual field. In the model, before the V1 signal is passed on, a normalization operation, called "soft-max," is implemented. The operation first performs a uniform exponential increase to the filtered signal from each V1 cell, then takes the sum of all of the modified responses, and finally divides each response by the sum. The resulting signal is "normalized" because the output for each patch, due to being divided by the summation, will be between zero and one. The result emphasizes "the differences between the units with the highest levels of activity” (Nowlan & Sejnowski, 1995, p. 1196), since units that don’t already have high levels of activity will be brought closer to a null output than those that do—this kind of “summation and normalization” is thus often used as a way of reducing noise. MT then receives this normalized signal.

In the model, MT itself consists of two distinct populations with very different properties, each of which contributes to motion perception—“local velocity” and “selection” networks. Each “unit” of the local velocity network consists of a number of “velocity-tuned” units, each with preferences for a particular spatio-temporal trajectory. Each of these receives input from a 9X9 section of the V1 level, and direction selectivity is implemented by getting input from all direction units tuned within 90° of the preferred direction (which was determined by the experimenters). Each local velocity unit combines these inputs according to a unique set of weights (see below). The setup implements the broader directional tuning and larger receptive fields of MT neurons, compared to V1 neurons. Local velocity cells, then, as their name implies, are meant to compute evidence for a particular velocity “hypothesis” for stimuli within their RFs.

Selection units similarly consisted in an array of inputs from specific parts of the modeled V1 cells, and similarly integrated them according to specific weights. They also had directional preferences. However, the job of the selection network was to gauge how *reliable* the assorted velocity signals were. The difference between networks was due to how softmax was applied for each. For local velocity cells, softmax occurred *within* each cell—i.e., on its inputs. This sharpened the estimation of local velocity by picking out the most probable velocity within the RF. However, for the selection network, softmax was applied across all of the cells. Since in some cells there will be more or less evidence for particular velocities (reflected in the strength of their responses), and since all of the probabilities in the model are forced to sum to one, the softmax implementation will “privilege” the cells at any velocity that have the highest amount of evidence. Finally, a decision level integrates the outputs of the selection and local networks—the decision is based on the local velocities that are most informative.

The model was tested on a range of motion stimuli. It successfully detected the velocities of a range of objects, including particular shapes, contrasts, and textures moving across the field, as well as random dot stimuli with particular levels of correlated motion. The power of the model was due to the interaction of local and selection units. To see this consider the stimulus below.

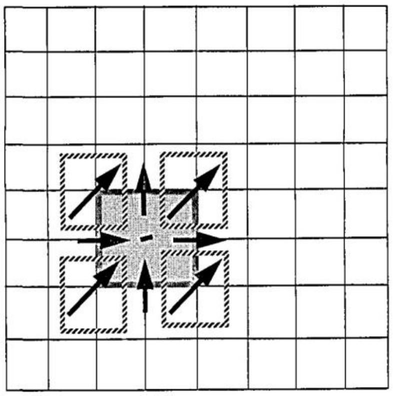

Figure 4.4. One of the stimuli from Nowlan and Sejnowski (1994).

The entire square moves in the same direction, but its parts provide different degrees of information about that direction. The middle, with no contrast, provides no information at all. The single edges, while they provide some information, are subject to the aperture problem (discussed in chapter 1)—the motion signal in cells responding to the different single edges is inherently ambiguous in terms of direction. But since the selection network privileges the areas with the greatest amount of evidence (in this case the corners, which are unambiguous), the overall network will be pushed to that interpretation.

The model fits a considerable range of MT responses. While there was no explicitly modeled inhibition, the model exhibited opponency, which arose in the local velocity cells due to the normalization procedure occurring at each location in the stimulus. The local and selection units, importantly, matched the distinction between component and pattern cells in MT. The selection units, while technically tracking reliability, ended up responding to the direction that had the most evidence it in the local velocity units—i.e., the pattern of motion in the stimulus. The recognition of pattern motion implemented a solution to the aperture problem, as discussed above. They also modeled a series of psychophysical tasks already known to involve MT, showing that the model produced similar responses as human subjects. For instance, Stoner, Albright, and Ramachandran (1990) studied overlapping grating stimuli similar to those discussed in section 3.2, but using only luminance, not color gratings. Whether component or pattern motion is perceived in these settings can depend on *transparency* levels, as shown below (Figure 4.5).

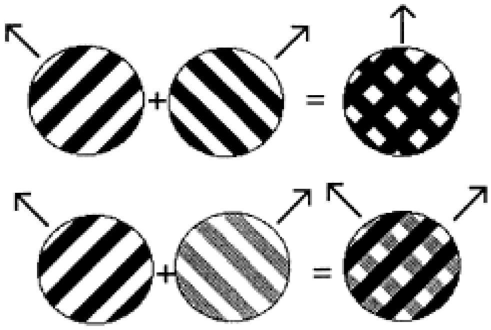

Figure 4.5. The grating stimuli of Stoner et al. (1990). From Koechlin, Anton, & Burnod (1999).

The top of the figure shows two gratings of similar transparency levels (in this case, no transparency), which are perceived as a coherently moving plaid. If one of the gratings shown as transparent, however, as in the bottom of the figure, then the stimuli will be perceived as two gratings moving in separate directions. Nowlan and Sejnowski showed that the model predicts these responses, privileging local unit activation in the transparency conditions, and selection unit activation without transparency. Further, in RDS stimuli such as those discussed in the last chapter, the model exhibited greater selection unit response with greater percentages of coherently moving dots, as has been shown both psychophysically and in MT cell responses. Finally, selection units show the ability to segment the scene into regions based purely on motion (i.e., without having to first isolate another feature of an object), as has also been suggested psychophysically.

A second model in this tradition was proposed by Simoncelli and Heeger (1998). Their model mimicked both the general structure and some of the computational principles of Nowlan and Sejnowski's—particularly the focus on summation and normalization—but differed in the detail and application. For one, MT cells did not simply take input from one region of the visual field, but from a distributed range of V1 cell inputs with the same orientation preference. Relatedly, Simoncelli and Heeger did not explicitly split local and selection networks in MT. As such, the MT cells were primarily concerned with pattern motion; component cells in MT were not modeled. Pattern responses were implemented summing over particular arrangements of V1 outputs, whose combined activity would disambiguate ambiguous velocity (the "intersection of constraints" solution to the aperture problem). Third, there was no explicitly modeled 'decision' stage in the model. The motion decision was determined simply as the peak activation level of the population code of MT cells. The computational detail also differed in several ways, although the model still implemented successive stages involving summation and normalization.[49]

The success of the model involved the ability to account for a wide range MT cell response properties, including their speed and direction preferences, via the manipulation of 3-4 free parameters. Simoncelli and Heeger's model provided more detail about the RF properties of MT cells than Nowlan and Sejnowski's, and accounted for some of the same stimuli. However, the differences are also important for our purposes. For one, Simoncelli and Heeger do not model component cells at all (although they suggest that component cells could be added; see below). As such, some of the results that could be accounted for in Nowlan and Sejnowski's model *based on* the distinction between local and selection cells, such as relative component and pattern responses to moving plaids, are outside of what Simoncelli and Heeger's

---

[49] In particular, the computation was implemented as a rectification step followed by summation. The division step was similar, with the response of each unit divided by the overall sum, but instead of being constrained to sum to 1, constants set bounds on maximum firing rates.

model can account for. Similarly, since each pattern signal is due to the whole population of MT cells, the model cannot segment the motion in the scene as Nowlan and Sejnowski's model can.

While the models are different in several respects, they both fall into the class of motion-energy models, and both employ summation and normalization as a basic type of computation. However, they model different aspects of the structure of MT (for instance, the distinction between component and pattern cells is included in Nowlan and Sejnowki's model, not in Simoncelli and Heeger's). I will discuss these aspects further below, after introducing one more kind of model in the motion-energy tradition.

4.3.3. Competition Models.

Krekelberg and Albright (2005) set out to physiologically measure and model responses of MT cells to apparent motion stimuli, including "phi"-motion and "reverse-phi" motion. Apparent motion phenomena are motion percepts in which discrete spatio-temporal displacements of a stimulus are perceived as continuous motion. Phi- and reverse-phi stimuli are diagrammed below.

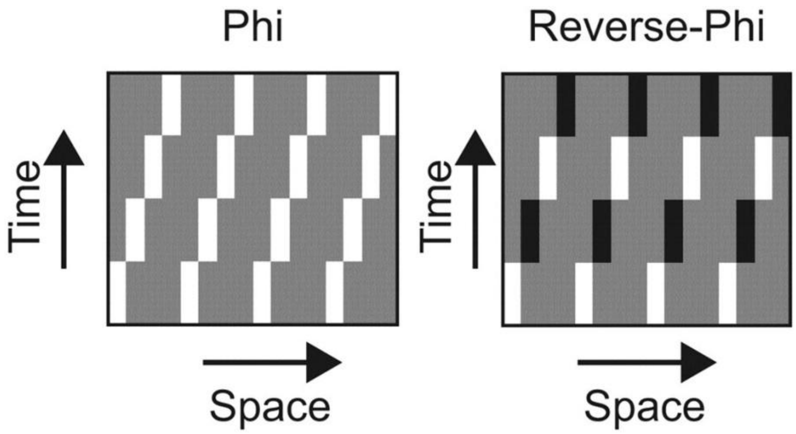

Figure 4.6. The phi- and reverse-phi stimuli from Krekelberg and Albright (2005).

Each panel shows a moving grating over several time steps. In the left panel, the grating moves to the right, and this stimulus produces a phi-motion perception to the right. In the reverse-phi stimulus in the right panel, the displacement is the same, but the grating switches luminance contrast (i.e., going from white to black, or vice versa) at each time step. In this stimulus, perceived motion is generally to the left. To see this, focus on the far right bar in the first time step of the figure on the right. In the second time step, subjects perceive the stimulus as though this bar had moved to the left, and was now the black bar second from the right. This perception matches a shift in the direction of the predominant Fourier energy in the stimulus, although there are still some components exhibiting rightward motion.

Krekelberg and Albright studied monkeys' behavioral and physiological responses to the phi- and reverse-phi stimuli. As expected, monkeys perceive rightward motion in the phi-stimulus and leftward motion in the reverse-phi. Interestingly, neurons in MT *also* tend to switch preferred directions. For example, if a neuron is directionally selective to phi-motion at 315°, a reverse-phi movement at

135° (i.e., 180° away) produced a similar (although somewhat lesser) response. However, a phi-movement at 135° or a reverse-phi at 315° each produced non-significant responses. This is an important result for several reasons. First, the switch in MT neuron direction-responses parallels psychophysical switches in the monkey, confirming the previously suggested view that MT is the neural locus of the reverse-phi perception. More importantly for current purposes, the results suggest that direction selectivity and opponency are more complicated in MT than traditionally thought. If cells really responded only to motion energy in one direction, and opponency occurred to motion energy in the opposite direction simpliciter, then MT cells should not exhibit strong responses to the reverse-phi stimulus. As Krekelberg and Albright note, the results are generally supportive of a motion-energy approach to MT, since in each case motion energy is determining the response. They do not support, however, a straight "subtraction" model of direction selectivity and opponency.

To tease apart how inhibition worked in this context, Krekelberg and Albright tested MT cell responses to a range of individual Fourier components *and* their combinations. Krekelberg and Albright measured MT neurons' responses to each of six Fourier components, and then to different combinations out of each. Unsurprisingly, cells responded most strongly to components in their preferred direction; however, they also occasionally responded above baseline to components in the opposite direction and to stationary flickering components. Responses to multiple between components did consist in linearly summing responses to individual components. For instance, some components dominated others, with responses no different in the combined condition than for the one alone. Others were averaged, and a variety of other, less easy-to-describe relationships occurred as well.

In order to account for the results, Krekelberg and Albright had to posit a different type of computation from the summation and normalization model proposed by Simoncelli and Heeger. Their “competition model” involved an excitatory *and* an inhibitory weight for each component. The responses of individual cells were determined by combinations of the weights the cell placed on each component, and each cell had a unique set of weights. Having both positive and negative weights for each component was unique to the competition model, and this element was essential—Krekelberg and Albright compared the competition model to several models, including Simoncelli and Heeger’s, and the competition model provided the best fit to the physiological data. Consider the results for a single modeled cell, shown below.

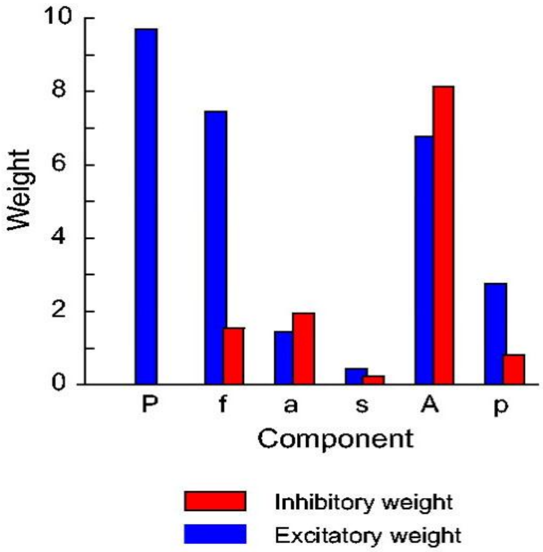

Figure 4.7. Competition model results for a single MT cell. From Krekelberg and Albright (2005).

Each blue line represents the excitatory weight the cell gives to the component labeled beneath; each red line represents the inhibitory weight given to that component. *P* and *p* are components with spatio-temporal energy moving in the cell's preferred direction; *A* and *a* in the opposite direction, and *f* and *s* in other directions. The responses to each individual component were determined by the *ratio* of excitatory and inhibitory weights. It was thus possible for two components with very different absolute values to produce similar responses in the cell. However, when the two different components were combined with *other* components, the absolute values in the weights would matter. This is how the model could account for normal inhibition as well as reverse-phi responses. The two opposite-direction components each produced similar results when presented individually, due to their similar excitation/inhibition ratios: a slight decrease of firing below baseline. However, they produced quite different responses when combined with other components.

Combining *A* with *f*, for instance produced a net inhibition, while combining *a* with *f* produced a net activation, due to the differences in the absolute values of *A* and *a*. Extending this principle allowed the model to account for the reverse-phi results. Consider *P*. If component *A* is paired with *P*, a significant inhibition will result, since the overall effect of inhibition is much greater. However, if *a* is combined with *P*, little change will occur. So, if in a reverse-phi stimulus the component that switches directions approximates the spatio-temporal frequency of *a*, then the physiological response to the stimulus will remain strong.

While still a motion-energy model, Krekelberg and Albright's model posits a very different *kind* of inhibition from traditional motion-energy models. Standard models posit that for each cell there is a set of components for which it has only excitatory weights, and others for which it has only negative ones, the negative influence being subtracted from the excitatory. On the competition model, however, inhibition *precedes* direction-selectivity, in that each component has both positive and negative weights. The standard assumption of subtractive opponency cannot account for the range of responses shown to diverse combinations of Fourier components. So, despite the success of motion-energy models in general, the competition model constitutes a significant amendment. In the next sections, I discuss the motivations for CA in light of these results.

4.4. Against the Anatomical Motivation for CA.

The anatomical motivation makes the following assumptions: (i) that there is a unique structure for each neural area; (ii) that there is a univocal computational description of that area that describes all and only its (functionally relevant) structural aspects; (iii) that the univocal description will functionally distinguish the area from other areas (as is required for successful decomposition); and (iv) that explanations of physiological results (i.e., what are to be “explained and predicted”) occur in virtue of describing that structure’s function in computational terms. I propose to grant assumption (i) for the moment. The burden in this section will be to argue that the remaining assumptions provide neither a good description of, nor good normative guidelines for, the modeling field in MT.

The first fact to note is that different seemingly successful models, which are each considered explanatory in their particular cases, model different aspects of known MT architecture. Simoncelli and Heeger’s model explicitly builds in only relatively few anatomical facts, focusing mainly on input from V1 cells. It does not model the distinction between component and pattern motion cells in MT, nor does it model an output stage. This is in contradistinction to other models, which build in these architectural facts. Nowlan and Sejnowski’s (1995) model, for instance, while still broadly both a motion-energy model and a summation/normalization model, does include these aspects, and in virtue of doing so can account for some motion phenomena that Simoncelli and Heeger’s cannot (namely, stimuli for which both component and pattern cell responses are important). Other kinds of models not discussed here, for instance Koechlin, Anton, and Burnod’s (1999) Bayesian model build in further architecture still, for instance explicitly modeling excitatory and inhibitory connections between MT cells.[50] Importantly, other models describe less architecture. The Krekelberg and Albright model, for instance, measures no MT architecture explicitly *at all*. It simply assumes that the different Fourier components in the stimulus are delivered as input to MT cells *somehow*.

The state of the field thus questions assumption (ii): different computational descriptions (e.g., summation/normalization, versus competition between Fourier components, versus Bayesian inference) describe different aspects of MT architecture depending on the goals of the study. Simoncelli and Heeger, for instance, modeled the aspects of MT architecture (i.e., V1 inputs, pattern but not component cells) that were relevant to the main points of the study, namely describing the receptive fields of MT cells selective for pattern motion. The natural response to this state of the field is to lean on the normative side of the claim. Perhaps different models do in fact describe different aspects of structure, but the *goal* should be to capture all of the relevant ones with a single computational description.

The normative move here, unfortunately, ends up controverting assumption (iii). Consider what would be involved in coming up with *one* computational description of all of MT structure. Presumably, the view would have to incorporate the successes of all of the different modeling projects in describing different aspects of its architecture. One way to do this would be to argue that "computing motion energy" sufficiently describes the function of MT's particular structure, and that the

---

[50] The model was Bayesian because the particular weights of intercellular connections were enforced to capture the conditional probabilities of spatially coincident motions in a training set of stimuli. The model could account for phenomena specifically due to intercellular connections between cells, such as the tendency for activation to "spread" down the length of a moving object.

successes of the particular models are derived from this general success. However, this is patently unhelpful. MT is not the only area that computes motion energy. V1, for instance, also does. And if there is a "decision area" (for instance, area LIP) which uses the MT signal to finalize a perceptual judgment, then that area will be responding to motion-energy as well. Picking out a general description that is intended to cover the successes of other models in describing pieces of MT architecture thus loses the ability to distinguish MT from other areas, a prime goal of functional localization.

It should be noted that this point is part of a more general worry about the role of computational models in functional localization. Currently, there is some enthusiasm for the idea that a "canonical computation" can be found that describe what, at base, *any* cortical circuit does. Summation and normalization, such as that employed in Simoncelli and Heeger's model, is in fact one candidate for a canonical computation (Carandini & Heeger, 2012), although it is not the only one. Summation and normalization models themselves are a subclass of a still more general set of models—linear/non-linear models—which include any model that posits sequential stages of non-linear and linear computations (the rectification step and summation and normalization steps, respectively, in Simoncelli and Heeger's model) . Now suppose that a canonical computation is found, and that *all* neural areas can be said to, at some level of generality, perform the same computation. The description of the canonical computation, then, clearly would not distinguish one area from another, which a type of description intended to localize function must do. This is just a more general version of the problem described with regards to motion-energy models and MT.[51] The situation is as follows: if we want models that in fact explain particular results, we will end up with different specific computational descriptions, describing only part of MT's structure. Attempting to generalize beyond this situation will lose the ability to describe unique MT structure and differentiate it from other areas, which is a prime goal of localization.

The normative move is also in tension with assumption (iv). Consider Krekelberg and Albright's model and its discoveries about inhibition in MT. The fact that each MT cell has both positive and negative weights for each Fourier component is a *major* revision to the traditional thinking about inhibition in motion-energy models. And the discovery was based on very minimal assumptions about architecture. Krekelberg and Albright write: "any model, regardless of its precise mathematical form, in which the response to a composite stimulus is determined only by the response to the components, cannot explain the full Fourier space summation properties of [MT cells]" (2005, p. 2916). The claim about mathematical form could be made just as well about architecture—no matter how much architecture is modeled, the model should not assume that inhibition is entirely due to subtraction of Fourier components. But if this is the case, then the explanatory value of Krekelberg and Albright's model is *not* based on its modeling any specific structural aspects, but is instead a principle of MT function that future more detailed architectural models

---

[51] Chirimuuta (2014) makes a roughly similar point with regards to the relationship between canonical computations and mechanistic explanation.

should attempt to incorporate. The normative claim, which ties explanatory success to describing structure, misses this key explanatory insight.

So, in sum, CA ties explanation too closely to modeling structure to either adequately describe the field of modeling in MT or to provide good normative guidelines. In the next section, I critique the motivation of holding up absolutism as an explanatory ideal in computational neuroscience.

4.5. Against the Absolutist Projectability of CA.

CA posits a single computational operation as inherent to each area. Given the nature of the modeling projects discussed above, and the idea that a functional explanation should explain and predict what an area will do in distinct contexts, CA suggests that given a successful model and new context, we should already be able to describe MT's responses in that context. There are several successful computational descriptions of MT in particular contexts. In this section I argue that the success of even very powerful models does not, on its own, supply explanations or predictions of what will happen in other contexts. Nor is it a reasonable explanatory ideal to hold for these models.

The general lesson that Krekelberg and Albright draw regarding inhibition in MT is that one cannot predict the responses of MT cells to multiple Fourier components in a stimulus just by knowing their responses to the individual ones. Fourier components need to be both measured individually and measured in combination. Krekelberg and Albright point out that they studied only a few Fourier components out of the entire set of possibilities, and that the cells they studied don't exhibit a clear pattern of inhibition and excitation in the components, instead being uniquely characterized. One of the things we want to know, then, is whether there *are* patterns of inhibition and excitation to different components across populations of MT cells, and how these relate to motion perception. The point is that the competition model, fit to the current data, *does not say anything about these patterns*. They write: "Sampling only 6 of the relevant points in [the Fourier domain] makes finding a common pattern of competitive interactions unlikely. To uncover a common pattern, we believe it to be necessary to measure interactions across a large part of Fourier space and additionally for components moving in other than the preferred or antipreferred directions" (2005, p. 2920).[52]

This is true *despite* the success of the model at fitting the data that was actually studied (which, recall, was greater than that of other models). In order to understand full MT functionality for motion stimuli on the competition model, neuroscientists would need to pursue studies of MT cell responses to a greater range of the Fourier domain. That is, they need to pursue a wider range of stimulus contexts, until patterns of Fourier component combination amongst MT cells are uncovered. While the competition model provides a way of analyzing the responses to any particular combination of Fourier components, it does not itself provide the explanatory principles for understanding MT responses as a whole. Nor is it intended to.

---

[52] Krekelberg and Albright note that there are pragmatic problems with the suggestion—namely the "combinatorial explosion" of needing to look at all of the Fourier components and their combination. Inevitably, this will involve some exploratory work. In the next chapter, I give some considerations on how such a search through contexts is generally systematic and intelligible despite not attempting a "complete" account of an area's function.

I take this to be a deep point about modeling and contextual variation. An extremely successful model for some stimuli does not, simply in virtue of that success, explain how the modeled area will respond in all contexts. This is true even *within* motion contexts—a model that successfully predicts what will occur in some motion contexts does not explain or predict what it will do in other ones, i.e. ones with different combinations of Fourier components. I will now show that this situation generalizes. Even for models designed to fit data for a very wide, unconstrained range of stimuli, successful fit does not explain or predict what will occur in other contexts.

Nishimoto and Gallant (2011) sought to extend the account of receptive field and pattern-motion properties of MT cells originally proposed by Simoncelli and Heeger. They had monkeys watch movies of naturalistic scenes (e.g., of animals moving across a prairie), and recorded MT responses across the entire stimulus set.[53] They thus hoped to explore MT responses to, if not the entire Fourier domain, at least an unbiased sample of it. They constructed a version of summation/normalization model that could account for MT responses to the entire range of stimuli.

The model was fit to MT responses over a range of movies, and then tested for a variety of receptive field properties. In particular, Nishimoto and Gallant's MT cells showed a physiologically realistic range of pattern motion responses, suggesting that a summation/normalization model can capture physiological responses in both constrained experimental settings and naturalistic stimuli. This is undoubtedly an

---

[53] The movies were "motion-enhanced", where the "enhancement" consisted of the random insertion of textured objects moving across the screen. The enhancement constrained the movies to meet certain statistics for spatial frequency, which I will not discuss in detail here. It should also be noted that the statistics of studies to natural-stimuli are different, with models generally accounting for less variance in responses than for more controlled stimuli. I will gloss over this detail as well.

interesting result, and it might be thought that having a model that fits an unconstrained data set sufficiently overcomes the problem of modeling limited contexts.

However, when compared with the results discussed in section 4.3, it's clear that the model does not explain what will happen in all possible contexts. Even the "naturalistic" stimuli are in black and white, and two-dimensional, so the model does not say anything about the sorts of robust and functionally relevant responses to depth that have also been discovered in MT. Not only does Nishimoto and Gallant's model not predict any of these effects, but it is not intended to—it is a study of motion receptive fields and pattern selectivity for motion stimuli that attempts to extend models of *those* responses to cover naturalistic stimuli. Understanding MT function as a whole, however, involves not just understanding its responses in motion-contexts, but also in depth contexts. As such, having a successful model for even a relatively unconstrained set of two-dimensional, black and white motion stimuli simply doesn't provide an explanation or prediction of what will happen in further contexts.[54]

One final point needs to be made here. Nishimoto and Gallant only model certain varieties of "naturalistic" stimuli, but they suggest that their model could be expanded to account for other aspects of MT responses, including depth influence. There is nothing unique in this: many other modelers employing different models

---

[54] Nor would it be helpful to model "totally naturalistic" stimuli, which contained all possible stimulus elements. Understanding cell responses to such stimuli would require knowing the statistics of the relevant aspects of the stimulus, and this is often difficult to discern in fully unconstrained settings. Drawing conclusions in such studies often requires having a model already in hand of what the cell is responding to (as was, in fact, the case for Nishimoto and Gallant's)—but, as already shown, a previous understanding of what cells do can be overturned in new contexts. Rust and Movshon (2005) discuss these points in an amusingly titled article, "In praise of artifice," which criticizes overly optimistic use of "naturalistic" stimuli in computational neuroscience.

make similar statements. This sort of claim, however, is entirely distinct from saying that a particular model explains what will happen in all contexts. What one would have to do in order to extend the model to account for how MT represents depth is to pursue physiological analysis of the responses of MT in depth contexts, then build into the model parameters accounting for that influence. But the need to continually explore contexts and modify our functional account when doing so is *exactly the kind of analysis that CA is supposed to obviate.* What we have, even in the case of a model fit to extremely general stimuli, is not an explanation of MT function in all contexts, but a model with enough flexibility that new parameters can be built into it when new contextual responses are uncovered. This is exactly how Krekelberg and Albright interpret their model, and Nishimoto and Gallant say nothing to controvert this reading. CA's claims to overcome contextual variability via computational description, then, are empty—they do not fulfill the goal for which absolutist descriptions are intended.

CA thus fails as either a descriptive or a normative explanatory ideal for modeling projects. Individual models are not explanations that, in point of fact, explain and predict what will happen in all contexts. Nor is it reasonable to expect them to. If even models fit to a very large range of stimuli cannot predict and explain all the ways in which MT responses might vary, then insisting that models do so is holding them up to an unhelpfully (because largely impossible) universal standard. Since the projects in computational neuroscience that are actually working on describing the function of MT are neither described nor normatively guided by CA, it is not a good account of localized neural function in this brain area.

4.6. Conclusion: Explanatory Unification versus Tools for Contextual Analysis.

I have argued that CA is neither a good descriptive nor a good normative principle for computational analyses of localized neural function. That does not mean that the motivations behind CA fail to describe *anything* currently going on in computational neuroscience. Neuroscientists *do* search for very general, powerful computational models, and occasionally hope that a wide range of results can be brought under them, including those describing the behavior of particular areas. The question is just whether these practices are good bases for *localizing* function. I conclude with a discussion of where to fit localization projects in the overall field of computational neuroscience.

In the computational neuroscience literature, one often finds appeals to unificationist considerations. A standard unificationist contention is that an explanation that subsumes more facts is a better one. The appeal of potentially discovering a canonical computation (see section 4.5), or even several, is partially due to the understanding that would be facilitated by seeing *all* neural computations as special cases of one (or a few) general type(s) of computation. There is no doubt that discovering a canonical computation would be a powerful result in neuroscience. The unificationist approach is reflected in the desire, professed by Price and Friston, for a single description that “encompasses” anything that an area might do in distinct contexts, as well as Friston’s (2010) subsequent claim that specific functional accounts should fall under “global” theories.

The unificationist ideal for computational neuroscience would be to understand the computations performed in specific contexts and by specific neural structures in particular contexts as special cases of more general computational principles. Summation and normalization models have been proposed as potential descriptions of a wide range of specific computations performed in the brain, and these can be further classified as a sub-type of linear/non-linear computation writ large. It might be hoped that the computations performed by each specific area could be typed somewhere in an overall framework of kinds of computation, and thus that even the specific behaviors of an area would be interpretable as examples of more general principles.

I will not discuss here whether such an ideal is in fact possible, nor do I intend to critique the unificationist impulse as such. The point I have tried to make is that this kind of project is no substitute for contextual analysis in localization projects, and does not eliminate the need to occasionally expand functional descriptions to include contextual variation. The very general kinds of descriptions that are amenable to unificationist analysis—e.g., that all areas are fundamentally involved in summation and normalization, in linear/non-linear transforms, or in Bayesian inference—do not on their own differentiate neural areas from each other. And even when these models are fit to the responses of areas in specific contexts, they fail to show what the area will do when the context is changed. Expanding or modifying a model to explain what occurs in a new context involves exactly the kind of contextual analysis that the

CA theorist seeks to avoid. That is, when we want to explain what particular neural areas do, we must be willing to change our description when the area operates in a new context.

A view that incorporates the unificationist intuition without making it central to localized function sees models as tools for contextual analysis, rather than as functional explanations that are supposed to meet absolutist strictures. That is, a model provides a way of assessing an area's behavior in a context, in a way that can clarify new functional principles revealed in that context. The more powerful the model, the more flexibly applicable it is. This is clearly much closer to Krekelberg and Albright's descriptions of the competition model than the absolutist ideal proposed by CA. It also arguably makes sense of the use of summation/normalization models in different contexts for different purposes, and the largely promissory claim, standard amongst modelers, that their particular models can be expanded or adapted to fit new contexts in future studies. While more would need to be said about this kind of role for models, reading the field in this way captures the importance of both developing very powerful general models *and* pursuing contextual analysis. Given the descriptive and normative failures of CA, such an alternative has considerable merit.

If neither TA- nor CA-style descriptions are good accounts of localized function, then embracing a contextualist alternative looks to be an appealing option. There has been relatively little development of the contextualist position, and showing that it is indeed the right way to go will require showing that it can meet the assorted goals of a general theory of localized function. I undertake to do so elsewhere

(Burnston, forthcoming). The current considerations have served to motivate the contextualist project. A reasonable meta-induction from the failure of multiple varieties of absolutism is that neural functions *are* context-sensitive—if functions are genuinely susceptible to variation in different contexts, then absolutist accounts are simply unlikely to describe them adequately. If this is the case, then a fully developed contextualist theory is a worthwhile goal for theoretical analysis in neuroscience to pursue.

Chapter 5. A Contextualist Theory of Functional Localization

5.1. Introduction.

If the arguments in the preceding chapters are correct, then absolutism, in any of its forms, is an inadequate theory of functional localization. Recall the desiderata from chapter 1: empirical adequacy, projectability, and decomposition. Recall also that absolutism, at its base, is a commitment that function ascriptions must project across all contexts. The arguments in chapters 3 and 4 can be summarized as follows. TA provides a simple way of decomposing, but that decomposition cannot meet its goals for projectability. That is, the decomposition proposed for areas such as MT, if read with the absolutist commitment to projection across all contexts, are empirically inadequate. CA, as discussed in the last chapter, fails as well: computational modeling projects, in general, are not in the business of attempting to explain all and only aspects of the structure of particular areas, nor do even very successful models manage to explain what an area will do in any given context.

Is there a way of describing functional localization without absolutism? I argue that there is. The problems with absolutism stem from its definitional commitment to projectability across all contexts. I propose to simply abandon this commitment. The result is a theory that embraces context—a "contextualist" theory of function—and which attempts to understand the desiderata in terms of contextually varying properties of particular neural areas. Making the move is simple enough. Arguing that it meets the explanatory goals of neuroscience and avoids collapsing into holism, however, is another matter.

Holism denies localization. So, to avoid it means to give an account that meets the desiderata on a theory of localization. That is what I will focus on here. The arguments in this chapter involve two major moves. The first is to re-construe how we should relate intuitive perceptual features—like motion, color, and depth—to brain functioning. The MFH theory takes these features as basic psychological units, and assumes that individual parts of the brain must be dedicated to processing them. Contextualists, alternatively, view them as parameters whose variation produces new contexts in the environment. This move is key to showing how contextualism can meet empirical adequacy and decomposition. The second major move is to reconsider the relationship between the desideratum of projectability and scientific practice. On the absolutist reading of projectability, a good theory tells us what will happen. That is, the right theory will generalize to any and all contexts of functioning. This is the kind of epistemic ideal to which absolutism is committed, and which I have argued is implausible. But this is not the only way of construing the epistemic import of successful theories. Another way of thinking about theoretical success is the ability of a theory to structure ongoing ingoing, rather than to tell us the outcome of an inquiry in advance. This move splits the notions of generalizability and projectability, which are collapsed in absolutism, into two distinct theoretical roles. Generalizability involves the actual set of cases that are explained by a particular function ascription. Projectability involves the applicability of current theory to future explanatory contexts—in this case, to other perceptual contexts that we might study. I argue that contextualist functional theories have generalizability, if not *complete* generalizability, in that particular function ascriptions often cover a variety of contexts. I also argue that contextualist theories are projectable. They are not projectable in the sense that they *tell us*, in advance, what a brain area will do in a new explanatory context. They are projectable in the sense that they *tell us how to determine* whether a new context can be accounted for within the current functional theory. To explain this type of projectability, I explain how the development of the functional understanding of MT has been informed by what I call "minimal hypotheses."

Minimal hypotheses are meant to allay a deeply intuitive worry about contextual variation and projectability. The idea is that, if we posit that the properties we are studying are *inherently* context sensitive, and that our explanations must change depending on context, our previous explanations will be *falsified* by the new context. That is to say, not only will our current explanation not explain or predict what will happen in the new context, it will be *irrelevant* to that context. We will simply have to abandon our previous explanations and start from scratch when new contexts are explored. Minimal hypotheses are basically a pragmatic assumption about contextual variation, which serves as a null hypothesis when discussing new contexts. The assumption is that, in a new context, the studied brain area will continue to perform the function we already know about, if it functions at all. This provides constraints on when to modify one's function ascription in new contexts—i.e., it provides the resources for determining how generalizable a particular function ascription is, and to what specific cases. It also, I suggest, shows how our theories can be general and powerful without their having to tell us what will happen in all contexts.

I begin, in the next section, by articulating the worries about contextual variation and projectability via an analogy to debates about contextualism in the philosophy of language. In section 5.3 I articulate the contextualist view, and show how it differs from the MFH view in its interpretation of intuitive perceptual attributives. Sections 5.4 and 5.5 show how the contextualist approach can meet the desiderata, and section 5.6 shows how the contextualist view accounts for a variety of other important aspects of neuroscientific practice. Section 5.7 concludes with a brief discussion of future directions.

5.2. Contextualism and Systematic Theory.

Contextualism about a property, X, is the view that the value of X changes depending on the situation in which X occurs. Roughly contextualist views have been proposed for many properties: values, knowledge, belief, morality, and, perhaps most fundamentally, semantics. On these views, the axiological/epistemic/ethical/semantic Xs being discussed categorically change their values with changes in the contexts external to the property.[55] I will discuss semantics here, both because several discussions regarding semantic contextualism mirror the discussion I wish to have about neural function, and because semantic contextualism is in some ways primary to other debates. That is, it is often taken to be evidence for contextualism of other sorts

---

[55] Egan (2012), who defends a contextualist position about value statements, provides a helpful discussion of the important role of semantic contextualism in other debates. Overviews of contextualism in a variety of debates can be found in several anthologies (Bouquet, Serafini, & Thomason, 2008; Preyer & Peter, 2005).

that *statements* regarding the other sorts of properties vary in their meanings, specifically in whether they are true or false, depending on the context of the utterance. In semantics, what are taken to categorically change with context are the *truth values* of statements—i.e., they are true in some contexts, and false in others. In neuroscience, I will suggest that what categorically changes is the *contribution* made to overall system functioning by a particular part of the brain (see the characterization of 'function' in chatper 1). Let me reiterate that I am only establishing dialectical similarity in the debates to clarify what is at stake in the discussion of neural function. I thus will not attempt to describe every facet of the debate here, nor will I attempt to solve any problems that are internal to debates about semantic contextualism.

The main motivation for semantic contextualism is that the truth values of statements *do* seem to change depending on the context in which a statement is uttered. This is intuitively obvious for statements involving indexical expressions such as 'I', 'here', and 'now'. However, some theorists have suggested that the set of context-sensitive statements is extremely large (perhaps even including all statements), and includes statements involving traditionally non-indexical expressions. Travis (1997) asks us to consider the sentence 'The leaves are green', as spoken about a Japanese maple (a kind of tree that normally has red leaves) whose leaves have been painted green. If one is speaking to a photographer looking for green objects to photograph, Travis suggests, the sentence is true. If one is talking to a botanist, the sentence is false. So, the meaning of the sentence is "compatible with various distinct conditions for its truth" (Travis, 1997, p. 89) in different contexts. And since meaning is standardly construed in terms of truth conditions, for the contextualist semantic properties are context-sensitive ones.

Those who deny contextualism are “minimalists” or “invariantists” (depending on which debate one is involved in) and hold a position roughly analogous to absolutism in debates about neural function. They suggest that each term (except for the standard indexical statement) has only *one* meaning, and that it has that meaning regardless of the statement or context of utterance in which it occurs. There are a variety of proposals given by minimalists about what stays constant as context changes.[56] What I want to focus on here is the argument, often advanced by minimalists, that admitting contextual variation in semantic properties undermines the goals of *theory*—that is, admitting context entails such an epistemic sacrifice that we should be strongly motivated to find an alternative.

Consider the following quotes.

MacFarlane:

> The worry is not so much that we’ll have too many parameters, but that there will be no end to the addition of such parameters. The worry is that such proliferation would make systematic semantics impossible. (2009, p. 246)

---

[56] One proposal given by Cappelen and Lepore is that what stays the same is the “disquotational” truth condition of a statement: for instance ‘x is tall’ if and only if x is tall. This proposal rules out contextual variation by claiming that anything that varies in context is not the proper domain of semantic theory. Semantic theory, that is, is not in the business of analyzing facts about the behavior of ‘tall’ in different contexts, such as information about relevant standards for tallness, the metaphysics of properties like *tall*, etc. Recananti (2005) gives a rather strong criticism of this view: it rules out the ability of semantics to explain any of the actual behavior of words and statements. I mention this here because it is at least roughly analogous to the proposal of individuating functions by extremely abstract computational principles, as discussed in the last chapter. One might argue that an area performs Bayesian updating, for instance, might be useful, but if one insists on ruling out any specific information processing as relevant to the function of the area, then one sunders the computational function individuation from any actual processing, a major concession that even CA theorists should want to avoid.

Cappelen and Lepore:

> Semantics is a discipline that aims to characterize systematically certain features of linguistic expressions and to do so in a way that captures general truths about languages, and not just truths about particular speakers in specific contexts. (2005, p. 58)

Stanley:

> My concern … is that the suggestions I am aware of for dealing with the additional complexity [entailed by contextual variation] essentially end up abandoning the project of giving a systematic explanation of the source of our intuitions. (2005, p. 222)

These concerns mirror the quote from Rathkopf, discussed briefly in the previous chapter: "there could be no grounds for claiming of any list of task-bound functions that it is complete. … Any functional contribution to behaviors yet to be taken into account would demonstrate the incompleteness of the list" (2013, p. 10).

The structure of the debate between contextualists and minimalists in the case of language is thus similar to that between contextualists and absolutists in the case of neural function. On the one hand, significant data exists that the properties of interest (in this case, the functional properties of brain areas) vary with context. On the other, there is sentiment that admitting this variation is too much of an epistemic sacrifice—we should explain it away somehow rather than admitting it.

Rathkopf, at least, doesn't *argue* for this conclusion. It is a straightforward assumption in his argument that contextual variation and projectability are incompatible. Proving the argument invalid, then, requires showing that there is some interpretation of contextualist theory that provides a reasonable analogue of projectability, but also accounts for the data. That is what this chapter will accomplish.

5.3. Contextualism.

The contextualist view states that the function of a neural area changes with context. That is, our description of a particular area such as MT must be able to change depending on the context which we are discussing (e.g. motion versus depth contexts). By a “contextualist account” of the function of some area, I mean a list of contexts and the functions performed by the area in those contexts. So, contextualism can be formulated as follows.

- *Contextualism*: for each neural area A, the appropriate functional characterization of A includes an open disjunction of functional properties, D, where each disjunct in D is indexed to a type of perceptual/behavioral context.

Contextualism argues that the particular features that a brain area represents in its physiological responses, as well as the role that those representations play in performing behavioral tasks, is dependent on the environmental context and the kind of task the organism is performing. That is, what contribution is made by a particular brain area to overall system functioning varies with context, and our theories must therefore take context into account in attributing functions. I will focus mostly on environmental context here. A context, in this case, is the combination of perceptible features given as input to the brain area, where this is constrained by the anatomy of the area and the relevant aspects of the environment (see chapter 1).[57] A type of context is one in which certain arrangements of perceptual features produce distinct responses in an area or its component cells.

To flesh out contextualism, consider the difference between how contextualism and absolutism treat intuitive perceptual features. It is natural to think that our everyday concepts of perceptual features are the basic perceptual categories, and that perceptual processing should be organized in such a way as to distinguish those features from one another, representing them independently (see chapter 2). The MFH view takes this intuitive picture and applies it to the workings of the visual system. Contextualism rejects this picture, and instead views intuitive perceptual features as parameters that can vary in the formation of specific perceptual contexts, which often (though not always) comprise combinations of these parameters at specific values. All contextualist views are committed to some characterization of the parameters that vary in specific contexts—in contextualism about knowledge claims, it is the stakes that are relevant in the epistemic situation; in linguistic contextualism, it is such factors as speaker intentions and the background knowledge of the listener. In perceptual neuroscience, the particular values of perceptible features are what vary, and what determine the information that a particular perceptual area responds to (along with, of

---

[57] One might object to this unargued importation of a notion of ecological relevance. However, in psychology and neuroscience it is vitally important to have some working notion of the ecological function of an object of study, in order to rule out rampant dispositionalism. We study the visual cortex in situations where it is presented with visual stimuli, not, for instance, when it is removed from the brain and kicked down a hill. There is ample evidence that the visual cortex is heavily involved in vision, and that the important parameters for vision include the ones I have discussed. An important aspect of contextualism is that it takes the notion that visual cortex serves a perceptual function as *defeasible.* We could always discover more relevant parameters, for instance if we were investigating multimodal influence in perception, or the interplay between perception and action. The view I give below makes sense of this gradual expansion of contextualist understanding.

course, the anatomical connections that provide the area with input regarding those features). So, perceptual contexts are determined by the combination of cues for certain features that are available in the scene—some of the ones we have discussed include those for disparity (depth), displacement (motion), wavelength (color)—and their arrangements (e.g., into a RDS or a moving plaid). The relevant *type* of context is uncovered during the course of investigation, in the discovery of particular combinations of stimulus parameters that produce a categorically different effect in the physiological responses of a given object of study (area, group of cells, etc).

Part of the import of contextualism in perceptual neuroscience is that it captures the particular epistemic practice in the field, and how that practice reveals new aspects of function. One of the specialized talents of perceptual scientists is combining stimulus parameters in distinct ways to reveal new types of contexts. Consider some of the ones we have discussed in the previous chapters. Within motion stimuli based solely on luminance, we have seen that RDS of assorted correlation values have been used to measure pattern motion responses, and a variety of overlapping moving stimuli, with varying degrees of transparency, have been used to test opponency and motion segmentation. Importantly, combinations of luminance and color parameters were employed by Dobkins, Stoner, and Albright (1998) to test moving stimuli whose pattern versus component motion is determined by color values. Similarly, depth contexts were explored via stimuli implementing a variety of disparity values (both relative and absolute), both individually, in combination with each other, and in relation to motion. These investigations have revealed new types of contexts in which MT functions—stimuli arranged in depth without motion, rotating objects (e.g., cylinders) whose direction is disambiguated by depth cues, etc. MT's responses in these contexts are explained as the representation of a particular parameter value or combination of parameter values, as appropriate.

This discussion elucidates a number of important points about how the methodology of perceptual neuroscience relates to absolutism. Notice that it is absolutely vital to vary the different parameters individually, without the others—one important type of context involves stimuli in which only a particular type of information is available. It is also natural that contexts of this sort should have been the first explored in detail. This is the kind of sentiment expressed by Zeki et al. (1991), who argue that the right methodological approach for uncovering function in perceptual areas is to expose them to variation along only one parameter, while holding the other elements of context fixed. Contextualism is not only compatible with this methodology, it takes it as extremely important. The absolutist error is not in being interested in contexts of this type, but in assuming that what is uncovered in these contexts is *the* function of a brain area, such that any other contextual variation must be explained away according to one of the strategies suggested in chapter 2. In fact, this is not even an in-principle error. It might have turned out that the absolutist description of the visual system was correct. The fact that (in my view) it hasn't is not a criticism of the original researches into, for instance, MT, but instead a spur to move beyond absolutism and into a contextualist account.

Take the case of MT and color. The fact that, when the color parameter is varied independently of other parameters, MT cells do not exhibit significant responses, is an important discovery. On contextualism, however, it does not mean that MT never responds to color information in a functional way; it means that there is a *type of context*—namely, contexts in which only color information is prevalent in the receptive field of MT cells—in which MT does not offer a functional response. This doesn't say anything about whether there are other types of contexts in which color information might play a role in MT function, and this is exactly what Dobkins, Stoner, and Albright (1998) uncovered.

According to contextualism, the appropriate functional description of an area is a disjunct of the types of contexts in which it has been shown to function and the particular representations it employs in those contexts. For instance, MT represents pattern motion in stimuli with symmetric luminance and color contrast, and component motion with asymmetric contrast. According to contextualism, these are two important disjuncts in the overall function ascription, which are not in conflict with other disjuncts for, e.g., depth representation.

I will now discuss how contextualism meets the desiderata. First, however, a word about behavioral contexts. As discussed in chapter 3, the relationship between perceptual and behavioral contexts is complex. We saw there that occasionally a behavioral task can disambiguate information that is carried at a particular area—for instance, motion information can occasionally be used despite the MT cells that carry it also being affected by depth. So there is not a one-one relationship between perceptual and behavioral contexts. As also discussed in chapter 3, there is a significant empirical question about how these perceptual and behavioral contexts interact, regarding which significant progress is needed. Focusing on perceptual contexts here, however, is justifiable. Consider Uka and DeAngelis's (2004), for instance, which explored how monkeys learn to perform depth tasks based on cells that respond to both depth and motion, it was important to know that MT carries information about motion and depth in contexts when both are present. Understanding the relationship between perceptual and behavioral contexts requires understanding each independently. I will thus focus on the perceptual side, with the caveat that eventually behavioral context will need to be considered in more detail.

5.4. Decomposition and Empirical Adequacy.

First, consider decomposition and empirical adequacy. Decomposition is the goal of distinguishing parts of the brain according to their functional properties. Contextualism posits that what distinguishes areas from each other is the particular informational functions they perform over a range of contexts in which they are active. MT, for instance, taken as a whole, can be said to represent motion, coarse depth, and fine depth (and combinations of those features) depending on the stimulus context. Read in this way, there is ample evidence that decomposition can be met. I will here briefly compare current physiological data on MT and V4 to show that, even if they are not individuated by each responding only to a particular feature, they still can be distinguished from each other.

The MFH implementation of TA is based on proposing that each visual area represents information about a particular perceptual property or feature. I have argued already that MT does not meet this kind of description. Now consider the summary description of V4's physiological properties from Roe et al. (2012): "V4 comprises cells that exhibit diverse receptive field preferences related to surface properties (color, brightness, texture), shape (orientation, curvature), motion and motion contrast, and depth" (p. 1). Offhand, not only does V4 not appear to be a feature-specific area, but the features it responds to have significant overlap with the featural responses of MT.[58] Recall, for instance, that even after MT depth responses were acknowledged, it was generally thought for a period of years that all fine depth (relative disparity) responses were located in V4. Sanada et al.'s (2012) and Krug and Parker's (2011) studies showed that this is not the case—MT cells respond to fine depth as well as coarse depth, and thus to the tilt and slant aspects of a stimulus. Similarly, color information affects MT responses in moving plaid stimuli, as Dobkins, Stoner, and Albright's (1998) study shows. MT is thus not "color blind." So, it looks like, if individual features are what we use to individuate the functions of areas, MT and V4 will not be clearly dissociable. This, I believe, is the garden path that leads to holism.

---

[58] To actually establish that V4 is not a feature-specific area would require undertaking the same kind of argument that I made regarding MT in chapter 3, and I will not attempt that here. I believe that such an argument can be made. It is also worth noting that Roe et al., despite discussing the many different feature interactions in V4, prefer to interpret the evidence in feature-specific terms, largely relying on the differentiation strategy discussed in chapters 2 and 3. I regard the following sets of statements to be vaguely inconsistent: first, Roe et al. claim that psychophysical considerations "suggest[] the presence of significant interactions between representations of different feature modalities (e.g., color and depth) in V4" (Roe et al., 2012, p. 16); then, however, they suggest "that … feature representations are tied to feature-specific domains within V4, [and] that domains of shared feature selectivity are anatomically and/or functionally linked into feature-specific networks (Roe et al., 2012, p. 24). That such positions are taken shows how deeply the MFH implementation of TA has taken hold in perceptual neuroscience, and how resistant some theorists are to viewing it in heuristic terms. The contextualist perspective I advocate is an attempt to provide an alternative.

If one treats ‘depth’ or ‘color’ as a basic functional category, and discovers that MT and V4 both “process depth” or “process color,” then it will appear that there is no functional dissociation between them. Further, if one believes that this lack of distinction is widespread, the result will be holism, or something close to it.

However, contextualism does not individuate functions by treating perceptual features as basic categories, but instead treats features as contextually varying parameters. On contextualism, it is perfectly fine for different areas to respond to some of the same features, so long as the range of contexts in which they do so differs. There is strong evidence that this is the case. Consider motion first. Most MT cells respond to motion in a variety of contexts, with specific velocity tuning curves, as well as pattern indices, that characterize the aspects of motion that drive the cells’ responses. Interestingly, a variety of motion influences have been found in V4 as well. What is important is that motion responses in V4 are considerably different in terms of the contexts in which they occur than are those in MT. For instance, it has been shown that roughly 33% of V4 cells show direction-tuning after sensory adaptation to a motion stimulus (Tolias, Keliris, Smirnakis, & Logothetis, 2005). Moreover, it is well known that a significant proportion of motion-sensitive cells in V4, maybe 10-15% of V4 cells in all, respond to kinetic edges in the stimulus—these are edges at particular orientations that are defined entirely by movement, e.g., within a RDS. Importantly, MT does not respond to kinetic edges. Moreover, so far as I am aware, no studies have established that V4 cells that do respond to motion exhibit pattern indices. Read in a contextualist sense, there is no problem in distinguishing

MT and V4's motion responses, because they respond to different motion contexts—e.g., those in which the motion forms a kinetic edge, and those in which they don't. Of course, there are many more motion contexts than just these, and there is no reason to suspect from the evidence that the specific contexts in which MT and V4 cells will respond to motion will co-align.

One can similarly distinguish V4's and MT's responses to color and depth. The contexts in which MT responds to color, so far as the current evidence goes, are extremely limited—the responses are dependent on color signals that segment moving stimuli. They may or may not vary depending on *what* colors are used to segment the scene. Recall that it was the sameness or difference of luminance for similarly colored grating lines that was vital for the effect. If this turns out to be the extent of color influence, then this means that the color responses of MT are restricted compared to V4 responses, where responses reflect categorical color assignments and exhibit color constancy (as discussed in chapter 2). That is, V4 cells exhibit a range of responses to contexts involving color information that differs from those in which MT responds. The fact that color affects MT responses as well as V4 responses does nothing to undermine a functional distinction between them. Finally, consider depth. On the current evidence, both V4 and MT cells respond to fine depth in certain contexts. In particular, MT cells are selective for the tilt and slant of the stimulus. V4 cells, however, exhibit a somewhat different depth response. In particular, V4 cells respond to curvature, whereas MT cells do not (Connor, Brincat, & Pasupathy, 2007). As Roe et al. point out, the combination of curvature selectivity and relative disparity selectivity allows for the perception of protuberances—parts of an object that extend towards the perceiver. Since MT cells are not curvature-selective, they are not particularly good for this aspect of perception, even if they are helpful for overall tilt and slant perception. The difference between the two in terms of their depth responses is *not* based on responding to fine depth or not responding to fine depth, but instead on the particular combinations of depth with other features—i.e., depth occurring in different contexts.

These are only a few examples; more could be added. The point to be made here is that a contextualist theory easily captures these patterns of variation, and builds them directly into the notion of decomposition. It is entirely sensible that a V4 cell should represent curvature solely in some contexts, depth solely in other contexts, and protuberance in contexts where both features are combined. Absolutism, as discussed, has problems dealing with this kind of variation. Thinking of decomposition in an absolutist way leads to the intuition that it is lost if areas respond to the same kinds of information.[59] Contextualism, I contend, provides an alternative. Read in a contextualist way, not only is there evidence for the decomposability of neural areas, but the evidence is rather preponderant. Thus, if we accept the contextualist reading of decomposition, there is currently no evidence that the brain cannot be decomposed.

There are two other points to be made concerning decomposition *within* an area. First, as argued in chapter 3, it is unlikely that the differentiation strategy will

---

[59] Notice that, while this does involve getting more fine-grained about the information represented, it is not *just* doing so. It is capturing real contextual variation in the behaviors of the parts of the brain. For instance, the very same cells MT cells will, in some contexts, respond solely to motion information, and in some cases to elements of disparity in the stimulus, and in some cases to both.

save absolutism for MT—one is unlikely to find significant parts of MT that respond only to motion, for instance, and never to depth. While I have not argued for it, a similar claim could be defended about V4. Consider, for instance, the fact that motion-selective cells in V4 are not isolated to a specific area within V4. If indeed 30%, or even something in that range, of V4 cells respond to motion (in particular contexts, of course), then those cells will certainly also have responses to the more traditional V4 features as well. Similarly, even if form and color *are* spatially dissociated in V4 (as according to the standard MFH reading discussed in chapter 2), this does not mean that form and color are the *only* things the cells in those divisions respond to. Motion, depth, and texture are other likely candidates.

On the other hand, the lack of feature-specific differentiation within areas does not mean that contextualism rules out meaningful decompositions of *any* type within areas. Consider such results as the discovery of disparity-selective cells distributed in an orderly fashion across MT, and the specific cue combinations for tilt in MT cells. A natural project is to consider whether particular clusterings of selectivity for relative disparity (as opposed to absolute disparity, and across different cue type) exist in MT—that is, whether there are particular cue-conjunction selectivities that are spatially grouped. Hypothetical discoveries along these lines would be important, not only as a clue for understanding the architecture of MT, but also for insight into which conjunctions of cues are important for perceiving (e.g.) relative depth. This kind of decompositional insight does not necessarily depend on feature-specificity or absolutist function ascription.

Finally, we do not necessarily give up the meaningfulness of the labels ‘MT’ and ‘V4’ themselves in embracing contextualism. An opposite worry to the claim that contextualism can’t allow decomposition within an area is the worry that contextualism supports *too much* decomposition. The concern here is basically that, if context is admitted into function ascription, then *some* contextual variation can be found to differentiate any two cells. Thus, the proposed problem is that contextualism does not allow decomposition into large brain areas, but forces it down to the level of individual cells. This worry is misplaced. Abandoning feature specificity does not entail that there is nothing important about comparing MT and V4. Indeed, the discussion of decomposition above suggests that the two areas are meaningfully distinguished, specifically in terms of the particular contextual variations of the cells that they contain.

Further, there are a variety of non-physiological ways of individuating areas, specifically by appealing to anatomy. This practice, of course, is as old as neuroscience itself. As shown in the discussions of Felleman and Van Essen in chapter 2, anatomy can be used to look for patterns in structural connectivity that can guide functional analysis. This extends the traditional process of looking for cytoarchitectural differences between areas as a clue to functional decomposition. A powerful modern application of this idea, discussed briefly in chapter 1, is graph-theoretic analysis (Sporns, 2011). One standard measure of connectivity employed by graph theory is to group areas of the brain into “modules” that have greater patterns of connectivity between their members than externally. Often, these align with areas identified by traditional physiology and lesion studies.[60]

With decomposition established, empirical adequacy comes along in rather short order. Notice, first, that contextualism *retains* the empirical success of the MFH theory, in a more flexible way that can accommodate new discoveries about contextual variation. No one should disagree that MT, for instance, is *a* motion area, an *important* area for motion, or even that it is *more* heavily motion selective, in general, than V4. These claims are still valid, and contextualism provides an easy reading for them. As discussed above, the range of color contexts contributing to MT function is relatively limited. Similarly, it is at least possible that the range of motion contexts affecting V4 responses could turn out to be similarly limited. Contextualism can read the relative prevalence of MT and V4 for motion and color as a difference in the *quantity* and *diversity* of motion/color contexts that affect the areas. It is important for the empirical adequacy of contextualism that the standard results in favor of the MFH view can be brought along in this way. However, since contextualism admits the possibility that exploration in new contexts can reveal new functional responses within MT—i.e., a particular set of functions and contexts is not meant to be "complete"—

---

[60] There are other measurements as well, including identifying "hub" areas that exhibit particularly high levels of centrality—that is, they are often in the shortest path between any two areas, defined anatomically. On these measures, V4 exhibits a significant hub-like structure, whereas MT is considerably less hub-like in its connectivity patterns (da F. Costa & Sporns, 2006). These divisions are not based on or committed to any previous view of the functions of these areas, but often pick out divisions that may have functional relevance. If, for instance, it is true that V4 is a major hub node in the visual cortex, there is reason to investigate the particular patterns of contextual variation of cells which comprise *it*, to see what physiological importance its hub-like structure may have. As such, contextualism supports and intersects with a variety of goals related to decomposition, including decomposition at multiple levels, without needing to rely on absolutist function ascriptions.

then new discoveries can be incorporated into the functional description of MT without sacrificing empirical adequacy.[61]

5.5. Projectability and Avoiding Holism.

The projectability of a theory can be decomposed into itself two aspects: (i) we want a theory that is *general*, in that does not apply to only one case, but to a range of cases, and (ii) we want a theory that tells us something about the cases we have not yet explored. Absolutism, of course, assumes that these goals can only be met by finding a single univocal description—that is, we must have a description that applies to *every* case, and therefore tells us what will happen in *every* case that we've yet to explore. The arguments in the preceding chapters have been at pains to argue that reading projectability in this way is a false hope. But does contextualism offer an adequate replacement? In this section, I argue that it does.

The argument for aspect (i) starts from contextualism's ability to incorporate the successes of the MFH theory. I have not argued at any point that the description of MT as a motion area is not a powerful generalization—indeed, chapters 2 and 3 showed that the strategies available to someone attempting to uphold the standard feature-specific description of MT function are very powerful, and can be applied successfully to quite a range of cases. Even in the case of color influence on MT, as discussed in chapter 3, some of the contexts where color might seem to exert an influence on MT are in fact interpretable in terms of luminance influence in the M-pathway.

---

[61] Importantly, this does not mean that contextualism is "unfalsifiable" in any interesting sense. Posits about particular contextual functions, and about the contexts which distinguish individual areas are highly falsifiable. See also the discussion of minimal hypotheses, below.

Read in these terms, aspect (i) is rather obviously non-problematic for contextualism. There are many contexts and instances in which representing motion is a perfectly reasonable functional ascription for MT. Moreover, the added functions of representing pattern motion in color-defined plaids (as opposed to component in black and white settings), representing coarse depth in stationary and moving objects, and representing fine depth in objects exhibiting tilt and slant, each cover a range of cases. That is, there are many different instances in which objects are segmented by depth or color, in which they exhibit tilt and slant, etc. Of course, contextualism by definition does not entail that *every* possible instance that can be described as one in which, for instance, coarse depth will be represented by MT. It is possible that there are further contextual factors that might mediate or modify what MT does in that particular context. That does not mean that representing coarse depth is not an important generalization about what MT does, as the field has come to appreciate.

The worry, then, seems to really focus around aspect (ii). As discussed above with regards to contextualism in philosophy of language, there is a concern that, once contextualism is admitted, we are theoretically adrift. I interpret the concerns about “systematic” theorizing in the following light: the claim is that, once we admit that our accounts of some property X can vary with context, we have no basis for hypothesizing about contexts beyond the ones that we have, in fact, explored. Not only would we not expect our current views to *tell us* what would happen in the new context, as absolutism requires, but we would not even know how to start investigating the question. We would be at square one.

The worry about systematic theory deeply misdescribes the development of physiological investigation into MT. As interest has expanded to include MT function in new contexts, previous accounts have not been irrelevant for exploring those contexts, but instead have shaped experimental analysis and hypothesizing at every turn. Importantly, this has been true *both* for the original MFH function attributions and for new, contextually expanded attributions, as I will describe in the remainder of this section. Previous functional accounts have shaped ongoing analysis by providing what I call *minimal hypotheses*, which are a kind of null hypothesis for contextual analysis of function. The idea is that one starts from an already accepted functional description, D, which has been proven adequate for explaining the function of area A in a variety of contexts (e.g., motion perception in MT). Given a new context C incorporating novel feature parameter F, the minimal hypothesis is simply the pragmatic, defeasible assumption that D will continue to hold in C, if A functions at all in C. While this sounds rather simple, the important thing to note is that the minimal hypothesis *sets the conditions* for whether one should modify D to include F in the case of C. That is, one should exhaust the options for interpreting the behavior of A in C in terms of the original D before adding a disjunct to one’s functional understanding of A, indexed to C. From this perspective, the strategies discussed in chapters 2 and 3 for maintaining feature specificity can be construed as attempts to uphold the minimal hypothesis that MT represents only motion. I will now briefly revisit the results discussed in chapter 3, and illustrate the role that minimal hypotheses have played.

The motivation for Dobkins and Albright to undertake the study of color responses in MT was converging evidence from anatomy and psychophysics that color information could affect motion representation (both in terms of perception of moving colors, and color segmentation of motion stimuli). Since MT was already posited to be a key area for motion perception, these results provided some reason to doubt that MT had no functional responses to color, as the MFH view of MT held. Hence, Dobkins and Albright undertook to find out whether there was influence of color on MT, and if so, what sort. Importantly, as shown in the initial study, their original hypothesis was that MT function could largely be explained in a manner that did *not* involve positing categorical color influence. The more conservative hypothesis in that study was precisely based on the standard notion of MT function, namely the MFH one. The minimal hypothesis is thus the hypothesis that, in new contexts, an interpretation can be found within the already-accepted view of an area's function. In their initial study, involving color-contrast grating stimuli, the minimal hypothesis was *not* overturned—the seeming influence of color on MT function could be attributed to color borders or luminance and not to color identity.

Just as importantly, however, Dobkins and Albright did not take the question as settled just because the minimal hypothesis could be maintained in one stimulus context. Given the particular interpretation that they had used to maintain the minimal hypothesis in the case of moving color gratings, they next asked whether there were contexts in which this interpretation would be unavailable, that is, cases where color-identity is the only factor that could explain a response. The grating stimuli that were segmented by color (into overlapping gratings or a single moving plaid) were designed with this goal in mind, and the discovery of genuine color influence in MT was specifically due to this manipulation. That is, since color borders did not disambiguate component from pattern motion in the stimulus, only color identity could explain the shift towards representation of pattern motion by MT cells in the overlapping gratings. Importantly, out of this process we do not come up with an attribution on which color *as such* affects MT, but one on which color identity in *specific contexts*—namely as a cue for object segmentation—influences MT. As discussed in the previous section, the specificity of the contextual attribution is part of what distinguishes MT from other areas that respond to color identity. Equally importantly, the researchers were *not* wandering in the dark in exploring the contexts. The minimal hypothesis shaped the experimental manipulations as well as the interpretation of the results. The result was scientific progress. Since the minimal hypothesis provides a rigorous condition for when to add a disjunct to the functional description of an area, and structures

A similar story occurred in the depth case, and importantly continued even *after* the consensus description of MT's function had been expanded to include depth representation. Recall that Maunsell and Van Essen, while noting the responses to disparity by MT cells, argued that no representation of depth is present there. Recall also that this reasoning was based on the idea of MT as purely a motion area. If depth representations are to occur at MT, they suggested, then it must be *motion* in depth that is represented. Failure to find responses indicating representation of motion in depth, they concluded that depth information was purely a modulatory effect on a motion signal, thus maintaining the minimal hypothesis. While it took a number of years for researchers to pursue it, the natural response to this interpretation was to look for contexts in which depth information could be teased apart from motion information, to see whether functional responses to depth could be teased apart from those to motion. The flood of interest in MT depth responses consisted in a variety of ways of doing just that. To recap briefly: DeAngelis et al. (1998) independently varied motion and disparity, establishing that MT had organized representations of disparity in the RFs of its cells, and that these representations were relevant for depth tasks. Palanca and DeAngelis (2003) established that many MT cells respond to disparity entirely in the absence of motion, and that the tuning curves for disparity were not correlated with particular tuning curves for motion. The independent manipulation of a depth parameter, contrasted with motion parameters, and the dissociability of MT responses to each, was inspired by the minimal hypothesis (which Maunsell and Van Essen supported) that MT only represents motion. The discovery of a type of context in which motion is not present at all, but in which MT still shows functional responses to depth is the strongest result speaking against the MFH view. But it is strong precisely because it most successfully resists any strategy for interpreting the results in accordance with the already-accepted description of MT as a motion area.

Importantly, the acceptance of coarse depth representation as a functional role for MT in certain contexts did not lead to unbridled expansion of function ascriptions. Instead, the inclusion set up a *new minimal hypothesis* regarding MT function, namely that MT represents either motion or coarse depth. Recall that the consensus that developed in the mid-2000s regarding MT function was that it could represent coarse depth, but not fine. The eventual questioning and overturning of this view proceeded along similar lines as the original expansion of the function ascription to include coarse depth representation. The Krug and Parker (2011) study discussed in chapter 3 independently varied absolute and relative disparity, and showed that a significant population of MT cells have reliable responses to relative disparity. Similarly, Sanada et al. (2012) discovered significant responses to relative disparity in contexts involving tilt and slant. Just as with the original addition of a disjunct for coarse depth, the addition of one for fine depth involved overturning a minimal hypothesis—in this case a disjunctive one suggesting that MT cell responses could be interpreted as representing *either* motion or coarse depth.

For current purposes, the vital part of this development is that the process of starting from a minimal hypotheses and exploring contexts that may require abandoning was the same in the two developments—one from the consensus motion-only description of MT to a motion-and-coarse-depth interpretation, and the second from the motion-and-coarse-depth interpretation to one that included fine depth. That is, in terms of approaching new contexts, *we are in no worse a position regarding experiment and theory in neuroscience after our functional understanding of an area has been changed to a disjunctive one, than we were while holding a univocal one*. In fact, the epistemic situation is largely the same when moving from a univocal description to a contextualist one as it is when moving from an already-contextualist description to one involving further contextual variation. In each case, the minimal hypothesis begins from the accepted attribution, and explores new contexts in the search for ones that would force adding new disjuncts to the ascription. The advancement of the field has been perfectly "systematic," in the sense that previous views of function shaped and guided subsequent analysis and new functional insights.

A potentially larger worry consists of how to individuate and count contexts. I have mostly relied on the developments of the field to show what the explanatorily relevant contexts are. The worry is that if there are not principled ways of counting contexts such that they can be reliably sorted into types, then generalizability will be threatened—every conceivable difference in context, no matter how miniscule, will be counted as a new type, and we will only end up with disjuncts that cover maximally specific contexts.

I will not be able to give a completely satisfactory answer to this concern, since understanding context is an ongoing project in perceptual neuroscience, for which the ultimate outcome is not already determined. What I can do, however, is show why the examples I've listed provide resources to show why the context-counting problem is not as worrisome as it might seem. Recall from chapter 1 that the parameters that go into contexts must be determined according to the particular explananda and empirical facts about the system being studied. Much of perceptual science is involved in uncovering the aspects of stimuli that make a difference in perceptual phenomena, and the current studies are no different. If we are willing to accept motion, depth, color, etc., as parameters, then it is not problematic to talk about *depth-without-motion* contexts, or *motion-disambiguated-by-depth* contexts. Same goes for color, etc. Moreover, parameters can be combined in different ways in the course of exploring a particular property. For instance, in the tilt study of Sanada et al., both relative disparity and relative velocity were used to define tilt properties. But since these are different parameters according to the field, it also does not seem problematic to discuss relative depth being represented in two types of contexts—relative depth as defined by disparity and as defined by motion. That MT represents tilt in *both* of these circumstances was a significant aspect of Sanada et al.'s findings, and an easy way to explain this significance, as I've argued, is that MT tilt responses range over several types of contexts.

Similarly, the very fact that, e.g., in Dobkins and Albright's study, it was considered up-for-grabs before the study whether motion-segmentation by color could be accounted for with a motion-only interpretation of MT implies that there is a categorical difference in the context between luminance-defined and color-defined motion. What the ultimate nature of this categorical difference is, and how to ultimately count contexts, is a question for the field to continue to explore. But showing that context is vital for functional explanation, and that contextualism can meet the desiderata, does not require an ultimate answer to the question. What it requires are examples in which analysis that is plausibly described as contextualist produces explanatorily potent accounts of function. That is what I hope to have established here, with the caveat that complete understanding of perceptual and behavioral contexts is a project that must continue.

A corresponding problem to the context-counting problem just discussed in what we might call the "many contexts problem." The concern is that there are just *so many* contexts we might study, and any particular ones we decide to work on are just random points in that (perhaps) limitless space. So, even the accounts of projectability and generalizability I have offered are not very satisfying counsel against despair. For instance, in discussing the role of intuitive perceptual properties in determining perceptual contexts, and defining MT function in terms of these contexts, haven't I imported the assumption that MT is a *perceptual* area? And a *visual* area to boot? What right have I to make those assumptions? And even if they are right, why should we expect that they will completely describe MT's function?

There are a few things to say here. First, the fact that there are many contexts does not mean that search through that space is unrestricted. One way in which neuroscientists restrict the space is to employ multiple distinct methodologies, and to use each to motivate search in the others. As mentioned, the primary reason Dobkins and Albright undertook their investigation into color and MT was the emerging anatomical and psychophysical evidence that color and motion might not be so segregated as was standardly assumed. And this kind of influence goes in the other direction as well; given the physiological evidence of detailed depth representation in MT, there is currently a project underway to determine the precise anatomical pathways that mediate this function (Anzai, Chowdhury, & DeAngelis, 2011). Viewed this way, we can suspect that further advances in the other methodologies will help to shape the way that functional investigation is pursued in electrophysiology and related projects. I have already discussed, at several points, the resources provided by graph theory, which measures the patterns and types of connectivity between parts of the brain using categories that are not, themselves, functional. Since graph theory is primarily based on the notion of connectivity, it is primarily concerned with relationships *between* parts of the brain, rather than what they do individually. As discussed in chapter 1, Sporns (2011) posits that graph theory provides ways of thinking about how functioning in the brain is "integrated" between different parts. But function must also be "segregated." To fully understand integration, it is necessary to have some understanding of what the individual parts whose functions are being integrated. Contextualism, I submit, is better placed that absolutism for pursuing the segregation side of the project, given the heavy degree of incerconnectedness between different parts of the brain.[62]

Second, the fact that I have characterized MT as a visual area, and used visual parameters as a way of articulating the description of its function, *does* reflect an assumption that it is a visual area. But this assumption is valid and well supported. The discovery that there is an identifiable visual cortex was one of the great neuroscientific advances of the early 20th century (Bechtel, 2008), and it is in no way unreasonable to take visual parameters as a starting point in understanding MT function (as the field has in fact done). In keeping with contextaulism, however, the fact that MT is a visual area does not guarantee that it might not contribute to other kinds of functions as well. For instance, there is ever-increasing evidence of a large

---

[62] Contextualism might also provide a corrective to those tempted to infer functional holism from graph-theoretic results (Silberstein & Chemero, 2013).

amount of interaction between different perceptual modalities, both at the psychological and neural levels (Stein & Stanford, 2008). It is entirely possible that MT contributes to multisensory interaction involving vision and other modalities in certain contexts. And, should such interactions be hypothesized, projectability based on minimal hypotheses should be applicable in investigating them; one should exhaust purely visual interpretations of MT before expanding its function ascription to count it as a genuinely multisensory area (in certain contexts, of course).

Ultimately, the many contexts problem will need to be solved with contributions from behavioral science. The constraints on what contexts are important are determined by which ecological factors are relevant for the production of functional behavior. Obviously, the full answers are far off. But any view of explanation in neuroscience that paints it as a near-complete science is fundamentally misdescribing the field. That there are many fundamental questions remaining to be answered in neuroscience is a fact that practitioners know all-too-well. The point of contextualism is not to itself answer these questions, but to show how a particularly fundamental explanatory aim—that of dividing up the brain into functional parts—can be pursued even in the face of widespread contextual variation.

Absolutists, like anti-contextualists in other areas of philosophy, will likely respond at this point with the charge that the contextualist has changed the question. The account in terms of minimal hypotheses and systematic progress is *not what the anti-contextualist means* by systematicity, projectability, etc. There is unlikely to be a non-question-begging conceptual argument in favor of one position over another.

Fortunately, in the case of neural function, we have other desiderata that help to tip the scales. The massive benefits of contextualism with regards to decomposition and empirical adequacy weigh strongly in its favor. If contextualism can offer a reasonable view of projectability that resists the standard idea that embracing contextualism has disastrous epistemic consequences, then the view succeeds, and should be accepted.

There is an inherently pragmatic element to successful theory on the contextualist view, and this is very much in keeping with viewing theories as heuristics—part of what makes a theory a good one, on the contextualist view, is that it provides the conditions which determine when it fails. This is what minimal hypotheses provide. However, given the above, we can see that there *is no conflict* between a heuristic approach to theories and realism. Rathkopf assumes that realism entails completeness, but this is false. There is no problem with being realists about each of the disjuncts in a contextualist function ascription, and hence no ontological worry about decomposition. Given that the contextualist view can also meet (reasonable analogues) of our epistemic aims regarding decomposition, there is every reason to consider contextualism a successful account of localization.

If the desiderata are met in such a way that we can be realists about the different function ascriptions made to different parts of the brain, then holism is avoided. We have functional categories for distinct parts of the brain, and those categories play a role in the systematic investigations of neuroscience. Contextualism, in addition, makes sense of the kinds of investigations that have occurred in the field, and the way in which those investigations have led to new discoveries about function. I have already articulated how the developments in MT investigation can be understood in terms of the posing and occasional overturning of minimal hypotheses. In the next section, I discuss some further aspects of this descriptive success.

5.6. Contextualism and Practice in Systems Neuroscience.

In chapter 1, I posed a problem for understanding explanatory frameworks in neuroscience: are they representational heuristics that are useful but limited, or are they fundamental truths about how the brain works? The contextualist reading of projectabilty makes sense of the problem and provides a novel perspective. The success of the MFH theory can be attributed to its adequate description of parts of the visual system in many instances. As I've stressed throughout, MT *is* a motion area, and nothing I've said denies that function attribution. Absolutist views mistakenly assume that, in order to contribute to successful functional localization, such ascriptions must project across all contexts. The contextualist view I have proposed accounts for both this success, and the subsequent use of the MFH theory as a heuristic guiding further analysis of MT. As Wimsatt (2007) stresses, even the most powerful heuristics are limited in their scope, perhaps inherently; theoretical frameworks, on his view, are simply not in the business of giving complete theories. What makes a heuristic useful, however, is not just how broadly applies, but also whether it can be productive when it breaks down—that is, do the *failures* of the heuristic point the way to more adequate theories? Reading projectability in terms of minimal hypotheses makes good sense of both how the MFH theory has contributed to the progress of the field, and how that progress has continued even after the field has moved more towards a contextualist approach.

The strategies in chapter 2 for maintaining the MFH theory in light of seemingly contradictory evidence set the heuristic borders of the MFH theory: evidence that can be accounted for with the strategies falls under the purview of the theory. In section 2.6, I discussed several results which can only be accommodated within the strategies with some strain (in particular, the results from Hegde and Van Essen, 2007), and in chapter 3 some results that, I argued, cannot be accounted for. In each case I've discussed, the standard motion-only account of MT function has played a key role in guiding the search for new function ascriptions, and specifying them. As discussed above, searching for new contexts that would overturn the minimal hypothesis has not only lead to the conclusion that depth and color can affect MT function, but elucidated the specific contexts and ways in which they do so. Without the minimal hypothesis as background, the search for new contextualist elements of function would be unconstrained; with it, clear hypotheses are generated and new results are interpretable. The transition from absolutist to contextualist accounts of MT function has been systematic, in a way that has vitally depended on the heuristic role of the standard, MFH theory.

Contextualism also provides a good description of some of the inherently pragmatic elements of systems neuroscience, and the ways in which different types of data interact in pursuing functional explanations. It is true that there are potentially limitless contexts that could be explored, and not all of them will be. There is thus a pragmatic element of choice in which contexts to explore, and which not, and these are often limited by available recording and behavioral techniques. The physiological studies I have discussed in monkeys, especially those involving behavioral measures, require considerable training time for the animals and are thus low-yield in terms of data. It is thus not possible that every conceivable context could be explored, and which ones actually are subjected to experimental analysis in the lab is, at some level, a personal choice of the experimenter. However, this choice is not groundless or unguided. I have stressed at several points how different kinds of data—particularly anatomical and psychophysical data—inspire particular physiological researches into new contexts. To recap just one example: despite the standard MFH view of MT as "color-blind," new anatomical results in the 1990s suggested that it in fact receives input from P-pathway cells in V1. This, combined with psychophysical results suggesting that perceivers can recognize motion defined entirely by color displacement, inspired Dobkins and Albright's research into potential color effects on MT.

The development can be read as follows: the standard account of MT provided a minimal hypothesis regarding color, namely that color never affects MT functioning. The psychophysical and anatomical results, however, suggested—without determining—that the minimal hypothesis might be incorrect, namely that there might be some contexts in which color information influences MT. The subsequent studies were an investigation of whether there were, in fact, such contexts, and if so what they were. A similar story could be told about depth. For instance, it was known for several years prior to the detailed exploration of depth representation in MT that MT receives input from disparity-selective cells in V2. Moreover, of course, it had been shown early on that disparity information affected MT cells. While Maunsell and Van Essen's application of the modulation strategy to cover this influence was very influential, there was still significant motivation in the physiological and anatomical data to pursue a more thorough exploration of depth representation in MT. While it took considerable time, these studies were eventually carried out, and led to new insights into MT's function.

Absolutism views these heuristic and pragmatic elements, which stem from the need to constrain and explore subsets of a potentially limitless number of contexts, as practical problems which theory must overcome. I have argued throughout that absolutist accounts cannot meet these goals. Rather than attempting to deny these aspects of the science, the contextualist account makes sense of the developments of in the field, adequately describes its successes, *and* in so doing meets the desiderata on a theory of functional localization. So much, at least, I have argued. It is true that, in meeting the desiderata, I have had to define projectability and decomposition in a contextualist way. To the absolutist, this will feel like giving up. But giving up on a project which has failed, and embracing one that informs the productive elements of the science, I contend, is better than holding projects in neuroscience up to a false explanatory ideal. Someone still motivated to deny contextualism will have to show that some vital aspect of explanation is missed by going contextualist. I have argued, that, at least as far as the standard desiderata go, contextualism covers the bases. In the final section, I discuss a range of subsidiary questions and some future directions.

5.7. Conclusion: Contextualism Going Forward.

The following complaint could be made about contextualism: the functional activities performed by parts of a system are supposed to be *productive* (Machamer, 2004). They are supposed to perform an operation that then has an effect on other parts in the system. It might be worried that contextualism is too descriptive to capture this productivity of functions. Contextualism, however, is not in principle different from the MFH theory in how it explains the influence of a part on the rest of the visual system. As on the MFH theory, MT's function is described in terms of its receiving input comprising certain forms of information, and integrating that information so that new information is reflected or represented in its responses. This new information then has further effects on the rest of the system. The difference is simply that absolutism suggests that one description must describe the process for each area in all contexts, while contextualism suggests that what information MT represents (and thus what effects it has on the rest of the system) can vary with context.

To see this more clearly, recall the "filtering" and "signaling" views of representation in cell responses, discussed in chapter 2. There is nothing in the idea of a filter that demands an absolutist view of what the filter does. Filters can of course be constructed to be selective for a range of properties—importantly, what the filter lets through in each case will depend on what is presented to it. The signaling view contends that there is some explicit encoding of the represented information in particular cell behaviors. Again, this is not incompatible with a contextualist view, since there is nothing wrong with the idea that different cell behaviors in different contexts could signal different kinds of information. Of course, it is then incumbent on some other property of the system to use the right information in the right ways—I discussed some of the challenges for understanding how this works in chapter 3. For now, however, suffice to say that the need for a new understanding of how signals from context-sensitive functional parts are used in the system is not a barrier to viewing those signals as productive (for instance, for making perceptual decisions) in the system.

A similar concern is the idea that a functional description should describe what an area does *intrinsically*, regardless of external conditions. One can perhaps see some of this idea at work in the motivations for CA, discussed in the previous chapter. Contextualism, it might be argued, fails to meet this intuition, since the functional ascription of an area can change depending on external conditions—e.g., on the perceptual stimulus. The notion of what is intrinsic to an area or part, however, admits of multiple interpretations. One can simply define it in an absolutist way, but then we are back in the same dialectic. The contextualist will respond by stressing the empirical inadequacy of these views and suggesting that how a part intrinsically functions depends on the set of particular contexts in which it functions, and the ways in which it does so.

The contextualist does, however, flatly disagree with the intrinsicality intuition in certain respects. Recall Anderson's (2010) view of why we need CA, discussed in chapter 3: he thinks that we must define the function of each part in a "functional complex" or network independently of the network (i.e., via a CA-style account) so that the network's behavior can then be explained in terms of the absolutist functions of its parts. Contextualism is friendly towards a different view, on which network properties get in on the ground floor of functional description—that is, we must explain some aspects of network properties in order to explain how a particular area functions, rather than the other way around. Consider the Dobkins and Albright results again. The impetus for the studies was (in part) the discovery of a new connection in the anatomical network of the visual system, namely between V1 P-pathway cells and MT. The switch between component and pattern motion representation in MT cells for color-segmented stimuli depends vitally on information (in this case, information about color identity) coming from this part of the network. Without that information, the responses remain unchanged. This suggests that what else is going on in the anatomical network in the visual system is vitally important for what an area does in a given context, *not* that the network should be explained in terms of what the particular areas do in any given context. Contextualism thus dovetails, quite intentionally, with the McIntosh's claim (discussed in Chapter 1) that "neural contexts" are vital in determining what an area does.

The first area for future exploration, then, is to more fully elaborate the relationship between contextualism and network views of brain anatomy. As mentioned briefly several times, graph theoretic approaches to brain anatomy are emerging as a powerful tool for understanding the organization of the brain. On their own, the categories of graph theory are not sufficient to localize function. Consider the claim that V4 is a hub node in the visual system (see fn. 5). In terms of the connection patterns that define a hub, V4 is similar to several other areas of the brain, including, for instance, prefrontal areas. So, just calling V4 a hub does not explain what it does distinctly from other hub nodes, and thus what its localized function is. Presumably, this will have something to do with the specific kinds of information that it receives as a hub node in the visual cortex, as well as the functional properties of the areas with which it is connected. That does not mean, however, that the designation of V4 as a hub node is irrelevant to uncovering its localized function. There may be general properties of how hub nodes distribute information, which can contribute to understanding how information spreads through the visual system in a contextually sensitive way. Not only is contextualism friendly to this possibility, it provides a normative claim for what should be done: namely, physiological accounts should look for contextual variation, and should do so in a way that considers the proposed functions of hub nodes in complex networks.

‘Considers’, alas, is merely a filler term at this point. Further work needs to be done to specify what kinds of contextual variation should be exhibited by hub nodes versus other graph-theoretic designations, and this is only one example of the kinds of designations that could be made. While much more research is needed, there is nothing mysterious about this process. I have already pointed out several instances in which anatomical and psychophysical data has constrained and inspired search for functional descriptions within physiological research. That contextualism extends this perspective is a point in its favor. The contextual view is inherently pluralist (Mitchell, 2002) in this respect, and further work should attempt to clarify how different explanatory frameworks are integrated under a contextualist perspective.

A second future direction involves extending the contextualist style of localization beyond the visual system. Given that the visual system is a textbook case of functional localization, any theory of localization should certainly account for it, but other systems raise other issues involving, for instance, intentional states (beliefs, desires, and intentions). There is hope, however, that even brain areas underlying these “higher” functions can be understood in contextualist manner. To take just one example: traditionally, neuroscientific theories of action and motor control have been modeled as a processing hierarchy analogous to the one posited in the MFH theory (although focusing more on the top-down aspect than the bottom-up). As I have argued elsewhere (Uithol, Burnston, & Haselager, 2014), there is reason to doubt this hierarchical view as well. Importantly, this claim is based on the view that what differentiates areas involved in action control is not that some correspond to abstract folk-psychological categories such as “intention” while others control specific parts of action, but instead that the different areas have different capacities in terms of the number, diversity, detail, and nesting of the action-situations they represent. Unsurprisingly, these discoveries have come through the manipulation of action-contexts—namely, variation in the properties of the action-situation to reveal new types of responses. While these categories are different than the ones varied in perceptual neuroscience, they are just as amenable to contextualist analysis.

Finally, there are a variety of ways of understanding the brain and perception that may need to be rethought on a thoroughly contextualist view of functional localization. For instance, can hierarchical views of object recognition in cognitive science and psychology be maintained if processing hierarchies are taken to be inadequate views of function in perceptual neuroscience? Or is there some sense of a processing hierarchy that can abandon absolutism, feature-specificity, and the use relation? More broadly, how should computational relationships between neural areas be construed if the signals stemming from each area are not feature-specific, and generally are context-dependent? The idea of a simple signal being encoded (or filtered) at each area, and then providing input to a variety of subsequent areas, as the MFH theory holds, is an easy way to understand computational relationships. However, it also may be false. Thinking about how computation works in the brain must be sensitive to what information is genuinely represented at each functional part of it.

These are questions for another day. They are difficult, and will require considerable elaboration of what contextualism about function is, how it relates to such notions as representation, and what these notions contribute to further notions such as computation. The difficulties, however, are also reasons for optimism. Given the shortcomings of absolutism, and the possibilities that contextualism affords for future advancements, it seems reasonable to conclude that a thoroughly contextualist perspective is worth a try.

References

Albright, T. D., & Stoner, G. R. (2002). Contextual influences on visual processing. *Annual Review of Neuroscience*, *25,* 339-379.

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *The Behavioral and Brain Sciences*, *33*(4), 245-266; discussion 266-313.

Anzai, A., Chowdhury, S. A., & DeAngelis, G. C. (2011). Coding of stereoscopic depth information in visual areas V3 and V3A. *The Journal of Neuroscience, 31*(28), 10270-10282.

Anzai, A., & DeAngelis, G. C. (2010). Neural computations underlying depth perception. *Current Opinion in Neurobiology, 20*(3), 367-375.

Ariew, A., Cummins, R. C., & Perlman, M. (Eds.) (2002). Functions: New essays in the philosophy of psychology and biology. Oxford: Oxford University Press.

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, *38*(6), 795-798.

Bartels, A., & Zeki, S. (1998). The theory of multistage integration in the visual brain. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 265*(1412), 2327-2332.

Bechtel, W. (2001). Decomposing and localizing vision: An exemplar for cognitive neuroscience. In W. Bechtel, P. Mandik, J. Mundale, and R. S. Stufflebeam (Eds.), *Philosophy and the neurosciences: A reader* (pp. 225-249). Malden, MA: Blackwell.

Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. New York: Routledge.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 36*(2), 421-441.

Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity: Decomposition and localization as scientific research strategies*. Princeton, NJ: Princeton University Press.

Bergeron, V. (2007). Anatomical and functional modularity in cognitive science: Shifting the focus. *Philosophical Psychology, 20*(2), 175-195.

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *The Behavioral and Brain Sciences, 30*(5-6), 481-499.

Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT. *Annual Review of Neuroscience, 28,* 157-189.

Bouquet, P., Serafini, L., & Thomason, R. H. (Eds.) (2008). *Perspectives on contexts.* Stanford, CA: CSLI Publications.

Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., & Movshon, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Visual Neuroscience, 13,* 87-100.

Broca, P. P. (1861). Loss of speech, chronic softening and partial destruction of the anterior left lobe of the brain. *Bulletin de la Société Anthropologique, 2,* 235-238.

Buller, D. J., & Hardcastle, V. G. (2000). Evolutionary psychology, meet developmental neurobiology: Against promiscuous modularity. *Brain and Mind, 1,* 307-325.

Cappelen, H., & Lepore, E. (2005). *Insensitive semantics: A defense of semantic minimalism and speech act pluralism.* Malden MA: Blackwell.

Chirimuuta, M. (2014). Minimal models and canonical neural computations: The distinctness of computational explanation in neuroscience. *Synthese, 191*(2), 127-153.

Chirimuuta, M., & Gold, I. (2009). The embedded neuron, the enactive field? In J. Bickle (Ed.), *The Oxford handbook of philosophy and neuroscience* (pp. 200-225). New York: Oxford University Press.

Churchland, P. S., & Sejnowski, T. J. (1992). *The computational brain*. Cambridge, MA: MIT press.

Clark, A. (2009). Perception, action, and experience: Unraveling the golden braid. *Neuropsychologia, 47*(6), 1460-1468.

Cohen, I. B. (1985). *The birth of a new physics*. New York: WW Norton & Company.

Connor, C. E., Brincat, S. L., & Pasupathy, A. (2007). Transformation of shape information in the ventral pathway. *Current opinion in neurobiology, 2,* 140-147.

Craver, C. F. (2007). *Explaining the brain*. Oxford: Oxford University Press.

Craver, C. F., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy, 22*(4), 547-563.

Cumming, B. G., & Parker, A. J. (2000). Local disparity not perceived depth is signaled by binocular neurons in cortical area V1 of the macaque. *The Journal of Neuroscience, 20*(12), 4758-4767.

Cummins, R. C. (1975). Functional analysis. *Journal of Philosophy, 72*(20), 741-765.

da F. Costa, L., & Sporns, O. (2006). Hierarchical features of large-scale cortical connectivity. *The European Physical Journal B, 48*(4), 567-573.

DeAngelis, G. C., Cumming, B. G., & Newsome, W. T. (1998). Cortical area MT and the perception of stereoscopic depth. *Nature, 394*(6694), 677-680.

DeAngelis, G. C., & Newsome, W. T. (2004). Perceptual “read-out” of conjoined direction and disparity maps in extrastriate area MT. *PLoS Biology, 2*(3), e77.

DeAngelis, G. C., & Uka, T. (2003). Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *Journal of Neurophysiology, 89*(2), 1094-1111.

Desimone, R., & Schein, S. J. (1987). Visual properties of neurons in area V4 of the macaque: Sensitivity to stimulus form. *Journal of Neurophysiology, 57*(3), 835-868.

DeYoe, E. A., & Van Essen, D. C. (1988). Concurrent processing streams in monkey visual cortex. *Trends in Neurosciences, 11*(5), 219-226.

Dobkins, K., & Albright, T. (1990). *Color facilitates motion correspondence in visual area MT.* Paper presented at the Society for Neuroscience Abstracts.

Dobkins, K. R., & Albright, T. D. (1994). What happens if it changes color when it moves? The nature of chromatic input to macaque visual area MT. *The Journal of Neuroscience, 14*(8), 4854-4870.

Dobkins, K. R., & Albright, T. D. (2004). Merging processing streams: Color cues for motion detection and interpretation. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (pp. 1217-1228). Cambridge, MA: MIT Press.

Dobkins, K. R., Rezec, A. A., & Krekelberg, B. (2007). Effects of spatial attention and salience cues on chromatic and achromatic motion processing. *Vision Research, 47*(14), 1893-1906.

Dobkins, K. R., Stoner, G. R., & Albright, T. D. (1998). Perceptual, oculomotor, and neural responses to moving color plaids. *Perception, 27,* 681-709.

Dodd, J. V., Krug, K., Cumming, B. G., & Parker, a. J. (2001). Perceptually bistable three-dimensional figures evoke high choice probabilities in cortical area MT. *The Journal of Neuroscience, 21*(13), 4809-4821.

Egan, A. (2012). Relativist dispositional theories of value. *Southern Journal of Philosophy, 50*(4), 557-582.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1*(1), 1-47.

Flourens, P. (1846). *Phrenology examined*. Hogan & Thompson.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT press.

Fodor, J. A. (1990). *A theory of content and other essays*. Cambridge, MA: MIT Press.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127-138.

Friston, K. J., & Price, C. J. (2011). Modules and brain mapping. *Cognitive Neuropsychology, 28*(3-4), 241-250.

Gall, F. J. (1825). *Sur les fonctions du cerveau et sur celles de chacune de ses parties*. J.-B. Baillière.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience, 3*(2), 191-197.

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 431.

Gegenfurtner, K. R., Kiper, D. C., Beusmans, J. M., Carandini, M., Zaidi, Q., & Movshon, J. A. (1994). Chromatic properties of neurons in macaque MT. *Visual Neuroscience, 11*(3), 455-466.

Grafton, S. T., & de C Hamilton, A. F. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Human Movement Science, 26*(4), 590-616.

Graziano, M. S. A., & Aflalo, T. N. (2007). Rethinking cortical organization: Moving away from discrete areas arranged in hierarchies. *The Neuroscientist, 13*(2), 138-147.

Griesemer, J. (2012). Formalization and the meaning of "theory" in the inexact biological sciences. *Biological Theory, 7*(4), 298-310.

Griesemer, J. R. (2005). The informational gene and the substantial body: On the generalization of evolutionary theory by abstraction. In M. R. Jones & N. Cartwright (Eds.), *Idealization XII: Correcting the Model—Idealization and Abstraction in the Sciences* (pp. 59-115). Amsterdam: Rodopi.

Grill-Spector, K., Sayres, R., & Ress, D. (2006). High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. *Nature Neuroscience, 9*(9), 1177-1185.

Gross, C., Bender, D., & Rocha-Miranda, C. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science, 166*(3910), 1303-1306.

Gross, C. G., Rocha-Miranda, C., & Bender, D. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. *Journal of Neurophysiology, 35*(1), 96-111.

Grunewald, A., Bradley, D. C., & Andersen, R. A. (2002). Neural correlates of structure-from-motion perception in macaque V1 and MT. *The Journal of Neuroscience, 22*(14), 6195-6207.

Haggard, P. (2005). Conscious intention and motor cognition. *Trends in Cognitive Sciences, 9*(6), 290-295.

Hanson, S. J., & Bunzl, M. (2010). *Foundational issues in human brain mapping*. Cambridge, MA: MIT Press.

Hardcastle, V. G., & Stewart, C. M. (2005). Localization in the brain and other illusions. In A. Brook (Ed.), *Cognition and the Brain* (pp. 27-39). Cambridge: Cambridge University Press.

Hartline, H. K. (1938). The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *American Journal of Physiology, 121*, 400-415.

Haxby, J. V. (2006). Fine structure in representations of faces and objects. *Nature Neuroscience, 9*(9), 1084-1086.

Hegdé, J., & Felleman, D. J. (2007). Reappraising the functional implications of the primate visual anatomical hierarchy. *The Neuroscientist, 13*(5), 416-421.

Hegdé, J., & Van Essen, D. C. (2007). A comparative study of shape representation in macaque visual areas v2 and v4. *Cerebral Cortex, 17*(5), 1100-1116.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology, 148*(3), 574-591.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology, 160*(1), 106-154.

Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology, 28*(2), 229-289.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology, 195*(1), 215-243.

Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences, 107*(25), 11163-11170.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience, 17*(11), 4302-4311.

Klein, C. (2012). Cognitive ontology and region- versus network-oriented analyses. *Philosophy of Science, 79*(5), 952-960.

Kohn, A., & Movshon, J. A. (2003). Neuronal adaptation to visual motion in area MT of the macaque. *Neuron, 39*(4), 681-691.

Kooi, F. L., De Valois, K. K., Switkes, E., & Grosof, D. H. (1992). Higher-order factors influencing the perception of sliding and coherence of a plaid. *Perception, 21*, 583-583.

Krug, K., & Parker, A. J. (2011). Neurons in dorsal visual area V5/MT signal relative disparity. *The Journal of Neuroscience, 31*(49), 17892-17904.

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology, 16*(1), 37-68.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23*(11), 571-579.

Lashley, K. S. (1929). *Brain mechanisms and intelligence: A quantitative study of injuries to the brain*. Chicago: University of Chicago Press.

Lashley, K. S. (1930). Basic neural mechanisms in behavior. *Psychological Review, 37*(1), 1-24.

Lashley, K. S. (1950). *In search of the engram.* Paper presented at the Symposia of the society for experimental biology.

Lee, T. S., & Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. *Proceedings of the National Academy of Sciences, 98*(4), 1907-1911.

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1959). What the frog's eye tells the frog's brain. *Proceedings of the IRE, 47*(11), 1940-1951.

Levy, A., & Bechtel, W. (2013). Abstraction and the organization of mechanisms. *Philosophy of Science, 80*(2), 241-261.

Levy, W. B., Hocking, A. B., & Wu, X. (2005). Interpreting hippocampal function as recoding and forecasting. *Neural Networks, 18*(9), 1242-1264.

Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science, 240*(4853), 740-749.

Locke, J. (1700). *An essay concerning human understanding*.

Love, A. C. (2012). Hierarchy, causation and explanation: Ubiquity, locality and pluralism. *Interface Focus, 2*(1), 115-125.

Machamer, P. (2004). Activities and causation: The metaphysics and epistemology of mechanisms. *International Studies in the Philosophy of Science, 18*(1), 27-39.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67,* 1-25.

Marr, D., & Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 211*(1183), 151-180.

Maunsell, J. H., & Van Essen, D. C. (1983a). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *Journal of Neurophysiology, 49*(5), 1127-1147.

Maunsell, J. H., & Van Essen, D. C. (1983b). Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *Journal of Neurophysiology, 49*(5), 1148-1167.

Maunsell, J. H., & Van Essen, D. C. (1987). Topographic organization of the middle temporal visual area in the macaque monkey: Representational biases and the relationship to callosal connections and myeloarchitectonic boundaries. *Journal of Comparative Neurology, 266*(4), 535-555.

McIntosh, A. R. (1999). Mapping cognition to the brain through neural interactions. *Memory, 7*(5-6), 523-548.

McIntosh, A. R. (2004). Contexts and catalysts: A resolution of the localization and integration of function in the brain. *Neuroinformatics, 2*(2), 175-182.

Meunier, D., Lambiotte, R., & Bullmore, E. T. (2010). Modular and hierarchically modular organization of brain networks. *Frontiers in Neuroscience, 4*, 200-200.

Millikan, R. G. (1989). Biosemantics. *The Journal of Philosophy, 86*(6), 281-297.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences, 6*, 414-417.

Mitchell, S. D. (2002). Integrative pluralism. *Biology and Philosophy, 17*(1), 55-70.

Müller-Linow, M., Hilgetag, C. C., & Hütt, M.-T. (2008). Organization of excitable dynamics in hierarchical biological networks. *PLoS Computational Biology, 4*(9), e1000190-e1000190.

Nassi, J. J., Lyon, D. C., & Callaway, E. M. (2006). The parvocellular LGN provides a robust disynaptic input to the visual motion area MT. *Neuron, 50*(2), 319-327.

Nguyenkim, J. D., & DeAngelis, G. C. (2003). Disparity-based coding of three-dimensional surface orientation by macaque middle temporal neurons. *The Journal of Neuroscience, 23*(18), 7117-7128.

Nowlan, S. J., & Sejnowski, T. J. (1994). Filter selection model for motion segmentation and velocity integration. *JOSA A, 11*(12), 3177-3200.

Nowlan, S. J., & Sejnowski, T. J. (1995). A selection model for motion processing in area MT of primates. *The Journal of Neuroscience, 15*(2), 1195-1214.

Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition, 107*(1), 179-217.

Palanca, B. J. A., & DeAngelis, G. C. (2003). Macaque middle temporal neurons signal depth in the absence of motion. *The Journal of Neuroscience, 23*(20), 7647-7658.

Parker, A. J., & Newsome, W. T. (1998). Sense and the single neuron: probing the physiology of perception. *Annual Review of Neuroscience, 21*(1), 227-277.

Penfield, W., & Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain, 60*, 389-443.

Phillips, C., Zeki, S., & Barlow, H. (1984). Localization of function in the cerebral cortex: past, present and future. *Brain, 107*(1), 328-361.

Piccinini, G., & Bahar, S. (2013). Neural computation and the computational theory of cognition. *Cognitive Science, 3*(7), 453-488.

Piccinini, G., & Scarantino, A. (2011). Information processing, computation, and cognition. *Journal of Biological Physics, 37*(1), 1-38.

Pitcher, D., Walsh, V., Yovel, G., & Duchaine, B. (2007). TMS evidence for the involvement of the right occipital face area in early face processing. *Current Biology, 17*(18), 1568-1573.

Poggio, G. F., & Fischer, B. (1977). Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *Journal of Neurophysiology, 40*(6), 1392-1405.

Preyer, G., & Peter, G. (Eds.). (2005). Contextualism in philosophy: Knowledge, meaning, and truth. Oxford: Oxford University Press.

Price, C. J., & Friston, K. J. (2005). Functional ontologies for cognition: The systematic definition of structure and function. *Cognitive Neuropsychology, 22*(3), 262-275.

Prinz, J. (2006). Is the mind really modular. In R. J. Stainton (Ed.), *Contemporary Debates In Cognitive Science* (pp. 22-36). Cambridge, MA: Blackwell.

Purushothaman, G., & Bradley, D. C. (2004). Neural population code for fine perceptual decisions in area MT. *Nature Neuroscience, 8*(1), 99-106.

Rathkopf, C. A. (2013). Localization and intrinsic function. *Philosophy of Science, 80*(1), 1-21.

Recanati, F. (2005). Literalism and contextualism: Some varieties. In G. Preyer and G. Peter (Eds.), *Contextualism in philosophy: Knowledge, meaning, and truth* (pp. 171-196). Oxford: Oxford University Press.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*(11), 1019-1025.

Roe, A. W., Chelazzi, L., Connor, C. E., Conway, B. R., Fujita, I., Gallant, J. L., . . . Vanduffel, W. (2012). Toward a unified theory of visual area V4. *Neuron, 74*(1), 12-29.

Rowlands, M. (2009). Situated representation. In M. Aydede & P. Robbins (Eds.), *The Cambridge handbook of situated cognition* (pp. 117-133). Cambridge: Cambridge University Press.

Sanada, T. M., Nguyenkim, J. D., & DeAngelis, G. C. (2012). Representation of 3-D surface orientation by velocity and disparity gradient cues in area MT. *Journal of Neurophysiology, 107*(8), 2109-2122.

Sanger, T. D. (2003). Neural population codes. *Current Opinion in Neurobiology, 13*(2), 238-249.

Savic, I., Gulyas, B., Larsson, M., & Roland, P. (2000). Olfactory functions are mediated by parallel and hierarchical processing. *Neuron*, *26*(3), 735-745.

Schiller, P. (1993). The effects of V4 and middle temporal (MT) area lesions on visual performance in the rhesus monkey. *Visual Neuroscience*, *10*(4), 717-746.

Shadlen, M. N., Britten, K. H., Newsome, W. T., & Movshon, J. A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *The Journal of Neuroscience, 16*(4), 1486-1510.

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society, 106*(6), 467-482.

Simoncelli, E. P., & Heeger, D. J. (1998). A model of neuronal responses in visual area MT. *Vision Research, 38*(5), 743-761.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience, 18*(1), 555-586.

Snowden, R. J., Treue, S., Erickson, R. G., & Andersen, R. A. (1991). The response of area MT and V1 neurons to transparent motion. *The Journal of Neuroscience, 11*(9), 2768-2785.

Spitzer, H., Desimone, R., & Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science, 240*(4850), 338-340.

Sporns, O. (2011). *Networks of the BRAIN*. Cambridge, MA: MIT Press.

Spurzheim, J. G. (1829). *Outlines of phrenology: Being also a manual of reference for the marked busts*. Treuttel, Wurtz, and Richter.

Stoner, G., Albright, T. D., & Ramachandran, V. (1990). Transparency and coherence in human motion perception. *Nature, 344*, 153-155.

Tanigawa, H., Lu, H. D., & Roe, A. W. (2010). Functional organization for color and orientation in macaque V4. *Nature Neuroscience, 13*(12), 1542-1548.

Thiele, A., Dobkins, K. R., & Albright, T. D. (2001). Neural correlates of chromatic motion perception. *Neuron, 32*(2), 351-358.

Thiele, A., Rezec, A., & Dobkins, K. R. (2002). Chromatic input to motion processing in the absence of attention. *Vision Research, 42*(11), 1395-1401.

Tizard, B. (1959). Theories of brain localization from Flourens to Lashley. *Medical History, 3*(2), 132.

Tolias, A. S., Keliris, G. A., Smirnakis, S. M., & Logothetis, N. K. (2005). Neurons in macaque area V4 acquire directional tuning after adaptation to motion stimuli. *Nature Neuroscience, 8*(5), 591-593.

Treue, S., & Maunsell, J. H. (1999). Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. *The Journal of Neuroscience, 19*(17), 7591-7602.

Treue, S., & Trujillo, J. C. M. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature, 399*(6736), 575-579.

Uithol, S., Burnston, D. C., & Haselager, P. (2014). Why we may not find intentions in the brain. *Neuropsychologia, 56*, 129-139.

Uka, T., & DeAngelis, G. C. (2003). Contribution of middle temporal area to coarse depth discrimination: comparison of neuronal and psychophysical sensitivity. *The Journal of Neuroscience, 23*(8), 3515-3530.

Uka, T., & DeAngelis, G. C. (2004). Contribution of area MT to stereoscopic depth perception: choice-related response modulations reflect task strategy. *Neuron, 42*(2), 297-310.

Uka, T., & DeAngelis, G. C. (2006). Linking neural representation to function in stereoscopic depth perception: roles of the middle temporal area in coarse versus fine disparity discrimination. *The Journal of Neuroscience, 26*(25), 6791-6802.

Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences, 11*(2), 58-64.

Ungerleider, L. G., Galkin, T. W., Desimone, R., & Gattass, R. (2008). Cortical connections of area V4 in the macaque. *Cerebral Corte, 18*(3), 477-499.

Uttal, W. R. (2001). *The new phrenology: The limits of localizing cognitive processes in the brain*. Cambridge, MA: The MIT Press.

Van Essen, D. C., & Anderson, C. H. (1995). Information processing strategies and pathways in the primate visual system. In S. F. Zornetzer, J. L. Davis, C. Lau, and T. McKenna (Eds.), *An introduction to neural and electronic networks* (pp. 45-76). San Diego, CA: Academic Press.

Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: An integrated systems perspective. *Science, 255*(5043), 419-423.

Van Essen, D. C., & Deyoe, E. A. (1995). Concurrent processing in the primate visual cortex. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 383-400). Cambridge, MA: MIT Press.

Van Essen, D. C., & Gallant, J. L. (1994). Neural mechanisms of form and motion processing in the primate visual system. *Neuron, 13*(1), 1-10.

Weismann, A. (1893). *The germ-plasm: A theory of heredity*. C. Scribner's Sons.

Wessinger, C., VanMeter, J., Tian, B., Van Lare, J., Pekar, J., & Rauschecker, J. (2001). Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *Journal of Cognitive Neuroscience, 13*(1), 1-7.

Wilson, E. B. (1900). *The cell in development and inheritance*: Macmillan.

Wimsatt, W. C. (2007). *Re-engineering philosophy for limited beings: Piecewise approximations to reality*. Cambridge, MA: Harvard University Press.

Young, M. P., Hilgetag, C. C., & Scannell, J. W. (2000). On imputing function to structure from the behavioural effects of brain lesions. *Philosophical Transactions of the Royal Society of London. Series B: Biological sciences, 355*(1393), 147-161.

Zeki, S. (1983). Colour coding in the cerebral cortex: the responses of wavelength-selective and colour-coded cells in monkey visual cortex to changes in wavelength composition. *Neuroscience, 9*(4), 767-781.

Zeki, S. (2001). Localization and globalization in conscious vision. *Annual Review of Neuroscience, 24*(1), 57-86.

Zeki, S., & Bartels, A. (1998). The autonomy of the visual systems and the modularity of conscious vision. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 353*(1377), 1911-1914.

Zeki, S., & Shipp, S. (1988). The functional logic of cortical connections. *Nature, 335*(6188), 311-317.

Zeki, S., Watson, J., Lueck, C., Friston, K. J., Kennard, C., & Frackowiak, R. (1991). A direct demonstration of functional specialization in human visual cortex. *The Journal of Neuroscience, 11*(3), 641-649.

Zeki, S. M. (1974). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *The Journal of Physiology, 236*(3), 549-573.

Zeki, S. M. (1977). Colour coding in the superior temporal sulcus of rhesus monkey visual cortex. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 195-223.

Zeki, S. M. (1978). Functional specialisation in the visual cortex of the rhesus monkey. *Nature, 274*(5670), 423-428.

Zhou, H., Friedman, H. S., & Von Der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *The Journal of Neuroscience, 20*(17), 6594-6611.

Zola-Morgan, S. (1995). Localization of brain function: The legacy of Franz Joseph Gall (1758-1828). *Annual Review of Neuroscience, 18*(1), 359-383.